

SSA analysis and forecasting of records for Earth temperature and ice extents

V. Kornikov, A. Pepelyshev, A. Zhigljavsky

December 19, 2016

St.Petersburg State University, Cardiff University

Abstract

In this paper, we continued the research started in [6, 7]. We applied the so-called Singular Spectrum Analysis (SSA) to forecast the Earth temperature records, to examine cross-correlations between these records, the Arctic and Antarctic sea ice extents and the Oceanic Nino Index (ONI). We have concluded that that the pattern observed in the last 15 years for the Earth temperatures is not going to change much, found very high cross-correlations between a lagged ONI index and some Earth temperature series and noticed several significant cross-correlations between the ONI index and the sea ice extent anomalies; these cross-correlations do not seem to be well-known to the specialists on Earth climate.

AMS [Primary]62M20

keyword: Singular Spectrum Analysis, stability of forecasts, long-horizon forecasting, retrospective forecasts, SSA vector forecasting

1 Introduction

This paper has the following purposes.

- (1) Further investigation of stability of SSA forecasts for the series with no apparent structure.
- (2) Learning about the behaviour of the major series representing the Earth temperatures during the last 35 years.
- (3) Evaluation of the quality of the SSA forecasts of the Earth temperature series that we have made five years ago.

- (4) Forecasting these temperature series for the next 5 years and assessing stability of the forecasts.
- (5) Investigating the Oceanic Nino Index (ONI) and studying inter-dependence between the ONI series and temperature records.
- (6) Application of SSA for the analysis and forecast of the series representing the Arctic and Antarctic sea ice extents and studying inter-dependence between these series and the series representing the temperatures on the North and South poles respectively.

The research on SSA analysis of Earth temperature records has started in our previous paper [6]. In the present paper, we assess the results published in [6] and advance the research further by analysing many more real-life series. By doing this, we get more confidence in the performance of SSA for the analysis and forecasting of series with no or little structure. Methodological and theoretical principles of the basic version of SSA, that we use in this research project, have been carefully considered in [3, 5]. In these books, however, the performance of SSA has been verified only on a series (real-life and simulated) with clear structure. For most of the series we analyse in this paper, there is no apparent structure and in this case many methods of time series analysis may not show any sense or stability. As we see in numerous graphs below, SSA performance is very stable and therefore SSA can be considered as a reliable tool for analyzing these kind of series.

In Section 2, we briefly consider SSA and our methodological principles. In Section 3, we consider the temperature records representing the average temperature on Earth and its parts during the last few decades. First, we evaluate the quality of the SSA forecasts of Earth temperature series that we have made 5 years ago and published in [6]. Then we make a large number of SSA forecasts of these temperatures for the following 5 years. To achieve the robustness of forecasts, we choose a wide range of parameters of SSA.

In Section 4, we correlate the main temperature series of Section 3 with ONI, the Oceanic Niño Index. We show that some of the temperature series are very highly correlated to a lagged ONI series. As is widely known (see, for example, [8, Chapter 6]) the origins, timings (except the observation that both El Niño and La Niña are usually strongest during December – April) and intensities of the El Niño and La Niña are still big mysteries for modern science so that the ONI series should be currently considered as unpredictable. This series, however, has the major effect on the dynamics of many of the Earth’s temperature series. An implication of the results of

Section 4 confirms one of the learnings from Section 3 saying that most of the Earth’s temperature series are lacking structure.

In Section 5, we use SSA for the analysis and forecast of the series representing the Arctic and Antarctic sea ice extents and correlate these series to the series representing the temperatures on the North and South poles respectively. The series showing Arctic and Antarctic sea ice extents have much structure related to their seasonality. This seasonality is very easy to extract with the help of SSA. After extraction of the main seasonality, in Section 5 we try to study the inter-dependence between the residuals from the ice extent data and temperature records, as well as the ONI series.

2 The methodology

2.1 Singular Spectrum Analysis

Singular Spectrum Analysis (SSA) is a well-known tool of analyzing and forecasting time series in climatology in general and the analysis of temperature records in particular, see e.g. [2, 4, 5, 9, 10]. It is a nonparametric method that does not require standard assumptions about the time series and analyses the data in the phase space. The non-parametric nature of SSA is a very attractive property for the analysis of the Earth temperature records as it seems that these records do not follow any simple pattern or model.

There are several versions of SSA. We shall use the most common one which is often called Basic SSA. A short description of this algorithm is given below, see [3] and [5] for details.

2.1.1 Short description of Basic SSA

Let x_1, \dots, x_T be a time series of length T . Given a window length L ($1 < L < T$), we construct the L -lagged vectors $X_i = (x_i, \dots, x_{i+L-1})^T$, $i = 1, 2, \dots, K = T - L + 1$, and compose these vectors into the matrix

$$\mathbf{X} = (x_{i+j-1})_{i,j=1}^{L,K} = [X_1 : \dots : X_K].$$

This matrix has size $L \times K$ and is often called ‘trajectory matrix’. It is a Hankel matrix, which means that all the elements along the diagonal $i+j=\text{const}$ are equal.

The columns X_j of \mathbf{X} can be considered as vectors belonging to the L -dimensional space \mathbb{R}^L . The singular-value decomposition (SVD) of the matrix $\mathbf{X}\mathbf{X}^T$ yields a collection of L eigenvalues and eigenvectors. For a

given integer r , $1 \leq r < L$, we choose the r largest eigenvalues and corresponding eigenvectors of $\mathbf{X}\mathbf{X}^T$. The chosen eigenvectors determine an r -dimensional subspace in \mathbb{R}^L ; call this subspace S_r . The L -dimensional data $\{X_1, \dots, X_K\}$ is then projected onto this r -dimensional subspace S_r and the subsequent averaging over the diagonals gives us some Hankel matrix $\tilde{\mathbf{X}}$, which we consider as an SSA approximation to \mathbf{X} .

2.1.2 SSA forecasting

There are several ways of constructing forecasts based on the SSA decomposition of the series described above, see Chapter 2 in [3] and Chapter 3 in [5]. The most obvious way is to use the linear recurrent formula which the last terms of the series reconstructed from $\tilde{\mathbf{X}}$ satisfy. We however prefer to use the so-called ‘SSA vector forecast’ ([3], Sect. 2.3.1). The main idea of this forecasting algorithm is as follows. Selection of r eigenvectors of $\mathbf{X}\mathbf{X}^T$ leads to the creation of the subspace S_r . SVD properties allow us to assert that the L -dimensional vectors $\{X_1, \dots, X_K\}$ lie close to this subspace. Consider the vectors Z_1, \dots, Z_K where Z_i defined as the projection of X_i onto the subspace S_r . The vector forecasting algorithm then sequentially constructs the vectors $\{Z_{K+1}, Z_{K+2}, \dots\}$ so that they stay in the chosen subspace S_r and the hankelization of the matrix $(Z_1, \dots, Z_K, Z_{K+1}, Z_{K+2}, \dots)$ gives the vector forecast.

2.1.3 Choice of SSA parameters

In the examples below, the length of the series is $T = 440$ and the forecasting horizon is $h \in [60, 100]$.

Basic SSA described above has two parameters: L , the window length and r , the dimension of the subspace S_r . Let us discuss the choice of these two parameters in the analysis of the temperature series.

Choice of L . If the structure of the series is assumed stable then large values of L , of the order $L \cong 80$, should be preferred to small values, of the order $L \cong 10$. We, however, do not make the assumption that the structure of the series is stable; on the contrary, we believe that this structure may be rather unstable. In this case, large values of L would make SSA too rigid and lacking flexibility. On the other hand, for very small values of L , SSA may be too sensitive to the noise and small variations in the trend. It is therefore natural to select values of L somewhere in-between. Our choice is $20 \leq L \leq 50$, which we believe is a rather broad range.

Choice of r . The choice of r should depend on what do we intend to

forecast. For example, if we observe some seasonal variations in the data and we want to forecast these variations, then we have to choose r large enough to capture these variations. There are several automatic procedures (see e.g. [5]) helping in choosing the most suitable value of r (roughly speaking, r should be the smallest among those values of r for which the residuals after signal extraction pass the chosen statistical tests for being a noise). These procedures, however, are often not very reliable as they require a few conditions for the time series to satisfy and are not well suited for the long and medium range forecasting.

In what follows, we choose $r = 5, 7$. Whatever the rule for selecting r , some values of r are too small, which implies that some part of the signal will be missed, but other values of r are too large, which implies that a significant part of the noise will be included into the ‘reconstructed signal’. This, however, goes in line with one of the main aims of our study which is checking the stability of the forecasts.

Summarizing, for all temperature series we choose parameters $20 \leq L \leq 50$ and $r \in \{5, 7\}$. This choice is independent of the structure of the series. This is done for consistency and for ensuring the fairness of the analysis. Despite a better forecast (with better stability) can be obtained if we optimize the domains of parameters L and r for each individual series, we have fixed the domains to show the robustness of results. Furthermore, the results of our study are very stable with respect to these domains.

2.2 Stability of forecasts

2.2.1 Forming the samples of forecasts

Assume that we have a family of SSA forecasts which is parameterized by a parameter $\theta \in \Theta = \{\theta_1, \dots, \theta_m\}$, where $\theta = (L, r)$ and m is the total number of chosen pairs of parameters (L, r) . For each time moment $t \leq T$, any $g > 0$ and any $\theta \in \Theta$ we can build a g -step ahead forecast $\hat{x}_{t+g}(\theta)$ based on the information x_1, \dots, x_t . Hence for any $t \leq T$ we may compute the following set of forecasting results:

$$\mathbb{X}_t = \{\hat{x}_{T+g}(\theta) : g \in [h_1, h_2], \theta \in \Theta\},$$

where $0 < h_1 \leq h_2$ are some numbers. The number of elements in the samples \mathbb{X}_t is $M = m(h_2 - h_1 + 1)$.

To summarize, our forecasting procedure gives us $T - T_0 + 1$ samples $\mathbb{X}_t = \{f_1^{(t)}, \dots, f_M^{(t)}\}$ at all $t = T_0, \dots, T$, where T_0 is the first time moment we make the forecasting.

2.2.2 Comparison of the samples

We now need to compare the samples \mathbb{X}_t ($t = T_0, \dots, T$) to evaluate the stability of the corresponding forecasts and decide whether at $t = T$ we have reached an acceptable level of stability.

The mean values of the samples \mathbb{X}_t are $\bar{f}_t = (f_1^{(t)} + \dots + f_M^{(t)})/M$. As the future is unknown, it is not possible to check whether the mean values \bar{f}_t give good approximations to the true values $\mathbf{E}x_{T+h}$. Hence the values \bar{f}_t do not provide much information about the quality of forecasts.

As measures of stability, we must consider the behaviour of some characteristics of variability of the samples \mathbb{X}_t . The most common among these characteristics is the (empirical) standard deviation of \mathbb{X}_t :

$$s_t = \left[\frac{1}{M-1} \sum_{i=1}^M (f_i^{(t)} - \bar{f}_t)^2 \right]^{1/2}.$$

Another important characteristic of the sample variability is the so-called range of the sample \mathbb{X}_t :

$$R_t = \max_{i=1, \dots, M} f_i^{(t)} - \min_{i=1, \dots, M} f_i^{(t)}.$$

3 Analysis of the Earth temperature records

3.1 Characteristics of forecast quality

We consider the Earth temperature records from the web-site <http://vortex.nsstc.uah.edu/> (National Space Science and Technology Center, NASA, USA). The records represent the temperature on Earth and some of its parts since December 1978. We did not present the analysis of longer temperature records as there is some controversy around earlier temperature records, see [8]. The series contains the temperature anomalies rather than the absolute temperatures.

Note that in [6] we have used the the temperature anomalies computed relative to the base period 1951-1980. Currently, the temperature anomalies have been recomputed and are given in <http://vortex.nsstc.uah.edu/> relative to the new annual cycle 1981-2010.

For a given series of monthly data records z_{y+m} (where y refers to the year and m ($1 \leq m \leq 12$) denotes the month) and a reference period of several years, the respected series of anomalies is defined as the difference between the original time series z_{y+m} and the series of monthly averages

\bar{z}_m , where the monthly averages \bar{z}_m ($1 \leq m \leq 12$) are computed for the reference period. If the reference period is not specified (as in Section 5.2) then the averages \bar{z}_m are computed for the whole period of observations.

Working with anomalies rather than with absolute temperature records is customary in climatology, see for example publications and web-sites of the Goddard Institute for Space Studies. In the present subsection we follow [6] and use the data from December 1978 to July 2015 so that altogether we have $T = 440$ data points. The first time moment we start the forecasts is January 2005 implying $T_0 = 314$. We forecast the series until 2020 (longer-term forecasts are very similar) by setting $h_1 = 97, h_2 = 99$. As mentioned above we use $L \in [20, 50]$ for the SSA window length L and choose the first $r \in \{5, 7\}$ eigenvectors.

Figure 1: Earth temperature. The time series (gray), the SSA approximation and the forecast for $L = 50$ and $r = 7$ (black).

Figure 2: Earth temperature. Top: averages \bar{f}_t (black), standard deviations s_t (light grey), ranges R_t (dark grey). Bottom: box-plots of the samples \mathbb{X}_t .

Figure 3: Forecasts for the Earth temperature at Jan 2018; $L = 20, 30, 40, 50$ and $r = 5$ (top) and $r = 7$ (bottom).

To illustrate our analysis, consider the series of the global temperature on Earth and Northern Hemisphere temperature. These two series of temperatures are discussed most often. We have done similar analysis for some other series; the results are presented at the web-site <http://earth-temperature.com/forecasting/>. For each of the three chosen temperature series we plot the following.

- (i) Figures 1 and 4: the series itself, the SSA approximation and SSA forecast for $L = 50$ and $r = 7$ computed at the last point $t = T$ (December 2009).

- (ii) Figures 2 and 5 (left): the series $-\bar{f}_t$, standard deviations s_t (light grey), ranges R_t (dark grey) for $t = 314, \dots, 430$ (the averages \bar{f}_t are always plotted with the minus sign for the purpose of clarity of display).
- (iii) Figures 2 and 5 (right): box-plots of the samples \mathbb{X}_t for $t = 325, 337, 349, 361, 373, \dots, 421, 433$.
- (iv) Figures 3 and 6: forecasts for the temperature at January 2018 using the series x_1, \dots, x_t for $L = 20, 30, 40, 50$, $r = 5, 7$ and all $t = 314, \dots, 433$.

Note that the markers on the x-axis in all plots correspond to Januaries. To compare the forecasted values of the temperatures with recent values, note the average values of these temperatures for 2000–2009: 0.222 for Earth; 0.312 for Northern Hemisphere.

Figures 2, 3, 5 and 6 are extensions of the corresponding figures we have provided in [6] to cover the time periods up to 2015.

Figure 4: Northern Hemisphere temperature. The time series (gray), the SSA approximation and the forecast for $L = 50$ and $r = 7$ (black).

Figure 5: Northern Hemisphere temperature: Top: averages $-\bar{f}_t$ (black), standard deviations s_t (light grey), ranges R_t (dark grey). Bottom: box-plots of the samples \mathbb{X}_t .

Figure 6: Forecasts for the Northern Hemisphere temperature at Jan 2018; $L = 20, 30, 40, 50$ and $r = 5$ (top) and $r = 7$ (bottom).

3.2 Assessing accuracy of forecasts made in 2009

In Figure 7 we can see that during the period Jan 2010 to Sep 2014 the global Earth temperatures roughly followed a typical forecast given in [6].

Figure 8 shows that for the global Earth temperatures, all forecasts made in [6] for the period from Jan 2010 to Dec 2020 were very similar and the

Figure 7: Global Earth temperature. Gray: the original time series from <http://vortex.nsstc.uah.edu/> for the period Dec 1978 to Sept 2014. Black: the SSA approximation until Dec 2009 and the forecast from Jan 2010 onwards. SSA parameters: $L = 50$ and $r = 7$.

actual global Earth temperature during the whole period from Jan 2010 to Sept 2014 was well inside the family of forecasts. Figure 8 also shows that the forecasts made in Dec 2009 do not show any significant change in the level of temperatures.

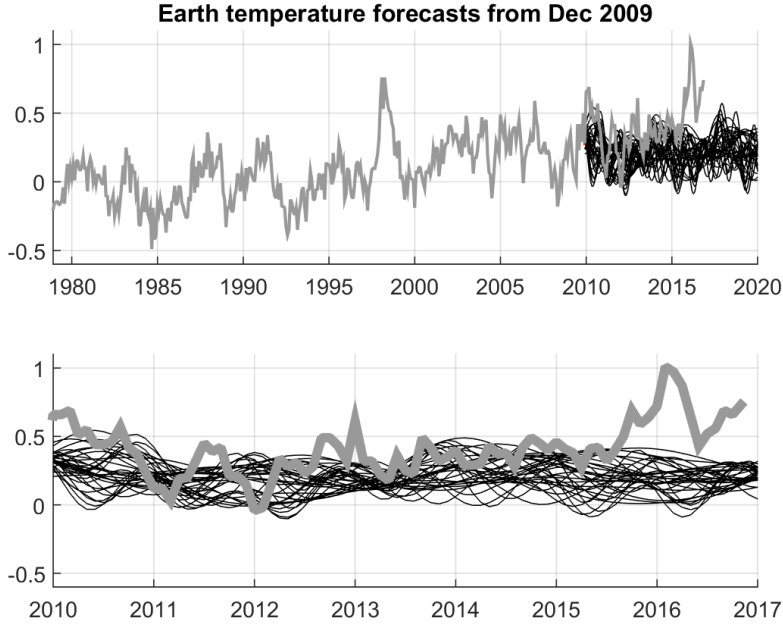


Figure 8: Global Earth temperature. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

Figure 9 is similar to Figure 8 but shows the North pole temperatures (rather than the global Earth temperatures). Again, all forecasts made in 2009 for the period from Jan 2010 to Dec 2020 were very similar and the

actual global Earth temperature during the whole period from Jan 2010 to Sept 2014 was well inside the family of forecasts. Unlike Figure 8, Figure 9 shows that the forecasts made in Dec 2009 indicate a very small increase in the level of temperatures in the North pole. This figure should be compared with Figure 16 where revised forecasts are shown. Note also that despite the forecasts were unable to catch the volatility of the actual series, these forecasts have shown the main trend (more precisely, the absence of any trend) quite accurately.

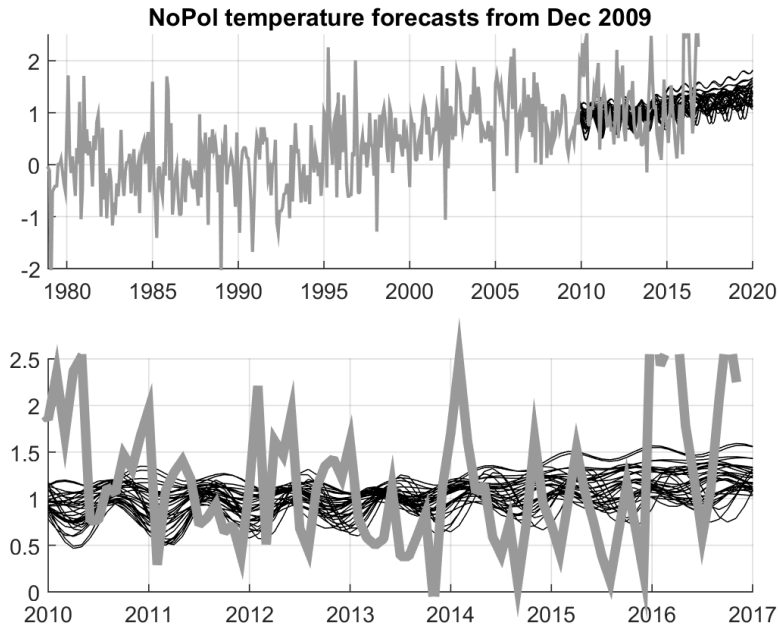


Figure 9: North Pole temperatures. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

Figure 10 is similar to Figures 8 and 9. It shows the South pole temperatures. Unlike the global Earth and North pole temperatures, the South pole temperatures never exhibited any trend and were very volatile. The volatility did not exhibit any patterns and this was reflected in the SSA forecasts: all these forecasts were very close to a constant. Figure 10 shows that the South pole temperatures continue to be volatile with no apparent

tendency.

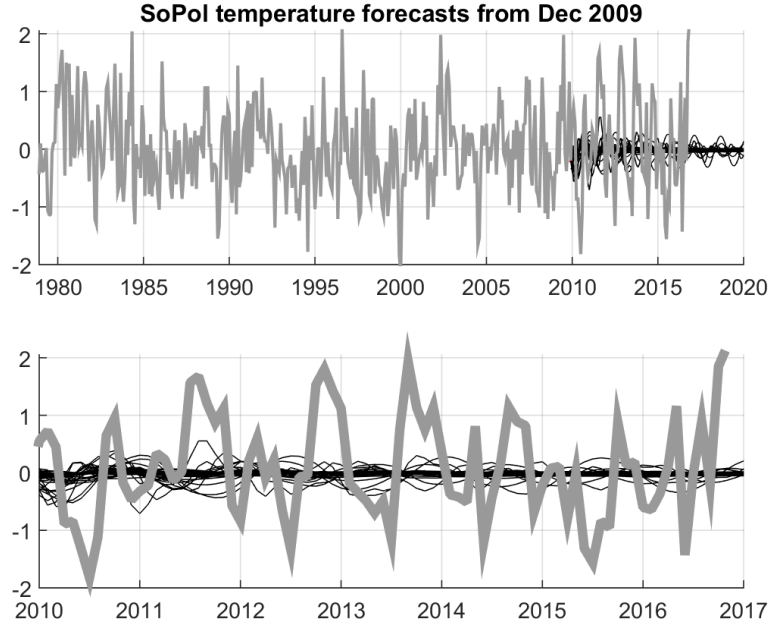


Figure 10: South Pole temperatures. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

Figures 11, 12, 13 and 14 are similar to Figures 8, 9 and 10 and deal with Northern hemisphere, Southern hemisphere, Earth land and Earth ocean temperatures respectively. These pictures are self-explanatory. They show the good quality of the SSA forecasts made in 2009 and the fact that the temperatures continued to be volatile but did not rise in the last three years.

3.3 Forecasting from the present time until Jan 2025

In the figures that follow we show the families of the SSA forecasts of temperatures of different parts of Earth from Aug 2015 until Jan 2020. To make the forecasts we use SSA with the same parameters as above; that is, $L = 20, \dots, 50$ and $r = 5, 7$. The data used is the corresponding temperature series from Dec 1978 until July 2015; thus, the length of all series is

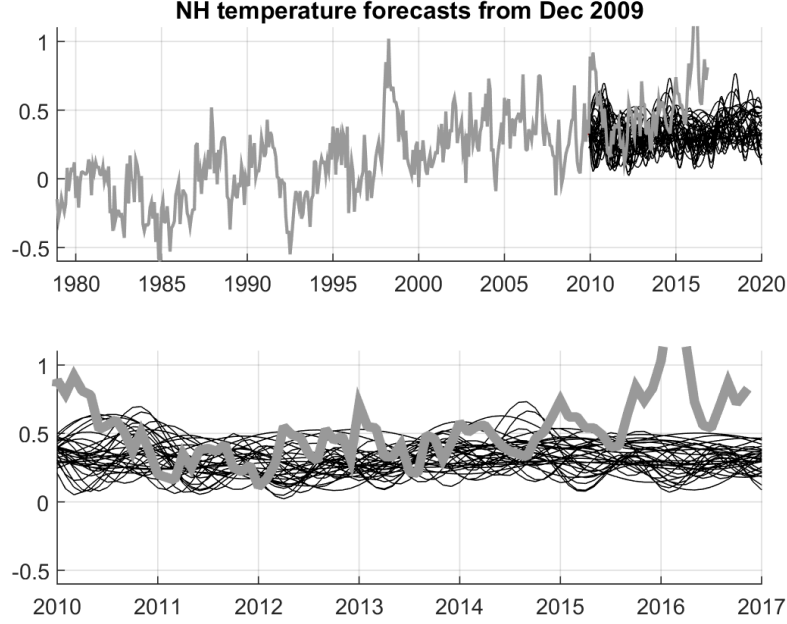


Figure 11: Northern hemisphere temperatures. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

440.

Figures 15, 16, 17, 18, 19, 20 and 21 show forecasts (constructed using the most recent data) of the temperatures for the Earth overall, North pole, South pole, Northern hemisphere, Southern hemisphere, Earth land and Earth ocean temperatures, respectively. The new forecasts have been made using the data until Sept 2014 (that is, the most recent data available at the time of submission of the present paper). These figures complement Figures 8, 9, 10, 11, 12, 13 and 14, respectively. All forecasts do not show any tendencies for temperature increase. There show, however, a lot of volatility and even a possibility of insignificant decrease of some temperatures.

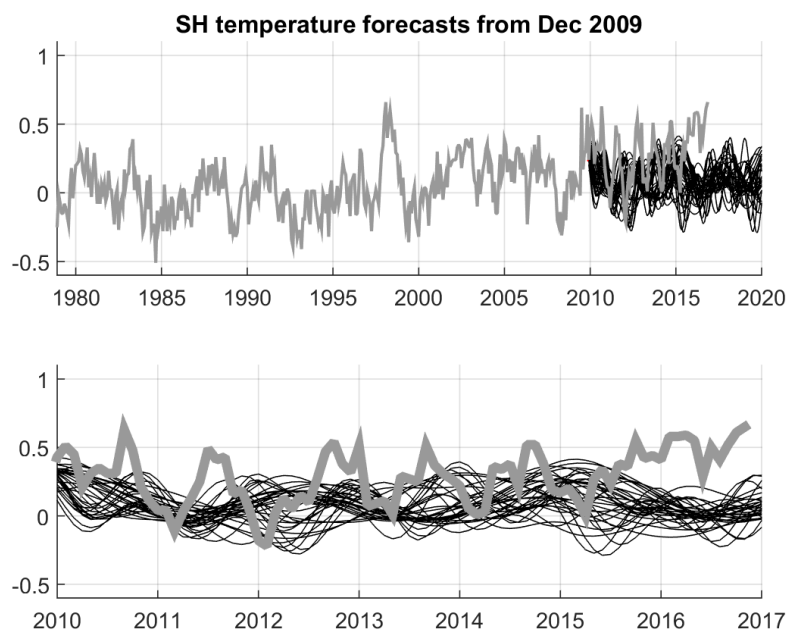


Figure 12: Southern hemisphere temperatures. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

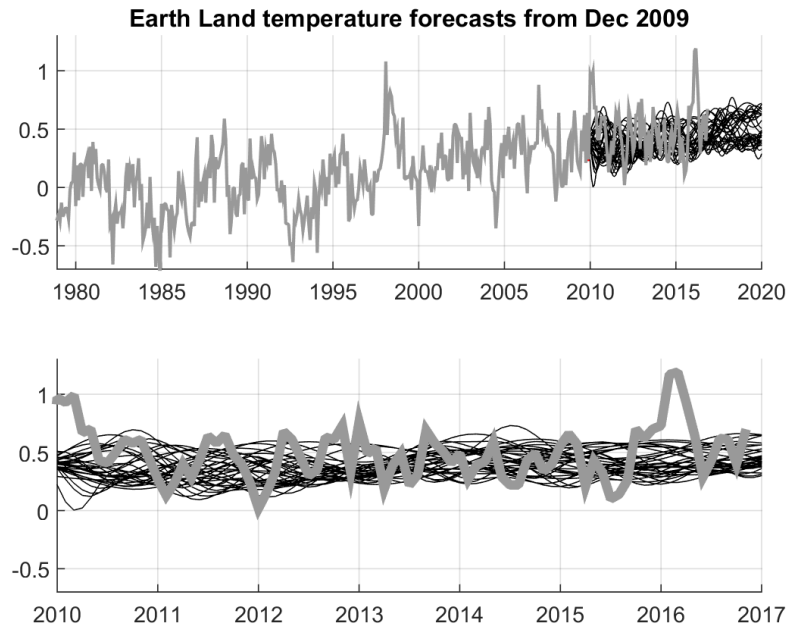


Figure 13: Earth land temperatures. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

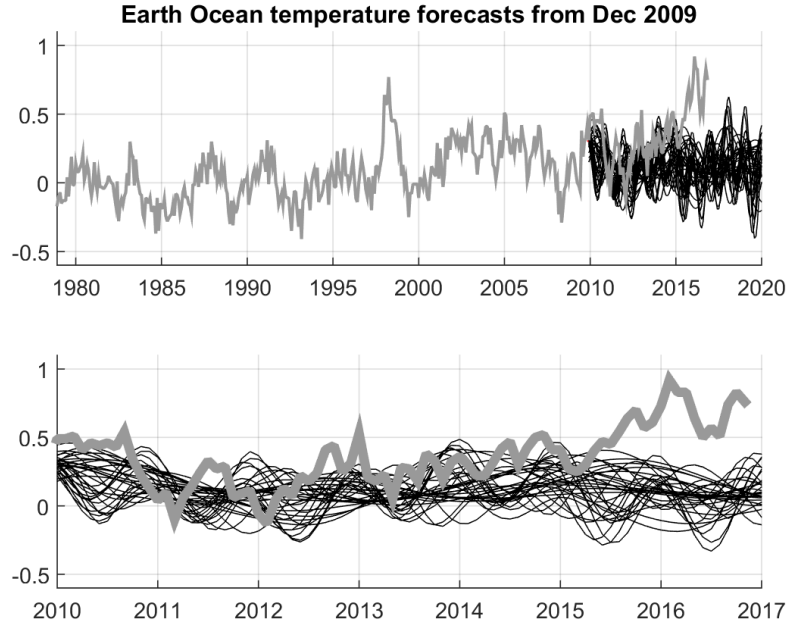


Figure 14: Earth ocean temperatures. Top: the original time series for the period Dec 1978 to Nov 2016 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \dots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2016.

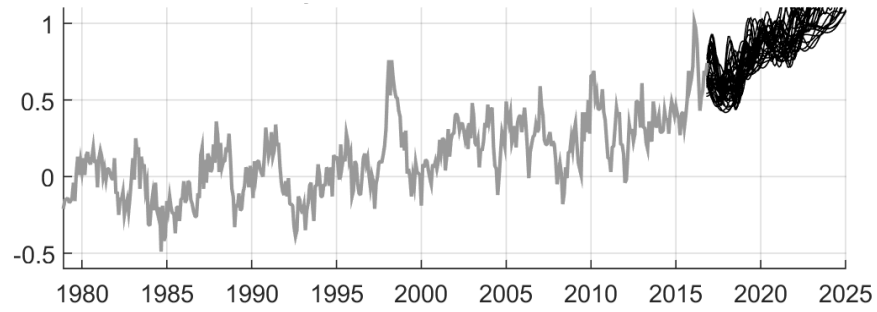


Figure 15: Earth temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016.

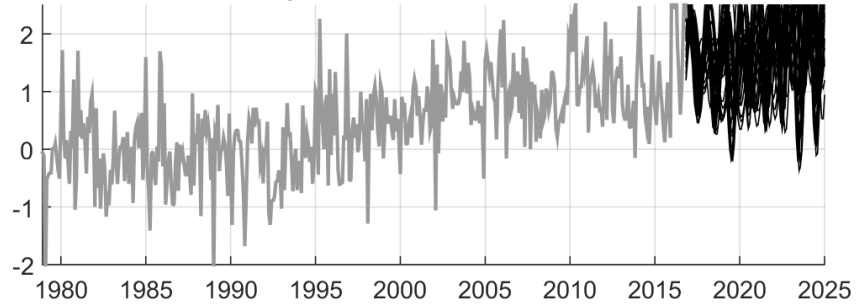


Figure 16: North pole temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016 for $L = 20, \dots, 50$ and $r = 5, 7$.

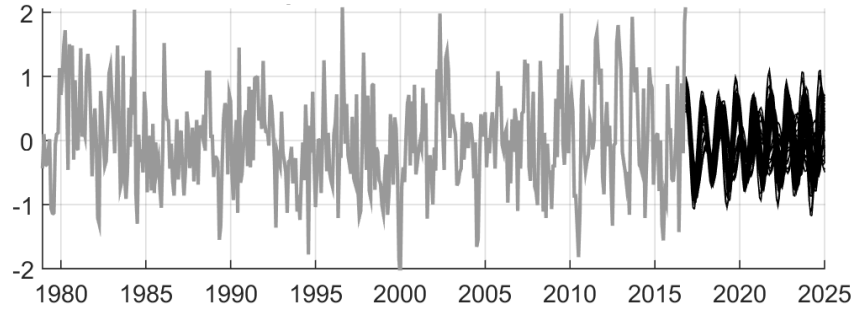


Figure 17: South pole temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016 for $L = 20, \dots, 50$ and $r = 5, 7$.

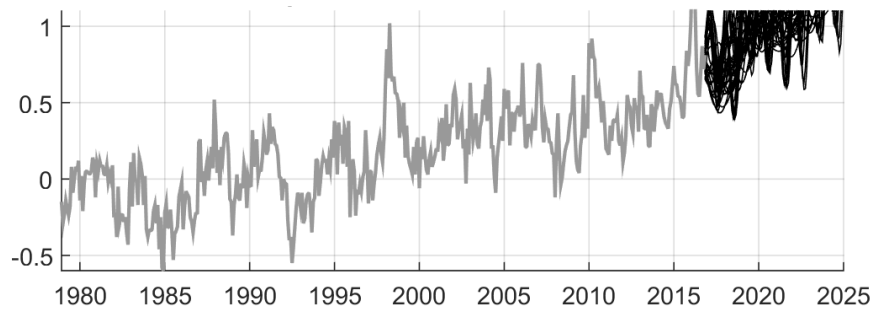


Figure 18: Northern hemisphere temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016 for $L = 20, \dots, 50$ and $r = 5, 7$.

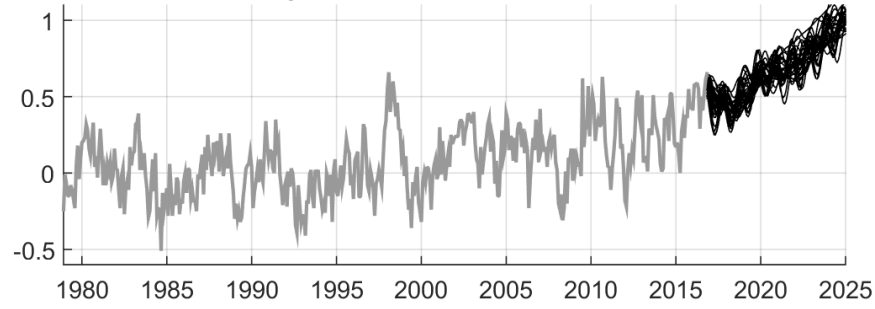


Figure 19: Southern hemisphere temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016 for $L = 20, \dots, 50$ and $r = 5, 7$.

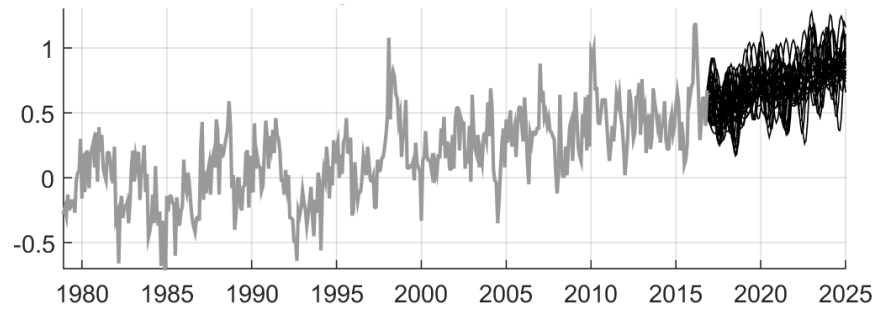


Figure 20: Earth land temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016 for $L = 20, \dots, 50$ and $r = 5, 7$.

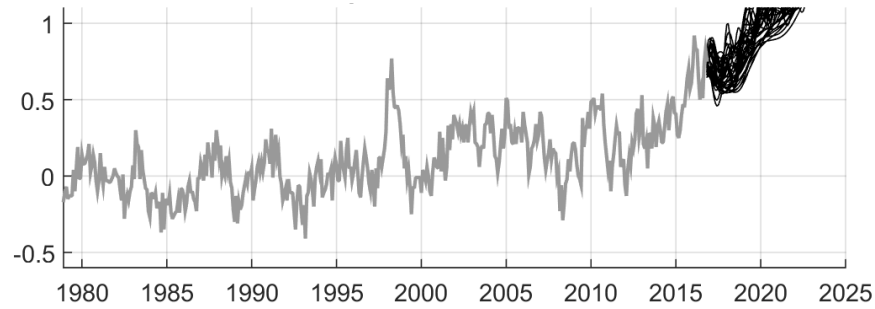


Figure 21: Earth ocean temperature. The time series (gray) until Nov 2016, the family of forecasts (black) from Dec 2016 for $L = 20, \dots, 50$ and $r = 5, 7$.

4 ONI index

The National Oceanic and Atmospheric Administration (NOAA) is a federal agency focused on the condition of the oceans and the atmosphere. De-facto, the standard tool used by NOAA for identifying El Nino (warm) and La Nina (cool) events in the tropical Pacific is the so-called Oceanic Nino Index (ONI) discussed in this section. It is the running 3-month mean SST anomaly for the Nino 3.4 region (i.e., 5°N - 5°S , 120° - 170°W).

Monthly data of the ONI index is taken from the website <http://ggweather.com/enso/oni.htm>. In Figure 22 we plot the ONI index.

Figure 22: The ONI index.

In Figure 23 we plot Earth temperature and the scaled ONI index. We can see an amazing similarity between these two time series. Note that the ONI index has highest correlation with the Earth temperature if apply a delay 3-4 months to the ONI series.

In Figures 24-30 we compare temperatures for different parts of Earth and the ONI index.

By looking at Figure 28 we observe an almost perfect similarity between Tropics temperature and the ONI index. Let $\rho_{\text{ONI},\text{Tr}}(k)$ be the cross-correlation between the ONI index and Tropics temperature for a lag k . The largest cross-correlation between Tropics temperature and the ONI index is $\rho_{\text{ONI},\text{Tr}}(3) = 0.7277$, while the largest cross-correlation between the ONI index and Earth temperature is $\rho_{\text{ONI},\text{Ea}}(4) = 0.3044$. Also note that the largest cross-correlation between the ONI index and smoothed Earth temperature is $\rho_{\text{ONI},\text{smoothedEa}}(4) = 0.3278$.

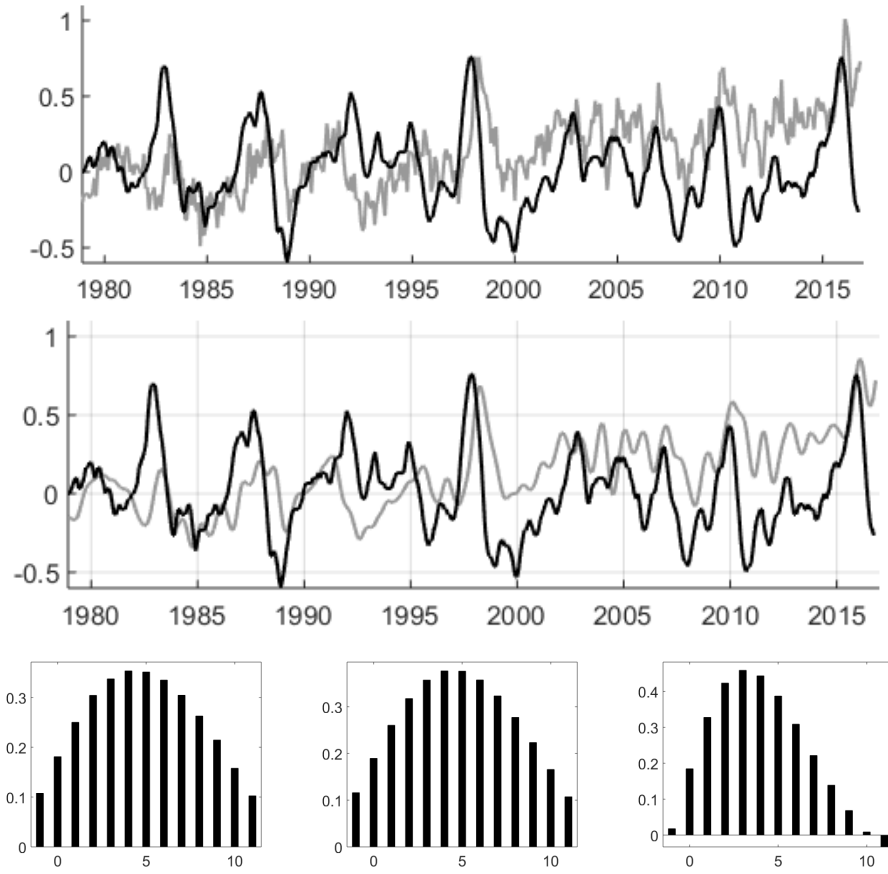


Figure 23: Earth temperature (gray) and the scaled ONI index (black). Middle: Smoothed Earth temperature (gray) and the scaled ONI index (black). Bottom left: Cross-correlation between the ONI index and Earth temperature. Bottom middle: Cross-correlation between the ONI index and smoothed Earth temperature. Bottom right: Cross-correlation between the differences of the ONI index and the differences of smoothed Earth temperature.

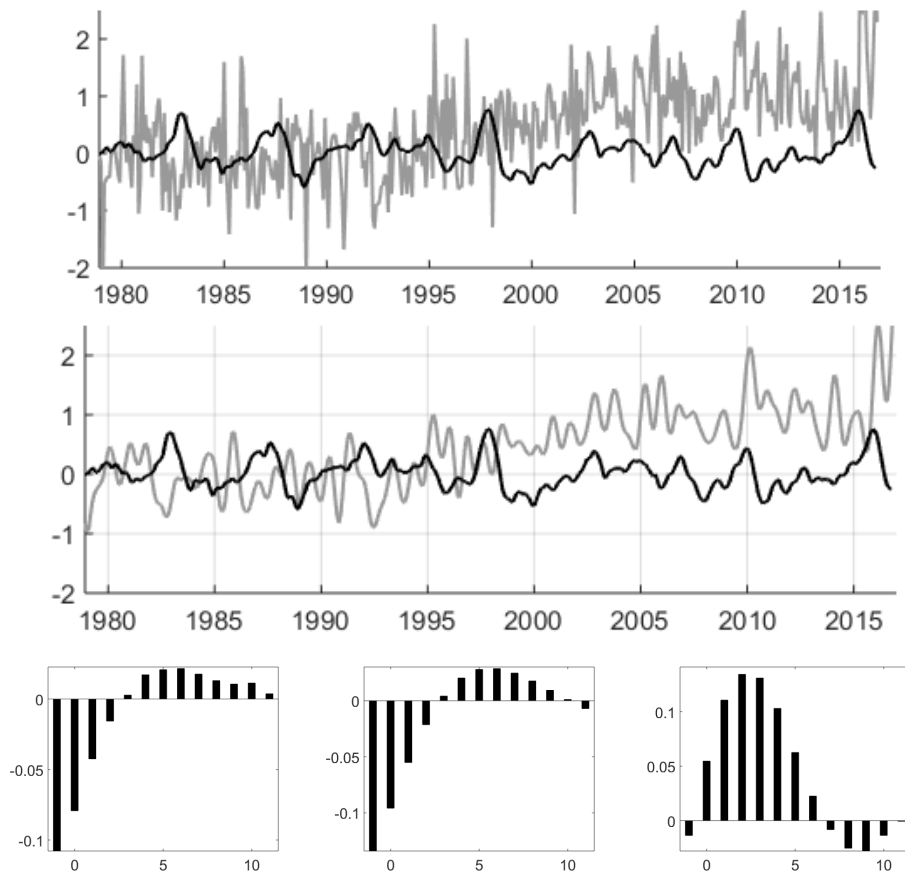


Figure 24: North pole temperature (gray) and the scaled ONI index (black).

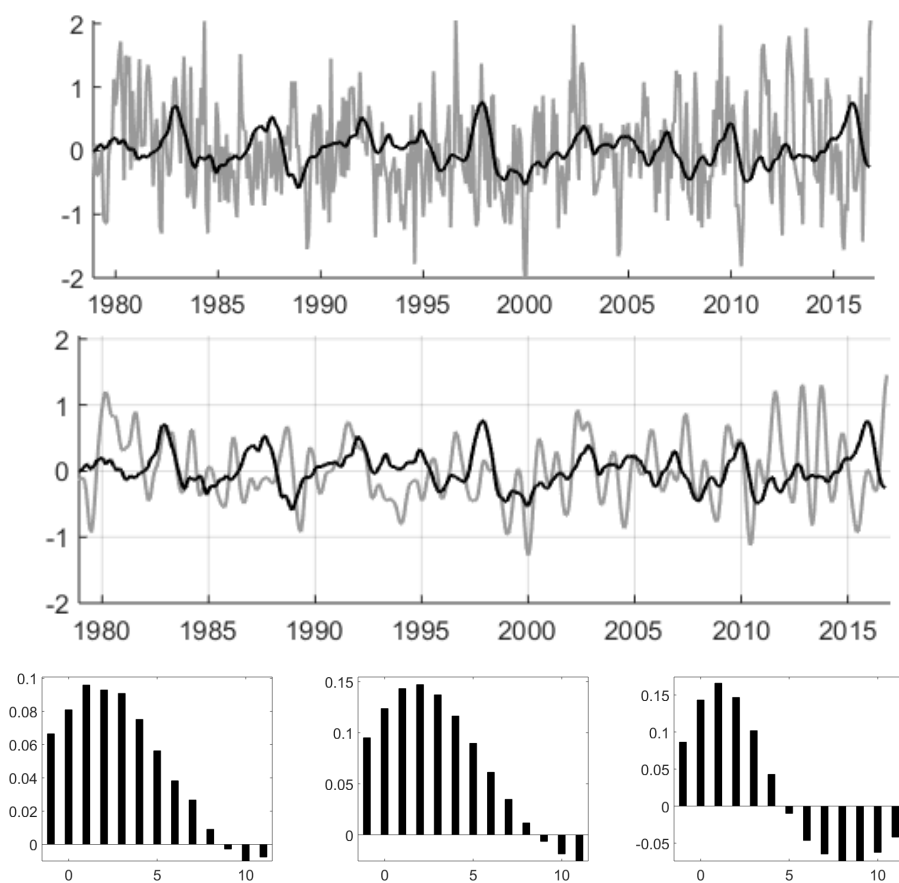


Figure 25: South pole temperature (gray) and the scaled ONI index (black).

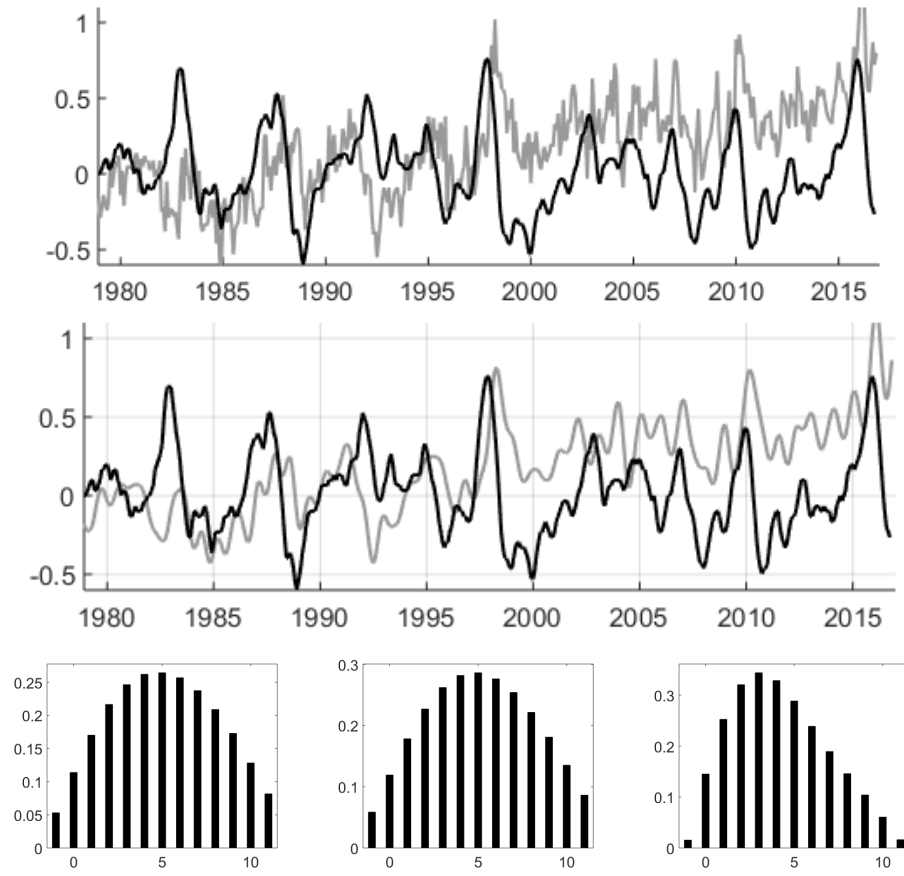


Figure 26: Northern hemisphere temperature (gray) and the scaled ONI index (black).

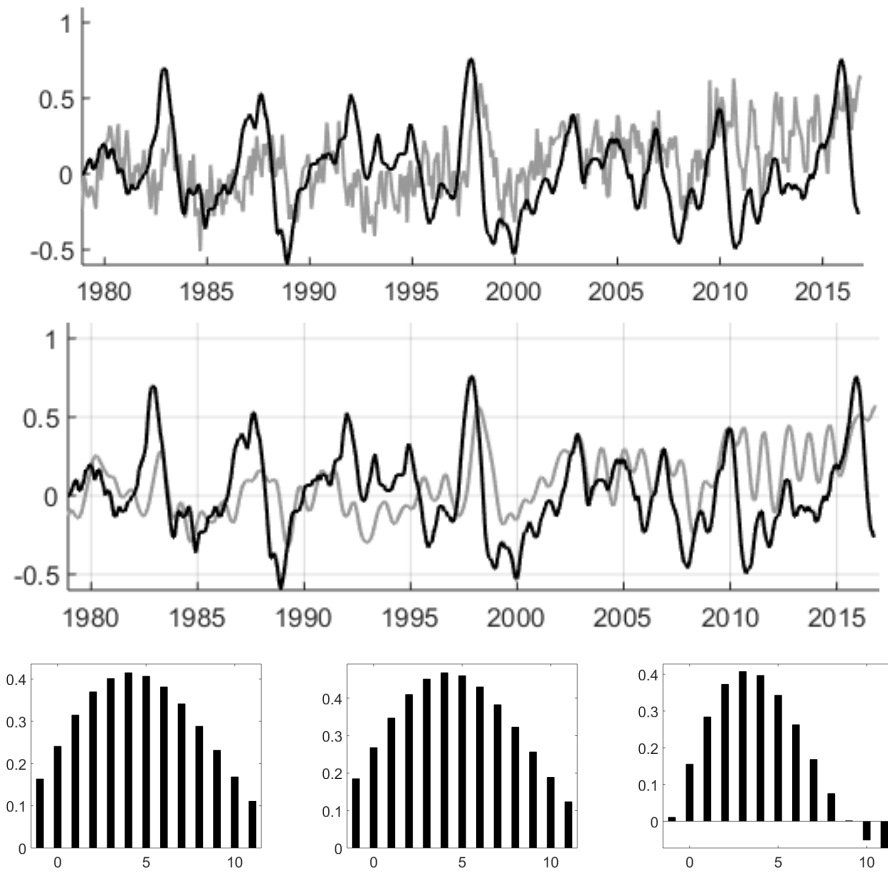


Figure 27: Southern hemisphere temperature (gray) and the scaled ONI index (black).

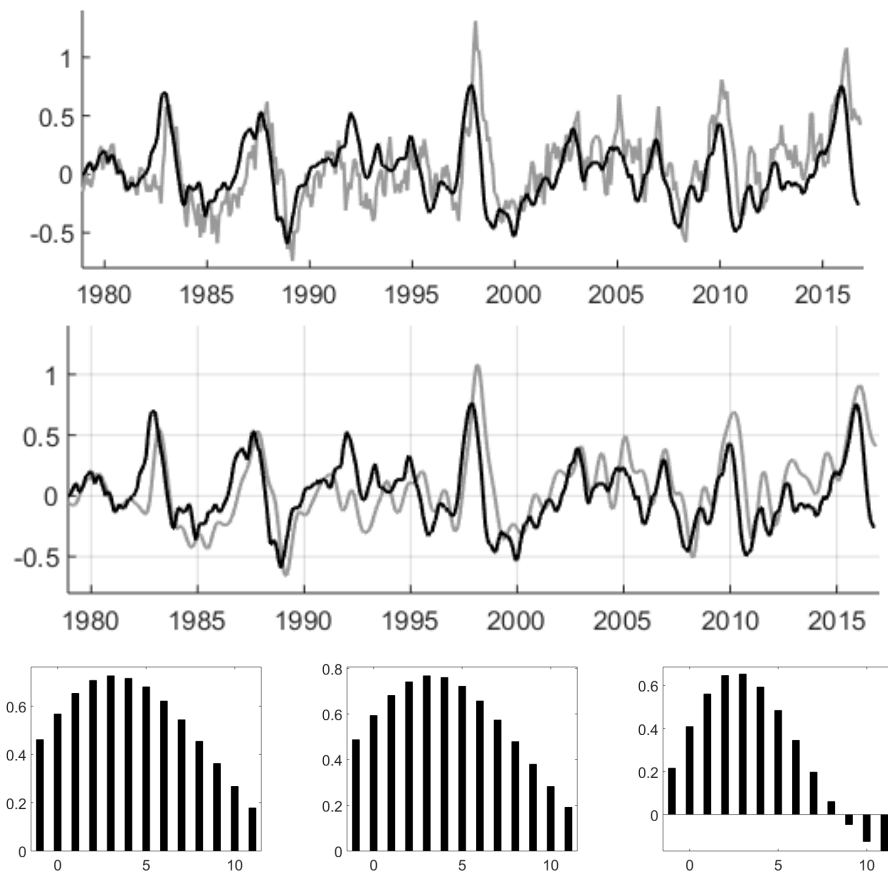


Figure 28: Tropics temperature (gray) and the scaled ONI index (black).

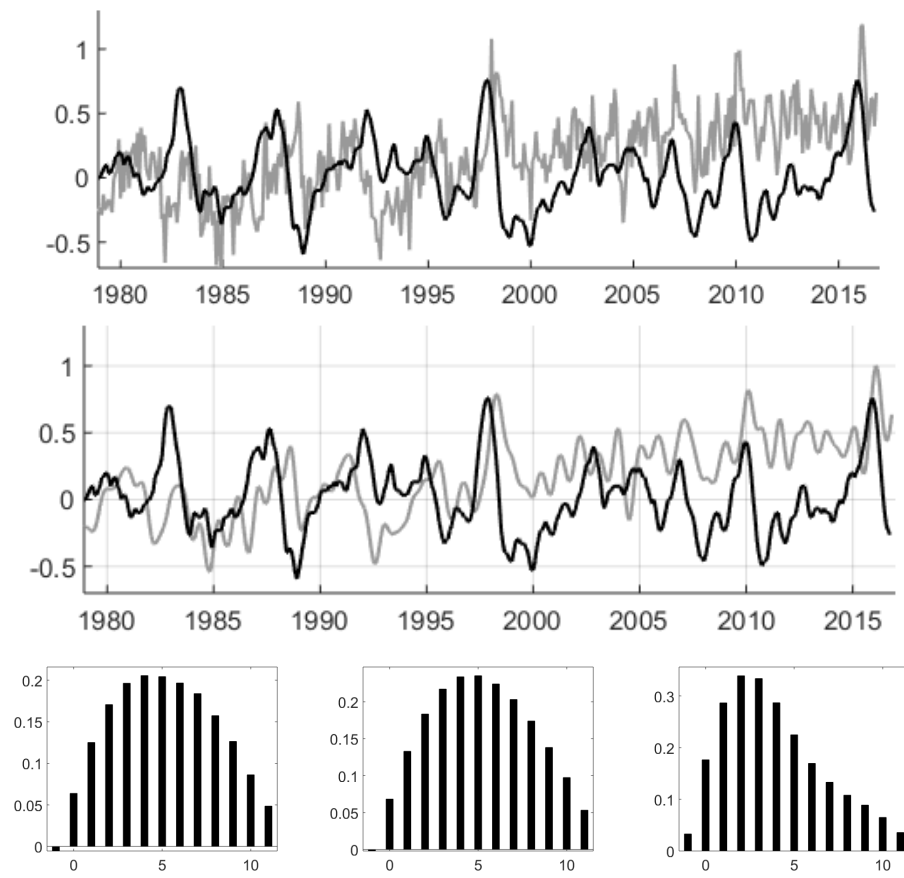


Figure 29: Earth land temperature (gray) and the scaled ONI index (black).

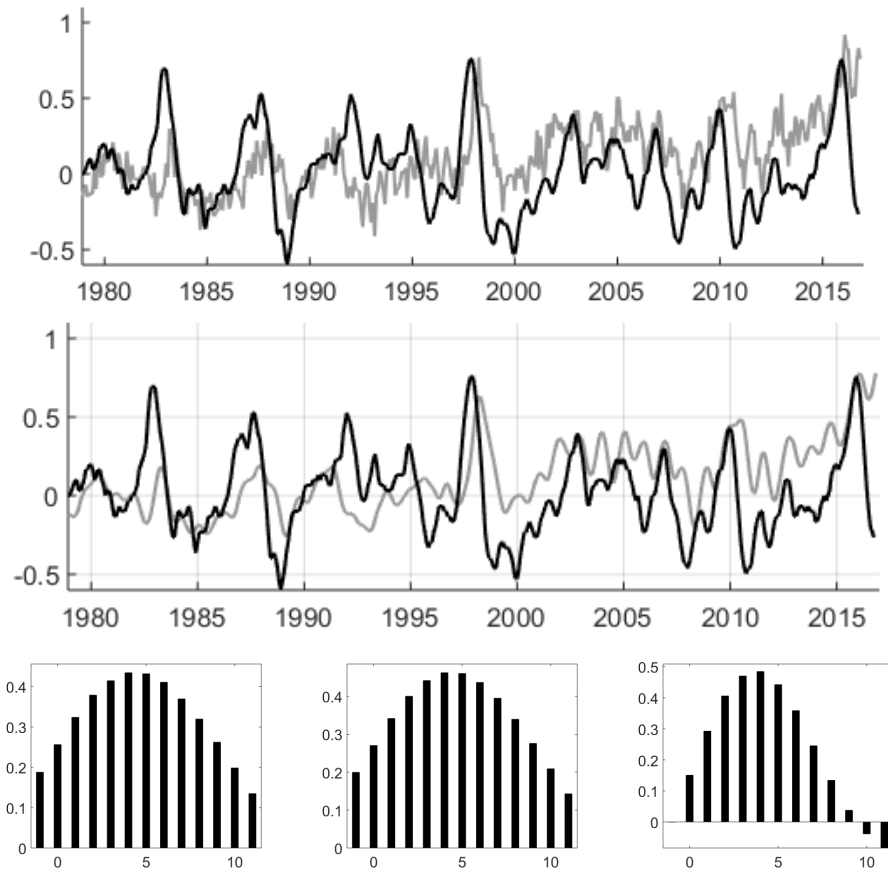


Figure 30: Earth ocean temperature (gray) and the scaled ONI index (black).

5 Arctic and Antarctic sea ice extents

5.1 Analysis and forecasting of ice extent series

In this section, we study Arctic and Antarctic sea ice extents from 1979 to present. The sea-ice extent (unit: 10^6 square Km) is calculated as the areal sum of sea ice covering the ocean where sea-ice concentration (SIC) exceeds a threshold (15% for AMSR-E). SICs are derived from various satellite-borne passive microwave radiometer (PMR) sensors using the algorithm developed and provided by Dr. Comiso of NASA GSFC through a cooperative relationship between NASA and JAXA. Daily data of Arctic and Antarctic sea ice extents is taken from

http://nsidc.org/data/seaice_index/archives.html

On the base of this daily data, we compute monthly time series by taking the average for each month.

In Figure 31 we can see that Arctic sea ice extent during summer months of last decade was smaller than Arctic sea ice extent for the period from 1980 to 2000. However, this decrease of Arctic sea ice extent is compensated by Antarctic sea ice extent so that their sum has very stable behaviour over all period from 1980 to the present time, as can be seen in Figure 38.

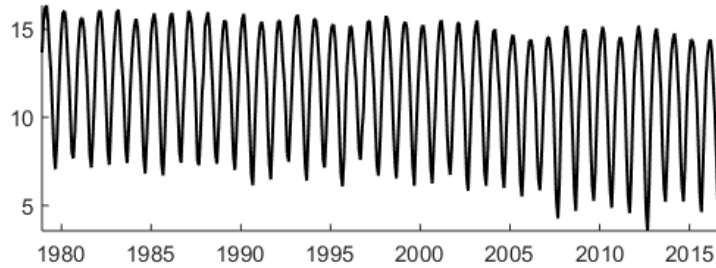


Figure 31: Monthly average values of Arctic sea ice extent from Nov 1978 to Nov 2016.

5.2 Ice extent anomaly

In this subsection we study the time series of the sea ice extent anomaly which is the difference between the original time series of ice extent and the series of corresponding monthly averages. As an increase of temperatures leads to a decrease of sea ice extent, we consider the inverted sea ice extent anomaly which is obtained by changing the sign of the sea ice extent

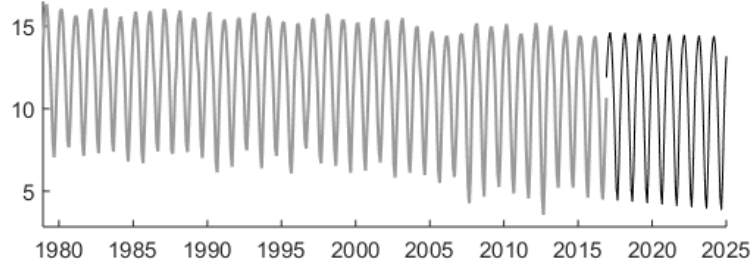


Figure 32: Monthly average values of Arctic sea ice extent from Nov 1978 to Nov 2016 (grey) and forecast with $L = 72$, $r = 7$ (black).

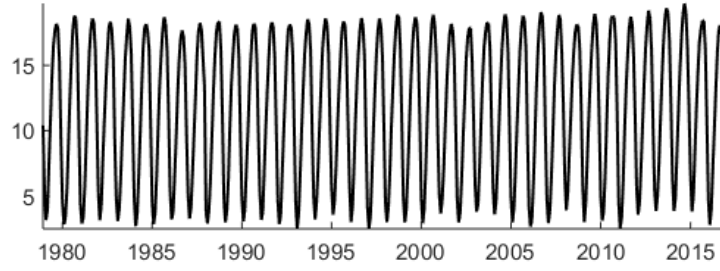


Figure 33: Monthly average values of Antarctic sea ice extent from Nov 1978 to Nov 2016.

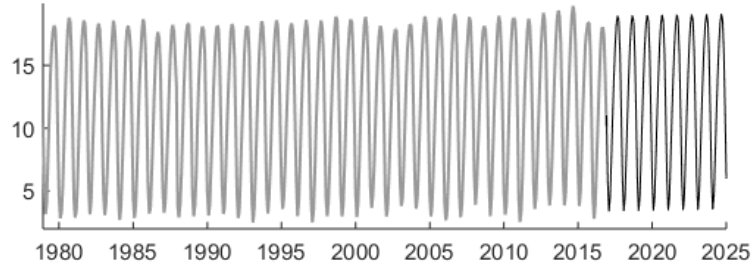


Figure 34: Monthly average values of Antarctic sea ice extent from Nov 1978 to Nov 2016 (grey) and forecast with $L = 72$, $r = 7$ (black).

anomaly.

We investigate the relationship between the inverted sea ice extent anomaly and other time series such as the ONI index (see Figures 39 and 40) and

Figure 35: The annual average values of Arctic sea ice extent from 1980 to 2014.

Figure 36: The annual average values of Antarctic sea ice extent from 1980 to 2014.

Figure 37: The annual average values of the sum of Arctic and Antarctic sea ice extents from 1980 to 2014.

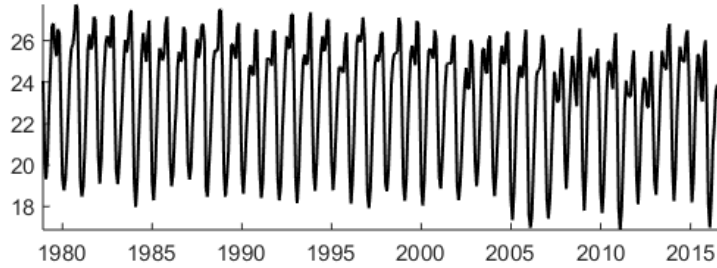


Figure 38: Monthly average values of the sum of Arctic and Antarctic sea ice extents from Nov 1978 to Nov 2016.

temperatures at North and South poles (see Figures 41 and 42).

In Figures 39 and 40 we can notice several significant correlations between the ONI index and the sea ice extent anomalies. We tried to find a confirmation of the existence of such correlations in the literature on Earth climate but we failed. This can be for one of the following three reasons: (a) we did not look hard enough (as we are not specialists in climatology), (b) the correlations found by us are spurious, and (c) the correlations are meaningful and hence they will be rediscovered and explained by the specialists in climate and earth science. We shall leave this to the specialists in climate to resolve.

From Figure 39 we can observe that the ONI index has small influence on the Arctic sea ice extent with delay 2 months. An increase of the ONI index anomalies is followed by a decrease of the Arctic sea ice extent at North Pole (delayed by 2 months).

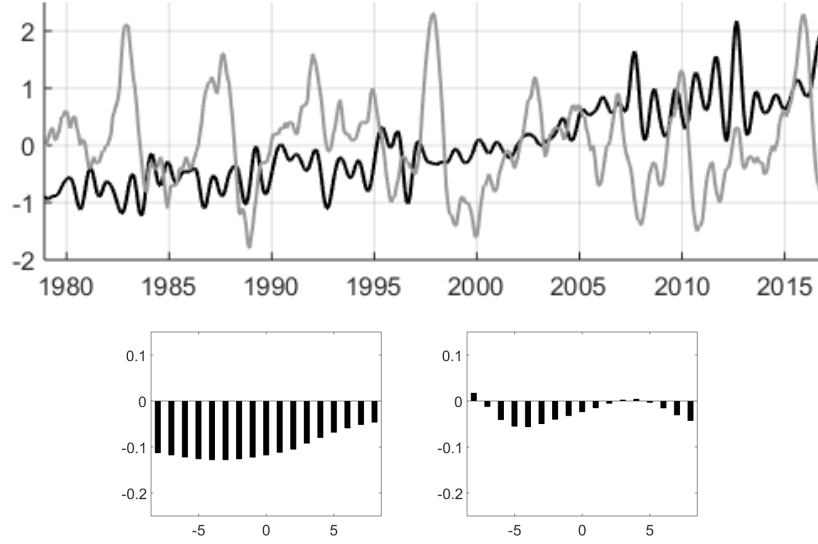


Figure 39: The scaled ONI index (gray) and the inverted smoothed Arctic ice extent anomaly (black). Bottom left: Cross-correlation between the ONI index and the inverted smoothed Arctic ice extent anomaly. Bottom right: Cross-correlation between the differences of the ONI index and the differences of the inverted smoothed Arctic ice extent anomaly.

In Figure 40 we can see that the Antarctic sea ice extent has influence on the ONI index with delay 2 months. A decrease in anomalies of the Antarctic sea ice extent is followed by an increase of the ONI index (delayed by 2 months).

In Figure 41 we can see that the North pole temperature has strong influence on the Arctic sea ice extent. This phenomena can easily be explained.

In Figure 42 we can see that the South pole temperatures have a moderate influence on the Antarctic sea ice extent with delay about 3 months. This can also be easily explained.

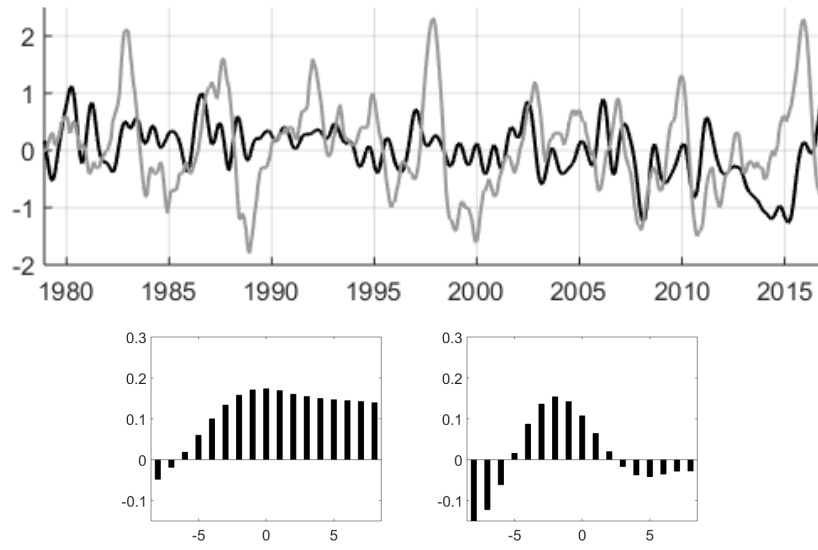


Figure 40: The scaled ONI index (gray) and the inverted smoothed Antarctic ice extent anomaly (black). Bottom left: Cross-correlation between the ONI index and the inverted smoothed Antarctic ice extent anomaly. Bottom right: Cross-correlation between the differences of the ONI index and the differences of the inverted smoothed Antarctic ice extent anomaly.

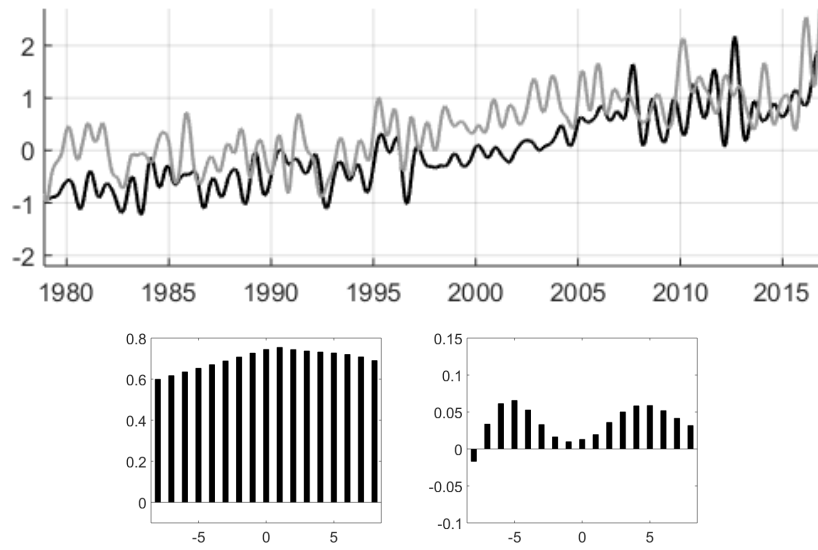


Figure 41: Smoothed North pole temperature (gray) and the inverted smoothed Arctic ice extent anomaly (black). Bottom left: Cross-correlation between the smoothed North pole temperature and the inverted smoothed Arctic ice extent anomaly. Bottom right: Cross-correlation between the differences of the smoothed North pole temperature and the differences of the inverted smoothed Arctic ice extent anomaly.

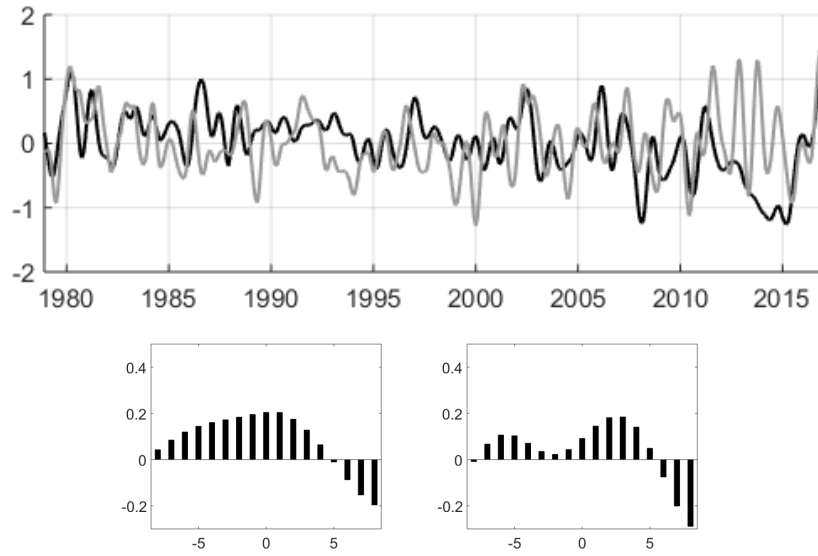


Figure 42: Smoothed South pole temperature (gray) and the inverted smoothed Antarctic ice extent anomaly (black). Bottom left: Cross-correlation between the smoothed South pole temperature and the inverted smoothed Antarctic ice extent anomaly. Bottom right: Cross-correlation between the differences of the smoothed South pole temperature and the differences of the inverted smoothed Antarctic ice extent anomaly.

6 Regression analysis of temperature data on the ONI index

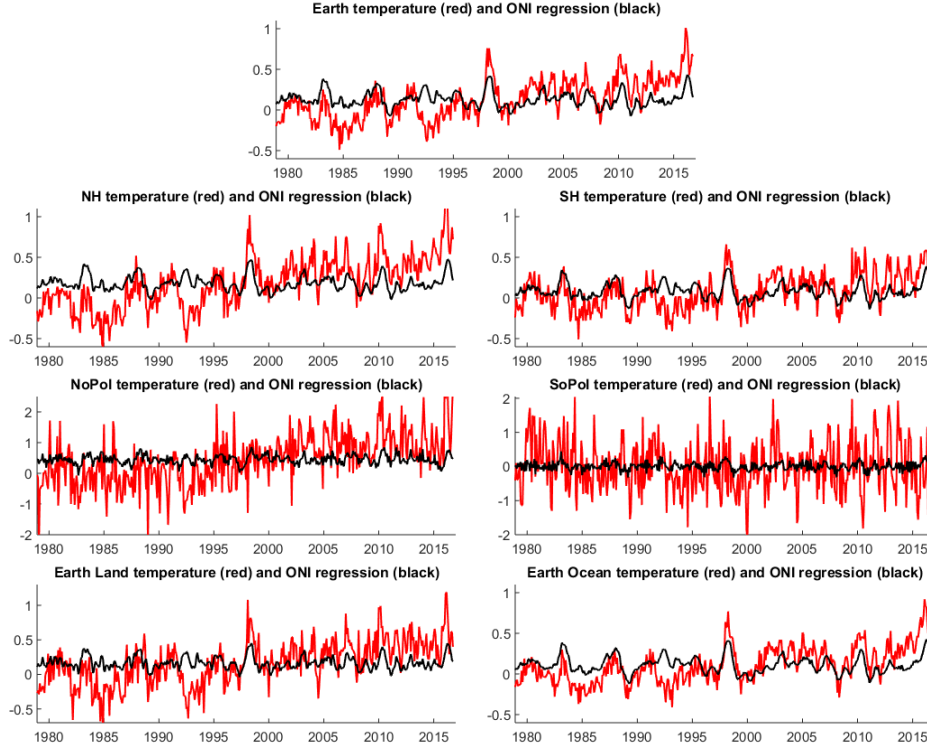


Figure 43: Temperature data and it regression on the ONI index.

Let $x(t_i)$ be a temperature at time t_i for a part of Earth. Let $O(t_i)$ be the ONI index at time t_i .

Consider the regression model

$$x(t_i) = \theta_6 + \sum_{j=0}^5 \theta_j O(t_{i-j}) + \epsilon(t_i).$$

We estimate $\theta = (\theta_6, \theta_0, \dots, \theta_5)$ via least squares, that is, as follows

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \left(x(t_i) - \theta_6 - \sum_{j=0}^5 \theta_j O(t_{i-j}) \right)^2.$$

We define residuals as $R(t_i) = x(t_i) - \hat{\theta}_6 - \sum_{j=0}^5 \hat{\theta}_j O(t_{i-j})$.

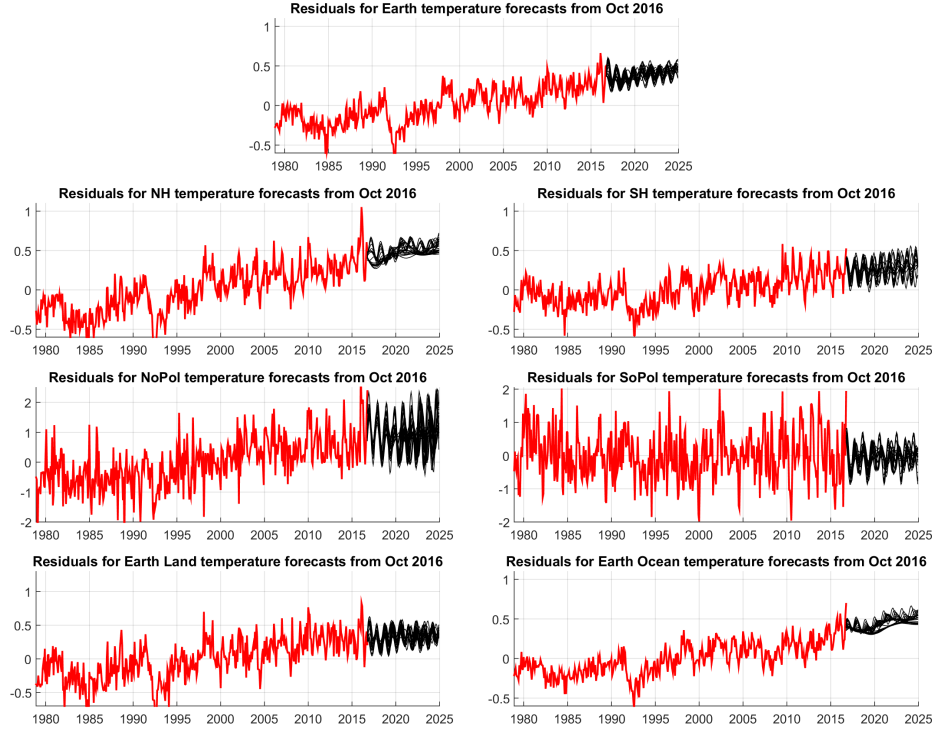


Figure 44: Residuals of regression for temperature data on the ONI index and its forecasts, $L = 30, 31, \dots, 50$ for all parts, $r = 9$ or $r = 5$ depending on a part of Earth.

Conclusions

The main results and finding of this paper are as follows.

- We have made a further investigation of stability of SSA forecasts for the series with no apparent structure. We have observed that the behaviour of SSA forecasts for such series does not change much when we vary the two SSA parameters within a reasonable range.
- We have evaluated the quality of the SSA forecasts of the Earth temperature series that we have made in [6]. We have found out that our forecasts made five years ago were very accurate.

- We have made forecasts of Earth temperature series for the next 5 years and assessed stability of these forecasts. We claim that the pattern observed in the last 15 years for the Earth temperatures is not going to change much. This implies, for example, that sharp rise of Earth temperatures is very unlikely.
- We have analysed cross-correlations between the Earth temperature records and the Oceanic Nino Index (ONI). We have found very high cross-correlations between a lagged ONI index and some Earth temperature series.
- We have applied SSA for the analysis and forecast of the series representing the Arctic and Antarctic sea ice extents and studied cross-correlations between anomalies of sea ice extents, the ONI index and the series representing the temperature anomalies on the North and South poles. In particular, we have noticed several significant correlations between the ONI index and the sea ice extent anomalies which are not well-known in the literature on Earth climate.

ACKNOWLEDGMENT

This research was supported by the Russian Science Foundation, project No. 15-11-30022 "Global optimization, supercomputing computations, and application".

References

- [1] Supplementary materials <http://earth-temperature.com/forecasting/>
- [2] Ghil, M. and Vautard, R. (1991). Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, 350, 324–327.
- [3] Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). Analysis of Time Series Structure: SSA and related techniques. Chapman & Hall/CRC.
- [4] Golyandina, N. and Osipov, A. (2015). Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series. *Statistics and Its Interface* 8, 277–294.

- [5] Golyandina, N. and Zhigljavsky, A. (2013). Singular Spectrum Analysis for Time Series. Springer briefs in statistics, Springer-Verlag.
- [6] Pepelyshev A., Zhigljavsky A. (2010). Assessing the stability of long-horizon SSA forecasting. *Statistics and Its Interface* 3, 321–327.
- [7] Pepelyshev A, Zhigljavsky A. (2017) . SSA analysis and forecasting of records for Earth temperature and ice extents. *Statistics and Its Interface* 10, 151–163.
- [8] Plimer I. (2009). *Heaven and Earth: Global Warming - The Missing Science*. Quartet books, London.
- [9] Schlesinger, M. E. & Ramankutty, N., (1994). An oscillation in the global climate system of period 65–70 years. *Nature*, 367, 723–726.
- [10] Vautard, R., Yiou, P. and Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signal. *Physica D* 58, 95–126.