Time Series Analysis of Arctic Sea Ice

Hector G. Flores Rodriguez*, Hernando Ombao†

Abstract

The decline of Arctic sea ice in recent decades has promoted great research for its effects on global climate change. This work aims to further investigate the ongoing trend of the decrease in sea ice concentration. We utilize a non-parametric approach to determine significant change points where the possiblity of a significant trend might occur. We noted the significance of the 1997 change point along with the 1997-1998 El Niño Southern Oscillation (ENSO). We then fitted a piecewise linear trend model to detrend the series to further investigate periodicity and any temporal correlations. From our detrended series, we modeled the seasonality component via harmonic regression. This led to developing an AR(2) model that captures the majority of the structure in the Arctic sea ice concentration data set.

Keywords: Climate modeling; Spline regression analysis; Arctic sea ice concentration; Change point analysis; Global warming, Autoregression, Climate forecasting.

*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

The retreat of the Arctic sea ice is one of the most visible indicators of global climate change (Zhang and Walsh, 2006). A potential reason as to why climate change is tied to the Arctic sea ice is hat climate change is amplified in the Arctic (Pithan and Mauritsen, 2014). The decrease in sea ice concentration as an indicator for climate change has stimulated current research into investigating the trends of sea ice. There is a near consensus among current climate models that Arctic sea ice extent will decline through the 21st century. The aim of this work is to further study in detail the temporal evolution of sea ice concentration. We aim to develop methods to determine significant change points in the time series along with developing a piecewise linear trend model that best fits the data. In doing so, we can remove any trends from the data to study the periodicity of the Arctic sea and develop an AR(2) model with the prospect of forecasting.

The structure of this paper is organized as follows: Section 2 provides background to our proposed models; Section 3 analyzes the Arctic sea ice concentration data set; Section 4 summarizes our results; Section 5 concludes this work.

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 (Ruppert et al., 2003), illustrated in Figure 1.

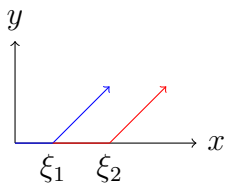

Figure 1: Linear spline basis.

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 predefined 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) (Burrus et al., 1994) 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

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

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 density (PSD)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 (Hayes, 1996). 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 (Shumway and Stoffer, 2005). 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.

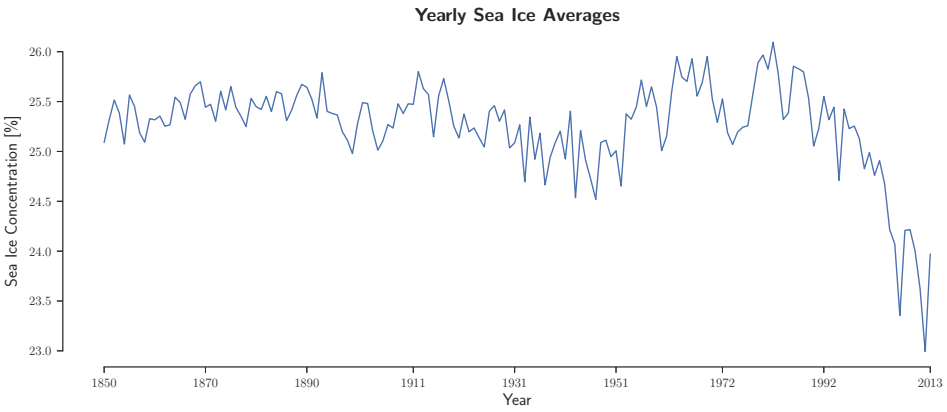

Figure 2: Yearly sea ice concentration averages.

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.

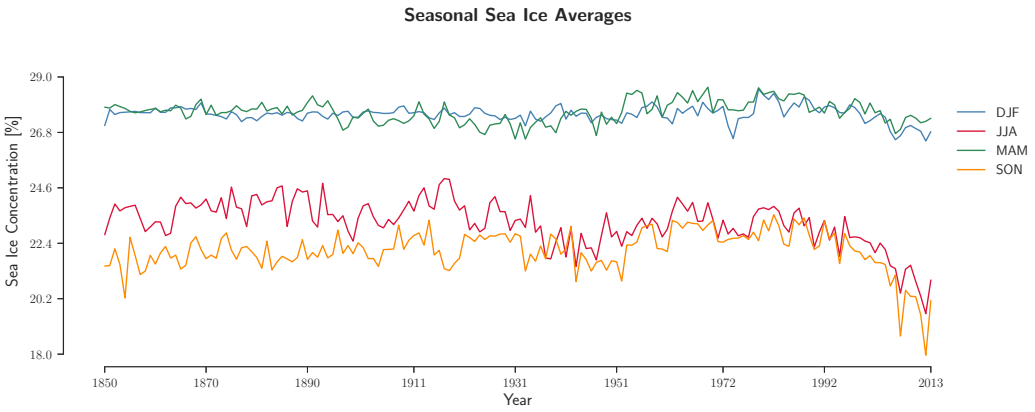

Figure 3: Seasonal sea ice concentration averages.

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

$$CP_{djf} = \{1935, 1944, 1980, 1997\}$$

$$CP_{mam} = \{1896, 1929, 1935, 1953, 1980, 1997\}$$

$$CP_{jja} = \{1917, 1943, 1963, 1979, 1997\}$$

$$CP_{son} = \{1915, 1950, 1980, 1997\}$$

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	$\{1935, 1980\}$
JJA	$\{1917, 1943, 1997\}$
SON	$\{1915, 1950, 1980, 1997\}$

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

The linear spline trends for summer months are shown in Figure 4

After fitting each season, we detrended each of the series by removing the fitted linear spline regression model from the original data to further study the periodicity of the Arctic ice concentration.

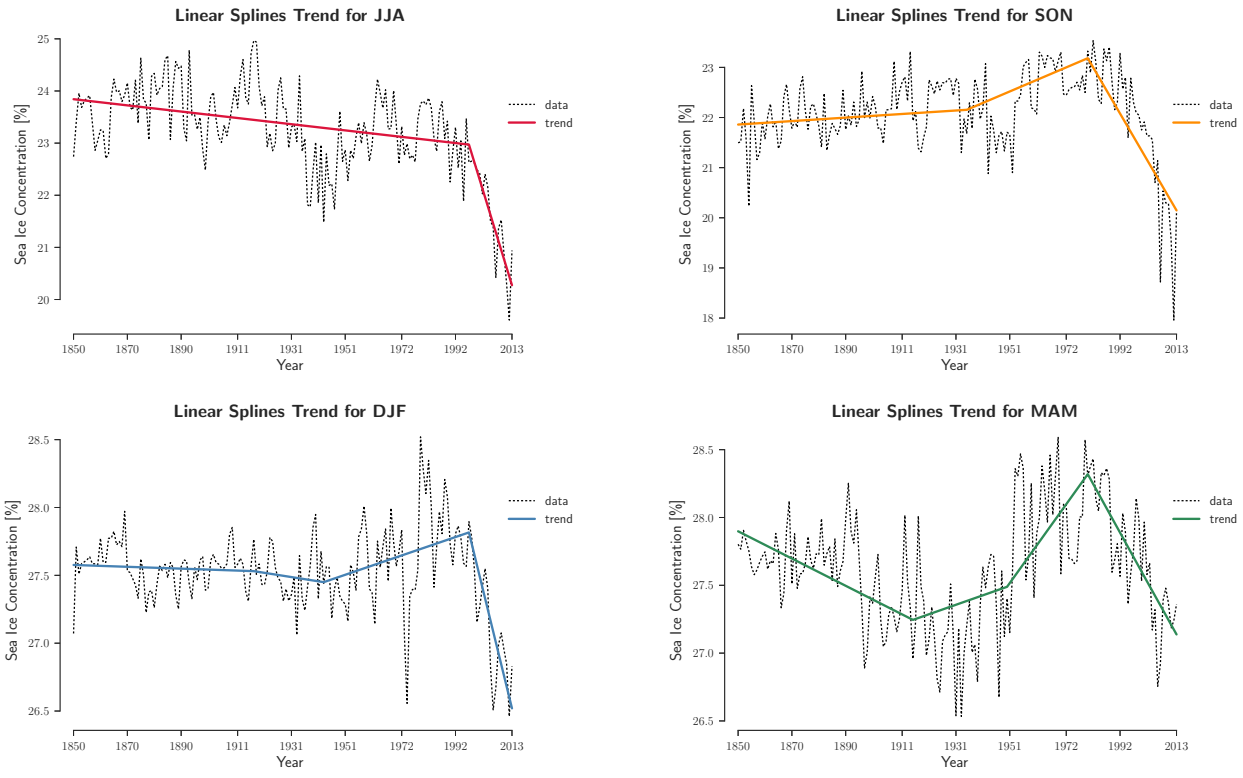

Figure 4: Seasonal sea ice concentration averages.

3.3 Harmonic Regression

The goal of harmonic regression is to first determine the dominant frequencies from our estimated power spectrum. Using the top dominant frequencies for each season, we can estimate our seasonality component of our model, as presented in equation (3). The power spectrum estimates for each season are shown in Figure 5 and the corresponding dominant frequencies are shown in Table 2.

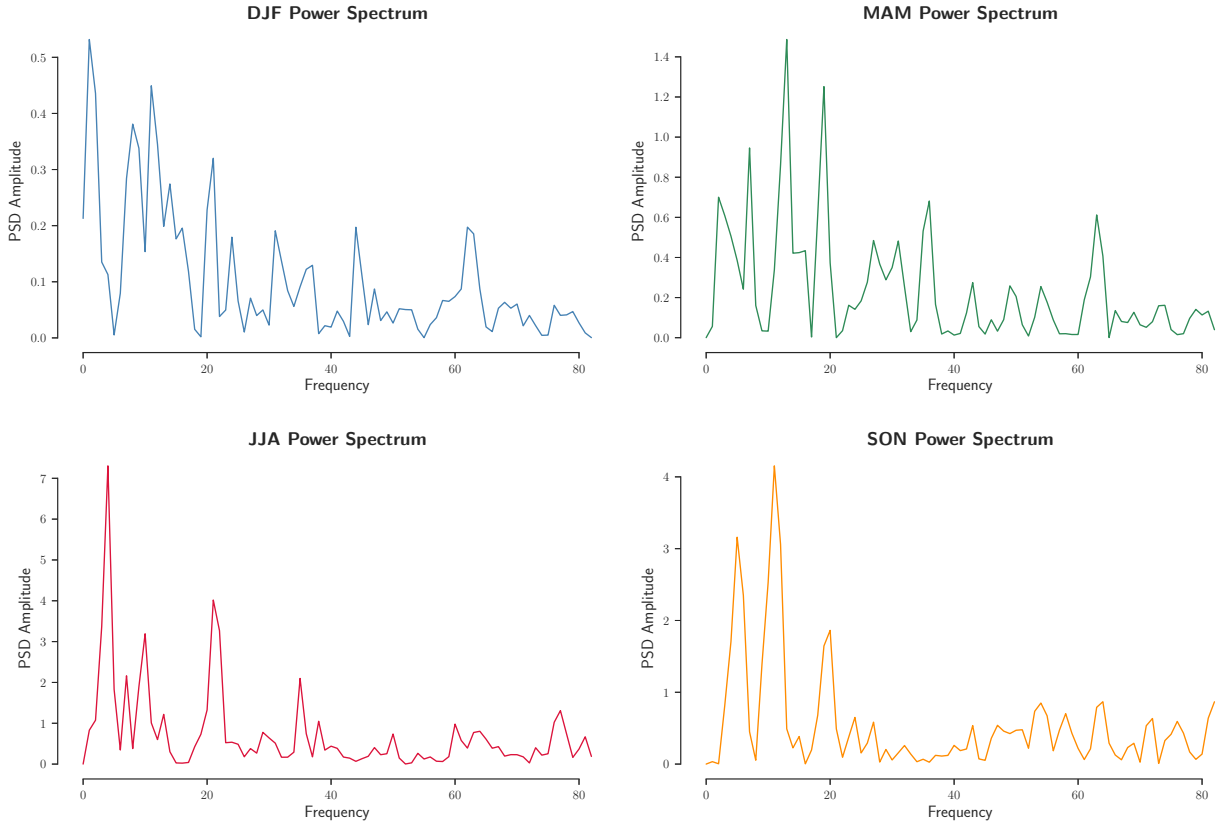

Figure 5: Power spectrum estimates for DJF, MAM, JJA, and SON.

Season	**Top 5 Dominant Frequencies**
DJF	$\omega_1, \omega_{11}, \omega_2, \omega_8, \omega_{12}$
MAM	$\omega_{13}, \omega_{19}, \omega_7, \omega_{12}, \omega_2$
JJA	$\omega_4, \omega_{21}, \omega_3, \omega_{22}, \omega_{10}$
SON	$\omega_{11}, \omega_5, \omega_{12}, \omega_{10}, \omega_6$

Table 2: Top five frequencies in order of magnitude.

The power spectrum for the warmer months differ vastly from the colder months. This result is consistent with the Figure 3, where the variability is much greater in the warmer months. Although the power spectrum estimates capture a lot of noise, we can use the top five dominant frequencies and determine the most prevalent harmonic components using a greedy backward selection, with a significance level of 5%.

Season	**Harmonic Frequencies**
DJF	$\cos(\omega_2), \cos(\omega_{11})$
MAM	$\sin(\omega_2), \cos(\omega_2), \sin(\omega_{13}), \cos(\omega_{19})$
JJA	$\sin(\omega_3), \cos(\omega_3), \sin(\omega_{10}), \cos(\omega_{10})$
SON	$\sin(\omega_5), \cos(\omega_5)$

Table 3: Resulting harmonic components after multiple comparison test.

The significant harmonics in Table 3 are used to estimate the seasonality component of each season. We then remove the seasonality component from our residuals, resulting in a detrended and deseasonalized series, equation (4). With the deaseasonalized residuals, we can then proceed to developing ARMA models and determine the best model that captures

the structure in the Arctic ice concentration.

3.4 Arctic Ice AR Modeling

The goal of developing an autoregressive model for the Arctic sea ice concentration is to determine if the present values can be plausibly modeled in terms of the past values. The process of selecting our final model begins with the investigation of the ACF and PACF of our residuals. The ACF and PACF will enable us to see if there is any structure that might suggest the type of model to use along with the order of the model.

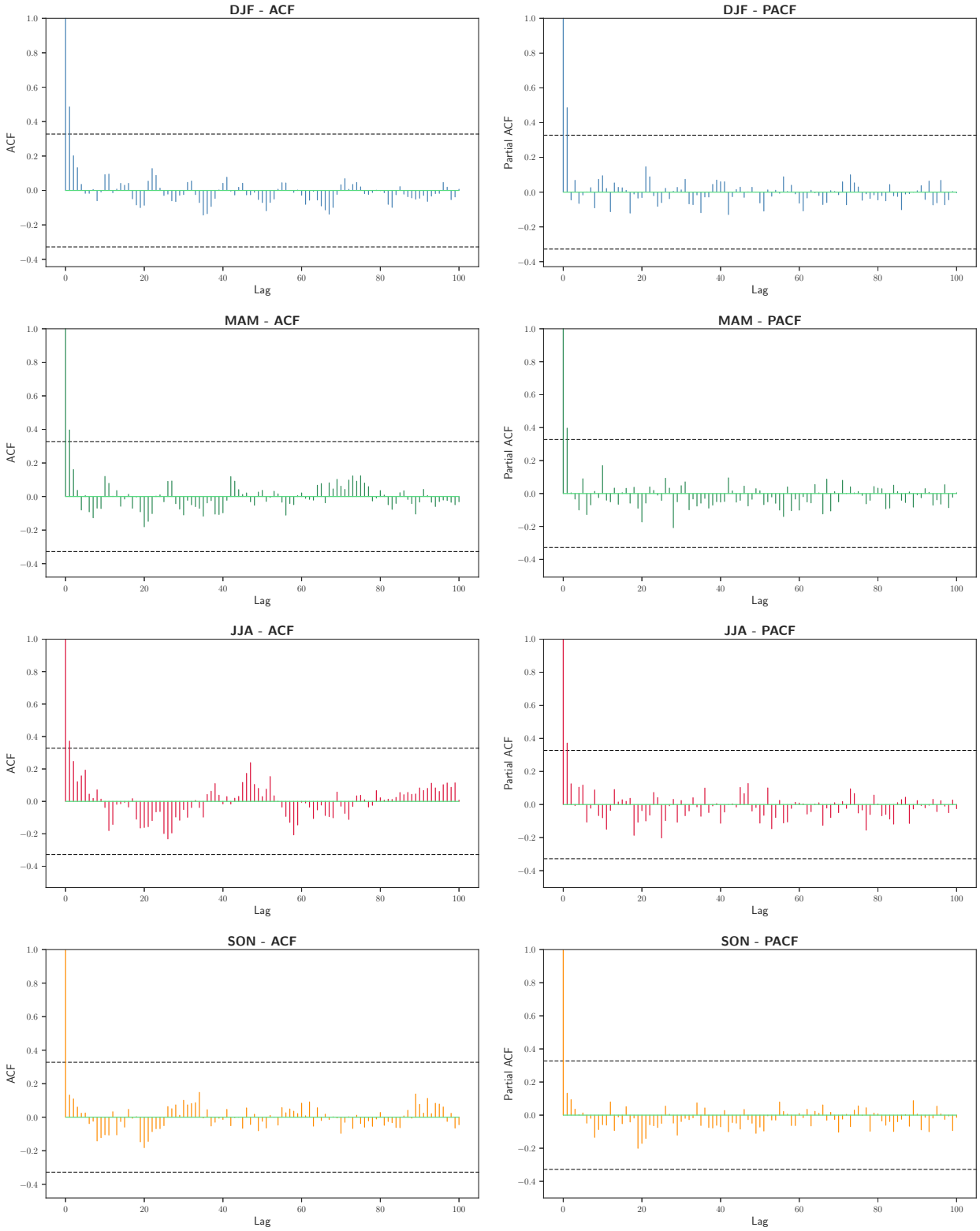

Figure 6: DJF - ACF and PACF

Figure 6 suggest AR(1) and AR(2). However, the ACF and PACF for SON has no significant lag values. Regardless, we fitted an: AR(1), AR(2), AR(3), and ARMA(2,1) for all seasons. Both AIC and BIC were used in our model selection. The resulting models were AR(2) for all seasons.

4 Results

A fundamental procedure in time series analysis is testing whether a series is white noise or its process contains a more complex structure. The Ljung-Box test, a well known procedure for checking whether a sequence is white noise or not (Ljung and Box, 1978), is defined as

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k, and h is the number of lags being tested.

To ensure our models were appropriate (i.e., good estimates of trend and periodicity), we tested the residuals for "whiteness" by:

- Analyzing the ACF and PACF plots for the residuals
- Performing the Ljung-Box test for up to 10 lags
- QQ plot for visually checking for normality

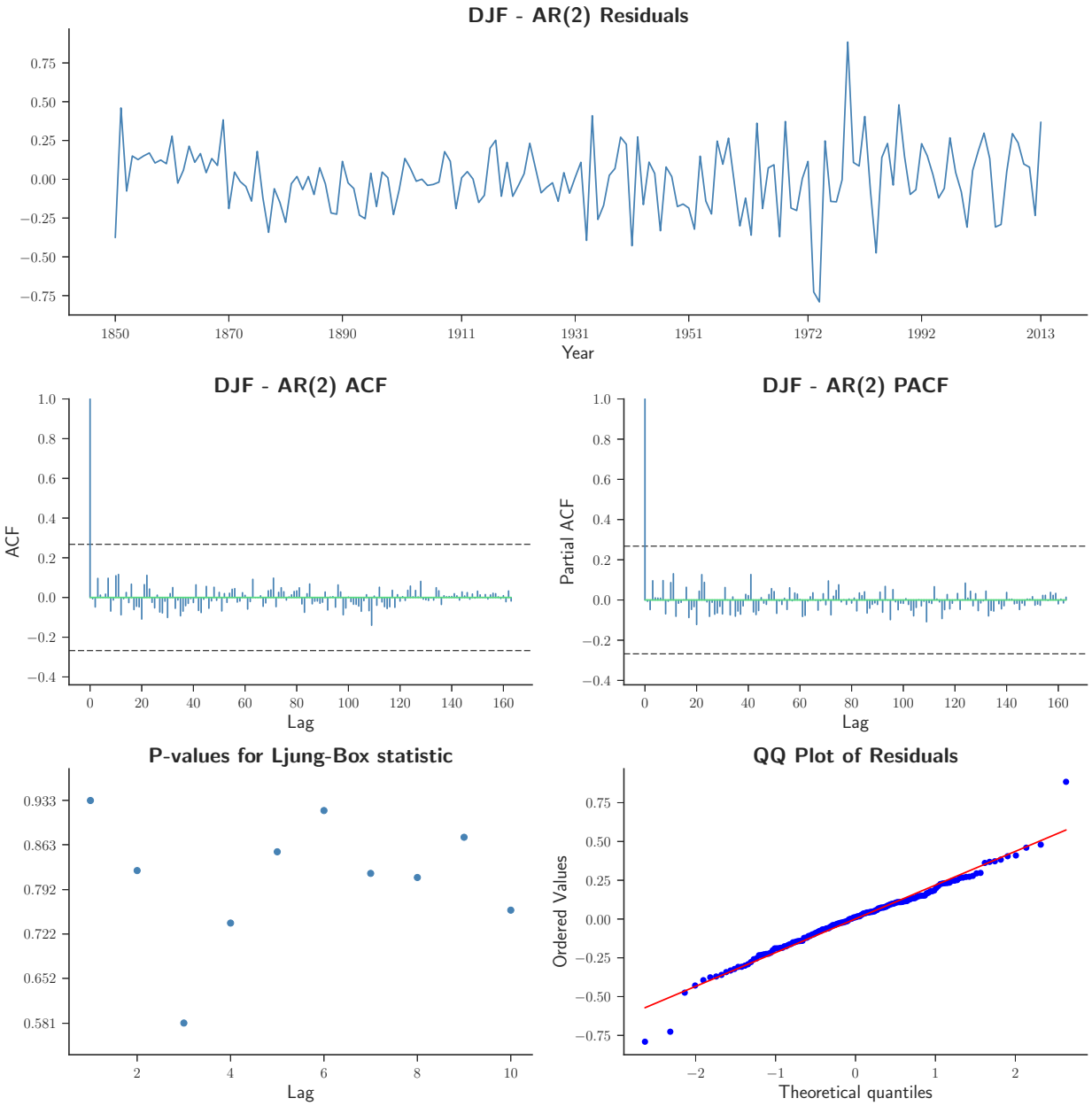

Figure 7: DJF residual diagnostics for AR(2) model.

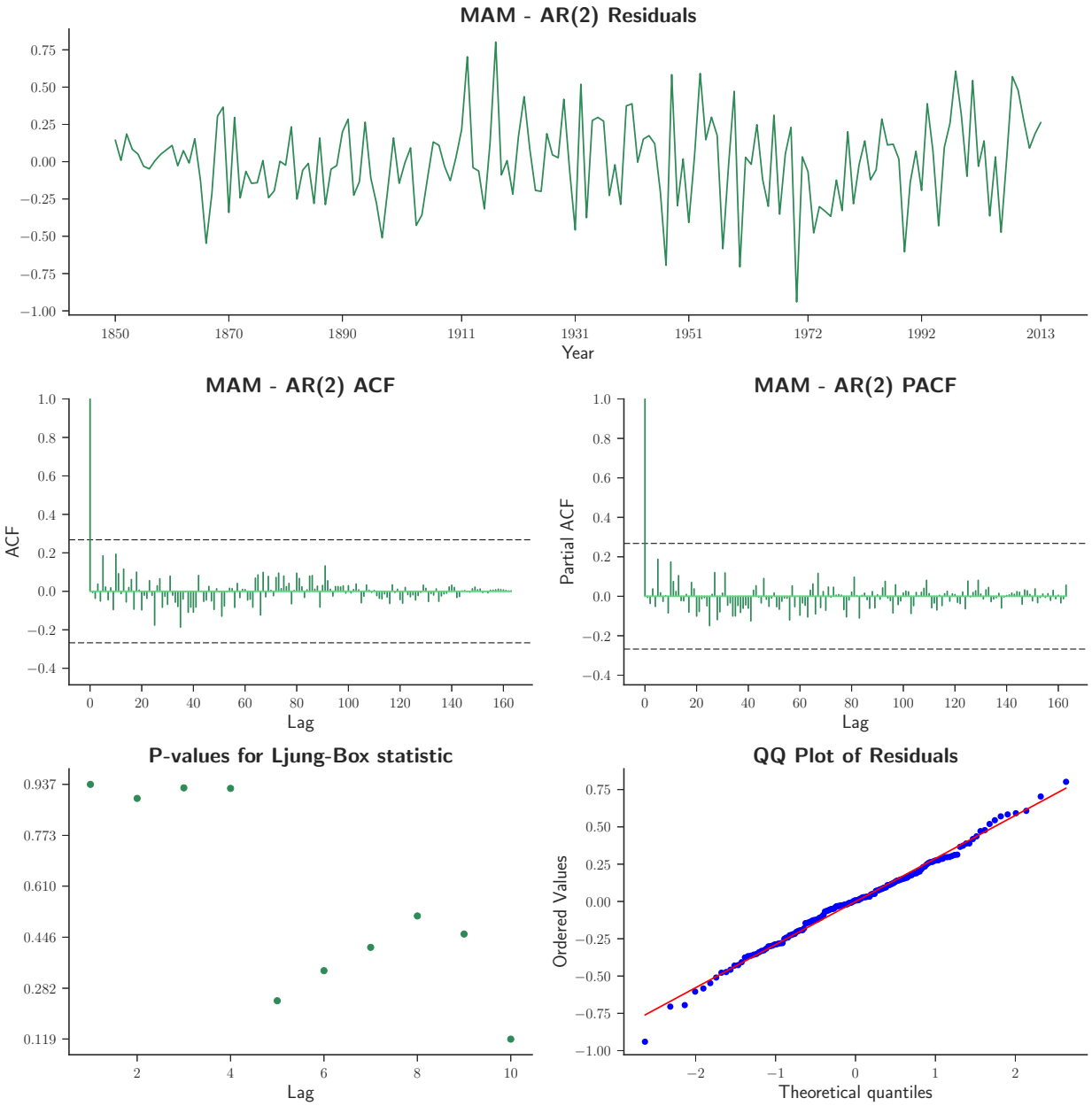

Figure 8: MAM residual diagnostics for AR(2) model.

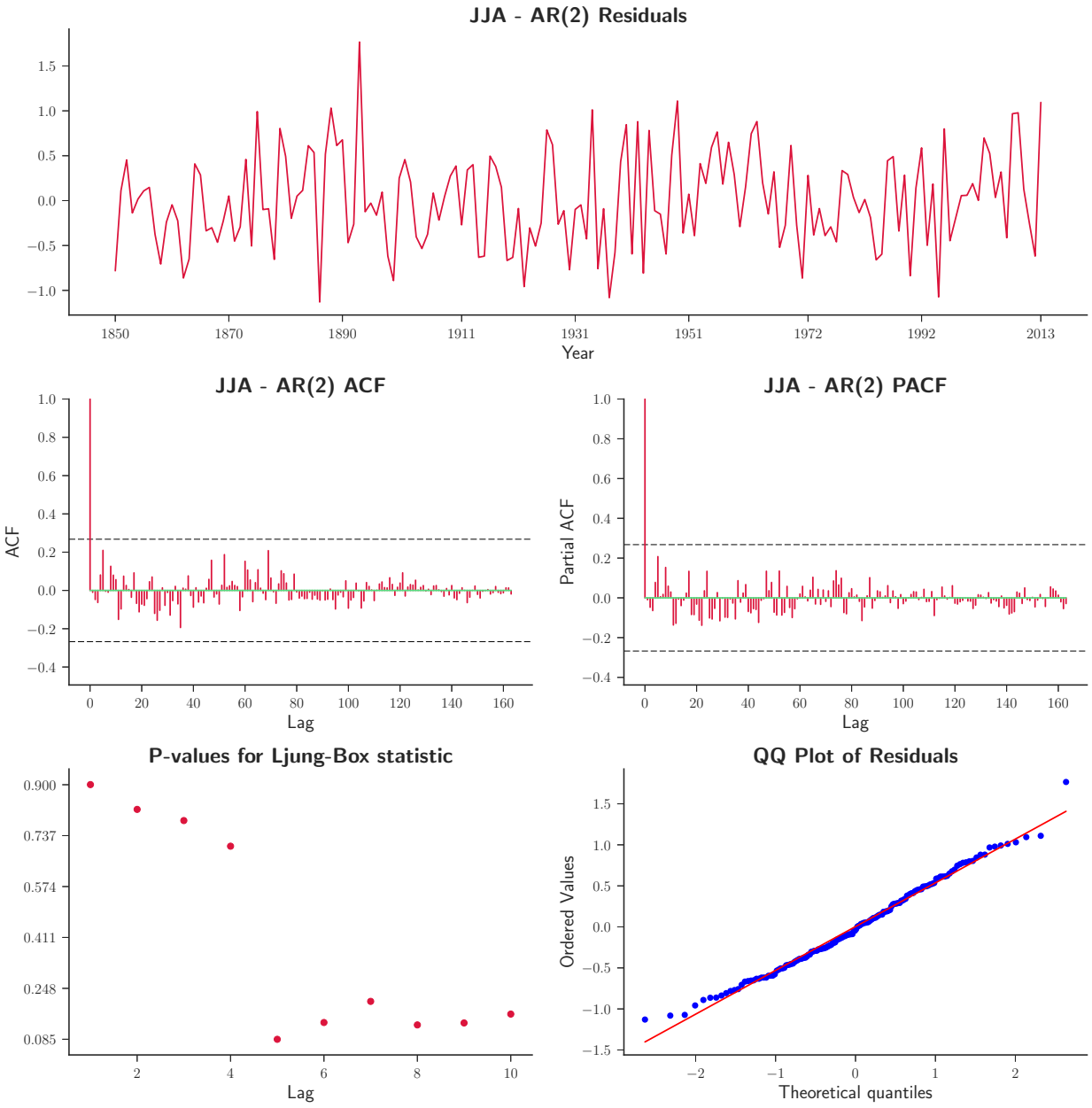

Figure 9: JJA residual diagnostics for AR(2) model.

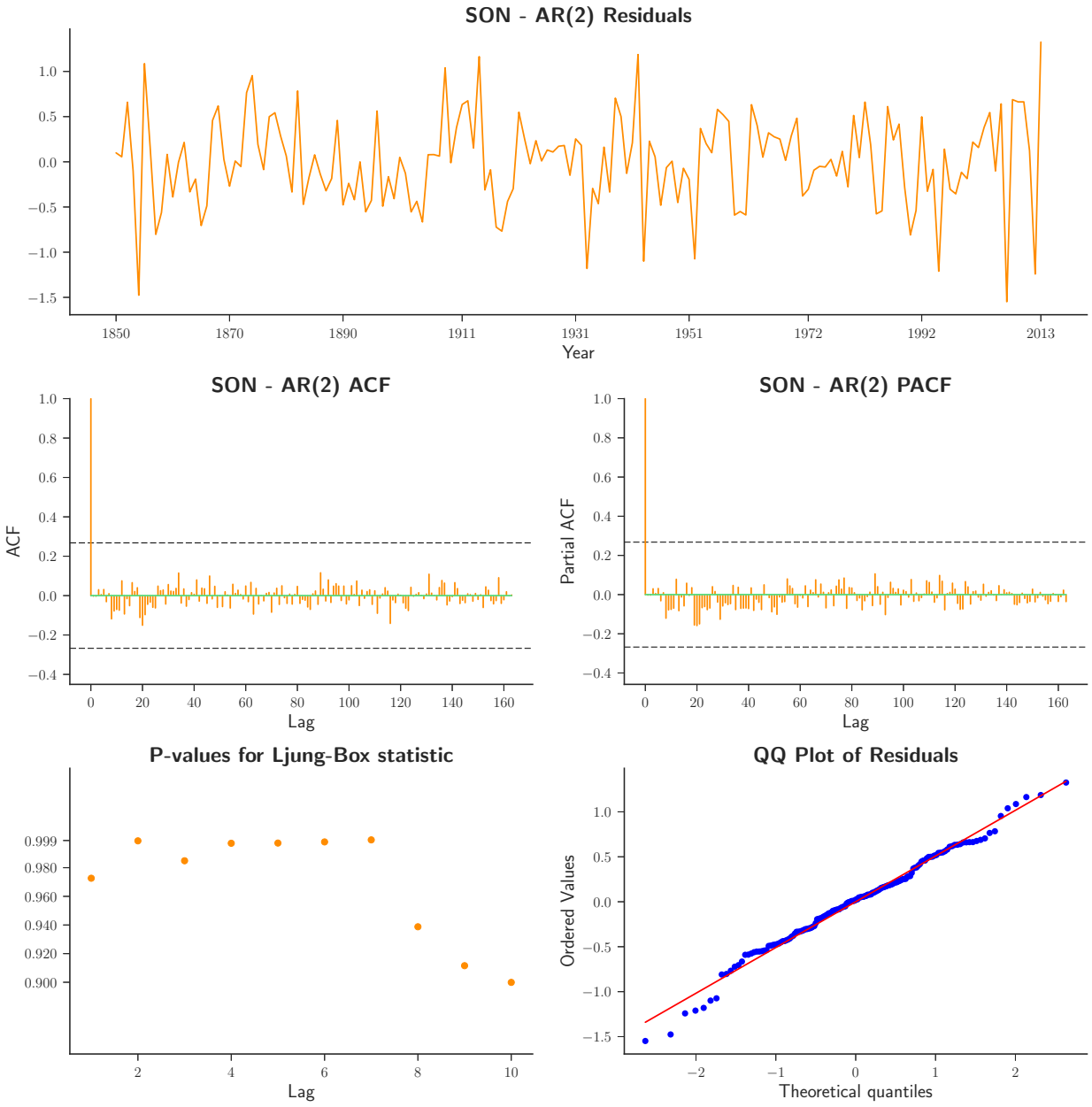

Figure 10: SON residual diagnostics for AR(2) model.

The autocorrelation and partial autocorrelation plots for DJF, MAM, JJA, and SON, shows that for all lags, all sample autocorrelations fall within the 95% confidence bounds indicating the residuals appear to be random. Furthermore, we applied the Ljung-Box test to the residuals from the AR(2) model fit for all the seasons to determine whether residuals

are random. At a significance level of 5%, the resulting Ljung-Box test suggest that there is strong evidence that the first 10 lag autocorrelations among the residuals do not exhibit serial correlation. The results indicates that the residuals resemble white noise and that the model provides an adequate fit to the data. However, the QQ plot of our residuals demonstrate that our residuals may not be normally distributed since the plots exhibit "heavy" tails with some slight skewness.

The results of our models capture the majority of the structure for the Arctic sea ice concentration data set for each season. With current research in global warming and its link to the Arctic sea (Pithan and Mauritsen, 2014), we propose a model that shows promising results for the possibility of forecasting.

5 Conclusion

In summary, we have demonstrated the use of non-parametric change point detection to detect interesting climate phenomenons that lead to significant changes in climate trend. The most noteable change point in 1997, where the greatest decrease in the trend occurs, can be furthered studied for any correlations between Arctic sea ice concentration and the 1997-1998 El Niño Southern Oscillation (ENSO) (Cobb et al., 2003). In addition to change point detection, linear spline regression was used to capture the linear trend in the Arctic sea ice concentration. This resulted in detrending our series to investigate the periodicity. We have found that the retreat of the Arctic sea ice can be modeling via AR(2) process.

We have studied the temporal evolution of sea ice concentration, however there leaves room for future work in investigating spatial patterns for the time series. Analyzing the

series for geo-location regions and studying any correlations are a future endeavor. Finally, including the ENSO data set into our work to determine how the two correlate with each other in space and time can potentially be an interesting research topic.

References

Burrus, C. S., Barreto, J. A., and Selesnick, I. W. (1994). Iterative reweighted least-squares design of fir filters. *IEEE Transactions on Signal Processing*, 42(11):2926–2936.

Cobb, K. M., Charles, C. D., Cheng, H., and Edwards, R. L. (2003). El nino/southern oscillation and tropical pacific climate during the last millennium. *Nature*, 424(6946):271.

Hayes, M. H. (1996). *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

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

Pithan, F. and Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geosci*, 7(3):181–184.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Shumway, R. H. and Stoffer, D. S. (2005). *Time Series Analysis and Its Applications (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

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.

Zhang, X. and Walsh, J. E. (2006). Toward a seasonally ice-covered arctic ocean: Scenarios from the ipcc ar4 model simulations. *Journal of Climate*, 19(9):1730–1747.