Труды Карельского научного центра РАН
№ 4. 2014. С. 155–161

УДК 004.01:006.72 (470.22)

STATISTICAL FORECASTING OF EARTH TEMPERATURE RECORDS

V. Kornikov[1], A. Pepelyshev[2], A. Zhigljavsky[2]

[1] *St.Petersburg State University*
[2] *Cardiff University*

In this paper, we continue the research started in [5]. We apply the so-called Singular Spectrum Analysis (SSA) to forecast the temperature records taken from http://vortex.nsstc.uah.edu/, the web-site of the National Space Science and Technology Center, USA, NASA. We compare the forecasts we made three years ago and published in [5] with the temperatures actually recorded. By doing so we demonstrate that our forecasts were quite accurate. We then forecast the temperatures for the next several years and show that the forecasts are not very different from the ones we made earlier: according to our forecasts, the Earth temperatures are neither increasing nor decreasing and will continue to be volatile.

K e y w o r d s: Singular Spectrum Analysis, forecasting.

В. В. Корников, А. Н. Пепелышев, А. А. Жиглявский. СТАТИСТИЧЕСКИЙ ПРОГНОЗ ГЛОБАЛЬНОЙ ТЕМПЕРАТУРЫ ЗЕМЛИ

Эта статья является продолжением исследования, начатого в работе [5], где метод Анализ Сингулярного Спектра применяется для прогноза временных рядов температуры Земли, взятых с сайта Национального Центра Космической Науки и Технологии, США, НАСА, http://vortex.nsstc.uah.edu/. В статье мы сравниваем прогнозы, сделанные три года назад и опубликованные в [5], с температурами, измеренными за эти три года. Сравнение показывает, что наши прогнозы были достаточно точными. Затем мы приводим прогноз температуры на несколько лет вперед и показываем, что эти прогнозы мало отличаются от построенных ранее. Согласно нашим прогнозам, температура Земли не увеличивается и не уменьшается, а будет продолжать колебаться.

К л ю ч е в ы е с л о в а: Анализ Сингулярного Спектра, прогноз.

We consider the temperature records representing the average temperature on Earth and its parts during the last few decades. We make a large number of forecasts of these temperatures. To make the forecasts we apply the so-called Singular Spectrum Analysis (SSA), which is a well-known tool of analyzing and forecasting time series in climatology. To achieve the robustness of forecasts, we choose a wide range of parameters of SSA. Results show that the forecasts are very stable. The most recent forecasts confidently predict that the level of temperatures in the next few years is going to be very close the current level of temperatures. Some additional results can be found at the website *http://earth-temperature.com/forecast/* and in our previous paper [5].

The methodology

Singular Spectrum Analysis

Singular Spectrum Analysis is a well-known tool of analyzing and forecasting time series in climatology in general and analyzing temperature records in particular, see e.g. [2, 4, 7, 8]. There are several versions. We shall use the most basic version. A short description of this algorithm is given below, see [3] and [4] for details.

Short description of SSA

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

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

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

The columns X_j of $\mathbf{X}$, considered as vectors, lie in the L-dimensional space $\mathbf{R}^L$. The singular-value decomposition (SVD) of the matrix $\mathbf{X}\mathbf{X}^T$ yields a collection of L eigenvalues and eigenvectors, which are often called Empirical Orthogonal Functions (EOF). Sometimes, the empirical covariance matrix of the series is used in place of $\mathbf{X}\mathbf{X}^T$. Let us choose the EOFs corresponding to r largest eigenvalues of $\mathbf{X}\mathbf{X}^T$, where r is a given number, $1 \leqslant r < L$. These EOFs determine an r-dimensional subspace in $\mathbf{R}^L$ (call this subspace S_r). The L-dimensional data $\{X_1, \ldots, X_K\}$ is then projected onto this r-dimensional subspace and the subsequent averaging over the diagonals gives us some Hankel matrix $\tilde{\mathbf{X}}$ which is considered as an approximation to $\mathbf{X}$. The series reconstructed from $\tilde{\mathbf{X}}$ satisfies some linear recurrent formula which may be used for forecasting.

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 [4]. The most obvious way is to use the linear recurrent formula which the series reconstructed from $\tilde{\mathbf{X}}$ satisfies. 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 give us a hope that the L-dimensional vectors $\{X_1, \ldots, X_K\}$ lie close to this subspace. The forecasting algorithm then sequentially constructs the vectors $\{X_{K+1}, X_{K+2}, \ldots\}$ so that they stay as close as possible to the chosen subspace S_r.

Choice of SSA parameters

In the examples below, the length of the series is $T = 373$ and the forecasting horizon is $h \cong 100$. If the structure of the series is assumed stable then large values of L of the order $L \cong 100$ should be preferred to small values of the order $L \cong 10$. We, however, do not assume that the structure of the series is rigid. In this case, large values of L would not give SSA enough flexibility. On the other hand, for 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 $20 \leqslant L \leqslant 50$, which we believe is a rather broad range.

The second SSA parameter to choose is r, the dimension of the subspace S_r. The choice of r should depend on what 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 procedures (see e.g. [3]) for 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 and are not well suited for long- and medium range forecasting.

In what follows, we choose $r \in [5, 7]$. We realize that whatever the rule of selection of r, some values of r are too small, which leads to us missing parts of the signal, but other values of r are too large, which means that we include a significant part of the noise 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.

Stability of forecasts

Forming the samples of forecasts

Assume that we have a family of SSA forecasts which is parameterized by a parameter $\theta \in \Theta = \{\theta_1, \ldots, \theta_m\}$, where $\theta = (L, r)$ and m is the total number of chosen pairs of parameters (L, r). For each time moment $t \leqslant T$, 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 $x_1, \ldots, x_t$. Hence for any $t \leqslant T$ we can compute the following set of forecasting results:

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

where $0 < h_1 \leqslant h_2$ are some numbers. The number of elements in the samples F_t is $M = m(h_2 - h_1 + 1)$.

To sum up, our forecasting procedure gives us $T - T_0 + 1$ samples $F_t = \{f_1^{(t)}, \ldots, f_M^{(t)}\}$ at all $t = T_0, \ldots, T$, where T_0 is the first time moment we make the forecasting.

Comparison of the samples

We now need to compare the samples F_t ($t = T_0, \ldots, T$) 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 F_t are $\bar{f}_t = (f_1^{(t)} + \ldots + f_M^{(t)})/M$. As the future is unknown, it is impossible 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 F_t. The most common among these characteristics is the (empirical) standard deviation of F_t:

$$s_t = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(f_i^{(t)} - \bar{f}_t)^2}\,.$$

Another important characteristic of the sample variability is the so-called range of the sample F_t:

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

Results

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 not present the analysis of longer temperature records as there is some controversy around earlier temperature records, see [6]. The series are the so-called temperature anomalies rather than the absolute temperatures (temperature anomalies are computed relative to the base period 1951–1980). Working with anomalies rather than with absolute temperature records is customary in climatology, see for example the publications and web-sites of the Goddard Institute for Space Studies. In the present subsection we follow [5] and use the data from December 1978 to December 2009 so that altogether we have $T = 373$ data points. The first time moment we start the forecasts is January 2005 implying $T_0 = 314$. We forecast the series until 2018 (longer-term forecasts are very similar) by setting $h_1 = 97, h_2 = 99$.

We select the domain $L \in [20, 50]$ for the SSA window length L and choose the first $r \in [5, 7]$ EOFs. Although 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.

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/forecast/. 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 $t = 314, \ldots, 373$ (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 F_t for $t = 325, 337, 349, 361, 373$.

(IV) Figures 3 and 6: forecasts for the temperature at January 2018 using the series $x_1, \ldots, x_t$ for $L = 20, 30, 40, 50$, $r = 5, 7$ and all $t = 314, \ldots, 373$.

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.

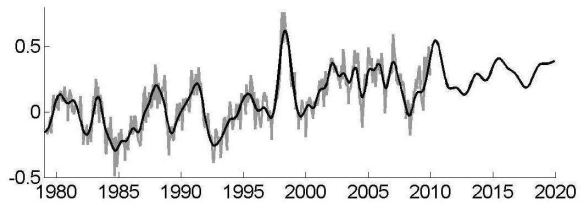

Fig. 1. Earth temperature. The time series (gray), the SSA approximation and the forecast for $L = 50$ and $r = 7$ (black).

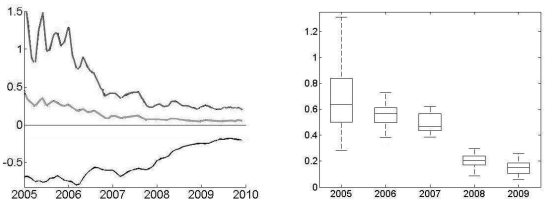

Fig. 2. Earth temperature. Left: averages $-\bar{f}_t$ (black), standard deviations s_t (light grey), ranges R_t (dark grey). Right: box-plots of the samples F_t.

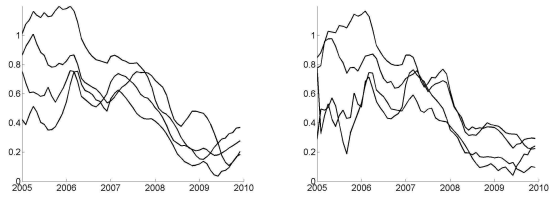

Fig. 3. Forecasts for the Earth temperature at Jan 2018; $L = 20, 30, 40, 50$ and $r = 5$ (left) and $r = 7$ (right).

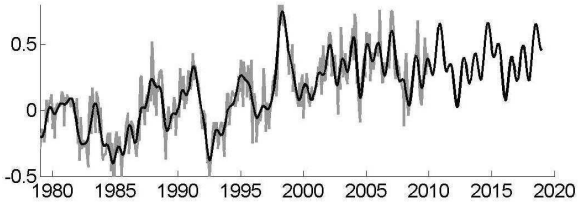

Fig. 4. Northern Hemisphere temperature. The time series (gray), the SSA approximation and the forecast for $L = 50$ and $r = 7$ (black).

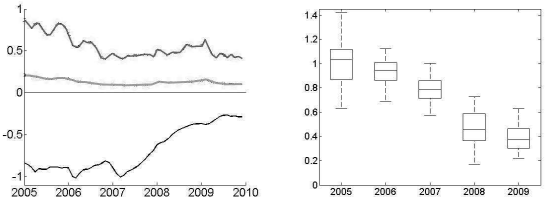

Fig. 5. Northern Hemisphere temperature: Left: averages $-\bar{f}_t$ (black), standard deviations s_t (light grey), ranges R_t (dark grey). Right: box-plots of the samples F_t.

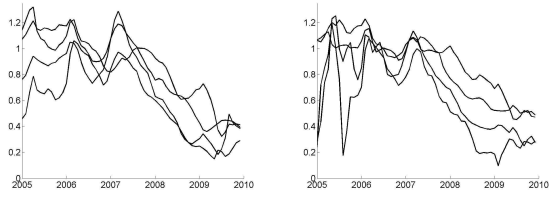

Fig. 6. Forecasts for the Northern Hemisphere temperature at Jan 2018; $L = 20, 30, 40, 50$ and $r = 5$ (left) and $r = 7$ (right).

Assessing the accuracy of forecasts made in 2009

In Figure 7 we can see that during the period Jan 2010 to Sep 2013 the global Earth temperatures roughly followed a typical forecast given in [5].

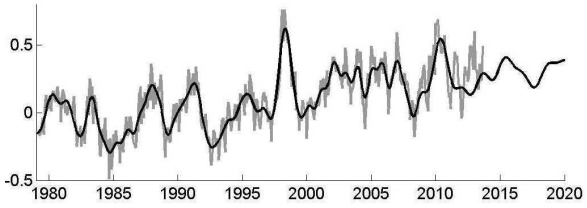

Fig. 7. Global Earth temperature. Gray: the original time series from http://vortex.nsstc.uah.edu/ for the period Dec 1978 to Sept 2013. Black: the SSA approximation until Dec 2009 and the forecast from Jan 2010 onwards. SSA parameters: $L = 50$ and $r = 7$.

Figure 8 shows that for the global Earth temperatures, all forecasts made in [5] for the period Jan 2010 to Dec 2020 were very similar and the actual global Earth temperature during the whole period Dec 1978 to Sept 2013 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.

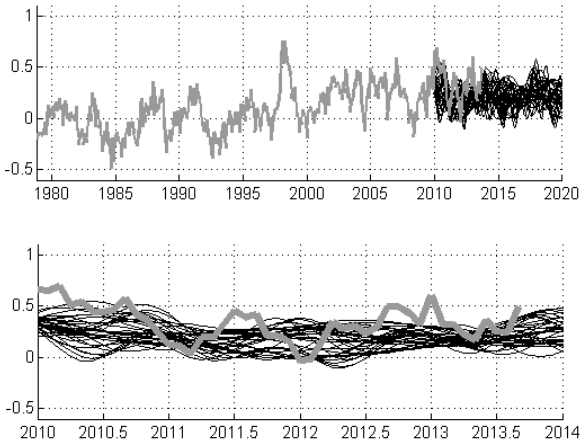

Fig. 8. Global Earth temperature. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

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 Jan 2010 to Dec 2020 were very similar and the actual global Earth temperature during the whole period Dec 1978 to

Sept 2013 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 although 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) very accurately.

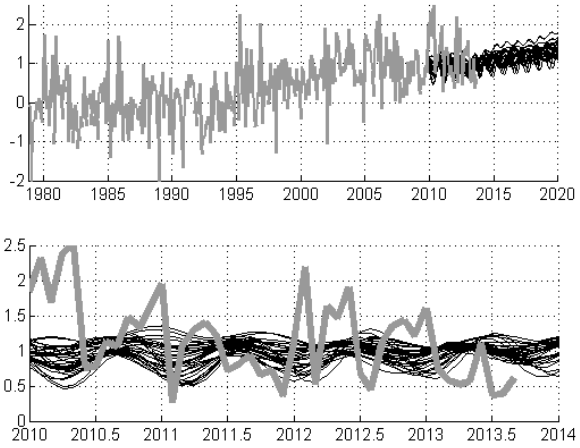

Fig. 9. North Pole temperatures. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

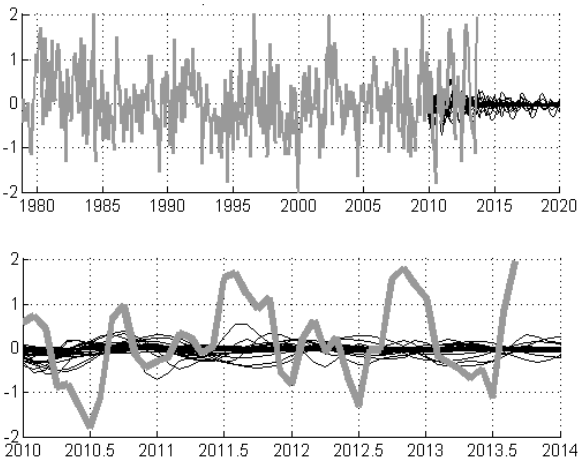

Fig. 10. South Pole temperatures. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

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 showed any trend but 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 tendency apparent.

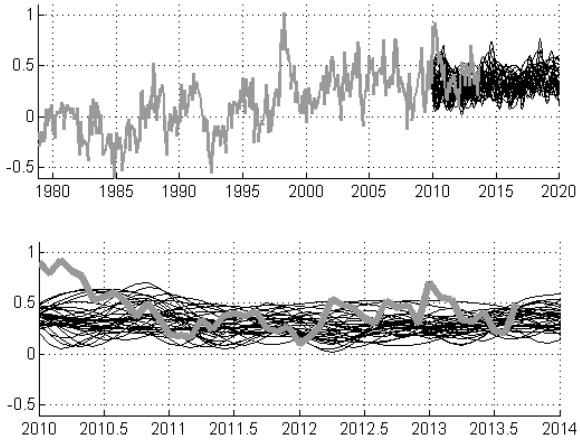

Fig. 11. Northern hemisphere temperatures. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

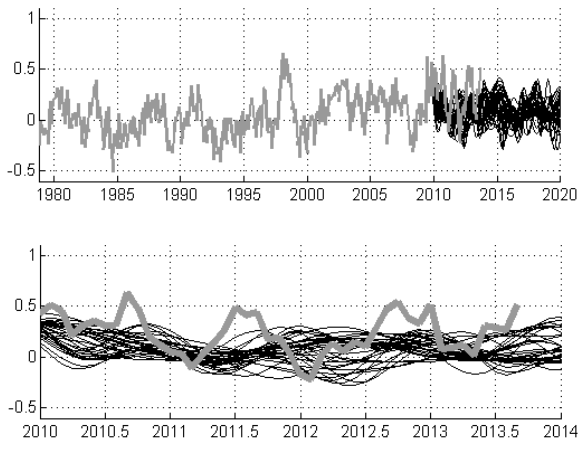

Fig. 12. Southern hemisphere temperatures. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

Figures 11, 12, 13 and 14 are similar to Figures 8, 9 and 10 and deal with Northern hemi-

sphere, 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.

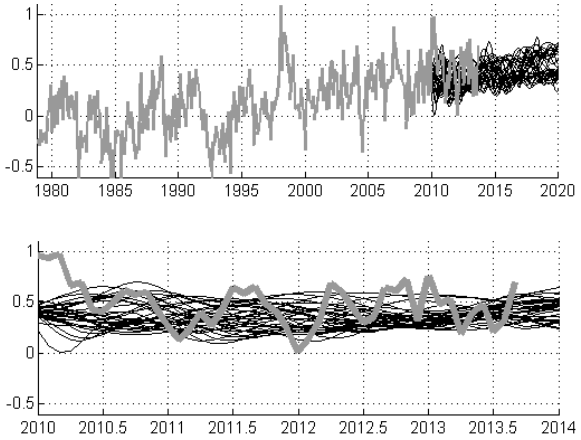

Fig. 13. Earth land temperatures. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

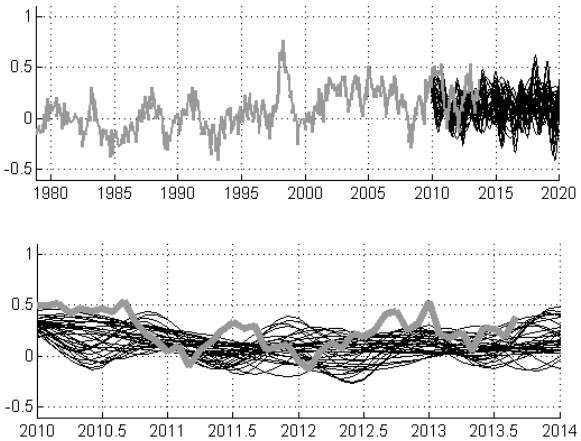

Fig. 14. Earth ocean temperatures. Top: the original time series for the period Dec 1978 to Sept 2013 (gray) and the family of forecasts from Jan 2010 defined by $L = 20, \ldots, 50$ and $r = 5, 7$ (black). Bottom: an extraction from the top graph showing the period Jan 2010 to Dec 2014.

Forecasting from the present time until Jan 2020

In the figures that follow we show the families of the SSA forecasts of different temperatures from Oct 2013 until Jan 2020. To make the forecasts we use SSA with the same parameters as above; that is, $L = 20, \ldots, 50$ and $r = 5, 7$. The data used is the corresponding temperature series from Dec 1978 until Sept 2013, thus the length of all series is 418.

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 were made using the data until Sept 2013 (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. None of the forecasts shows any tendencies for temperature increase. They show, however, a lot of volatility and even a possibility of an insignificant decrease of some temperatures.

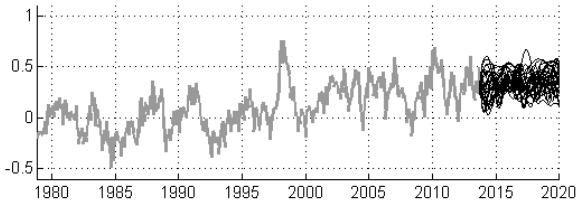

Fig. 15. Earth temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013.

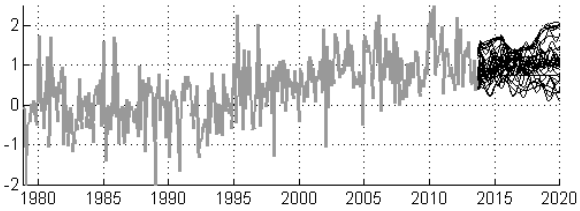

Fig. 16. North pole temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013 for $L = 20, \ldots, 50$ and $r = 5, 7$.

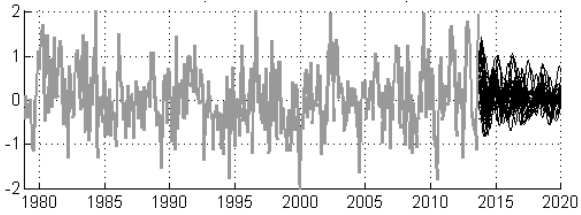

Fig. 17. South pole temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013 for $L = 20, \ldots, 50$ and $r = 5, 7$.

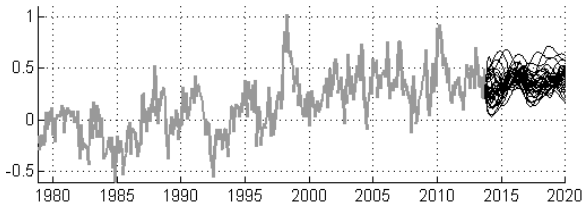

Fig. 18. Northern hemisphere temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013 for $L = 20, \ldots, 50$ and $r = 5, 7$.

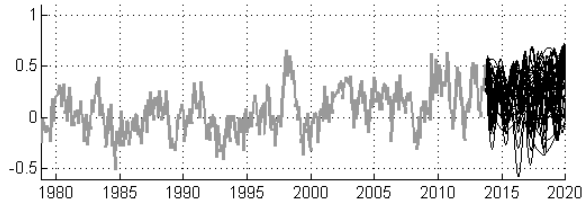

Fig. 19. Southern hemisphere temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013 for $L = 20, \ldots, 50$ and $r = 5, 7$.

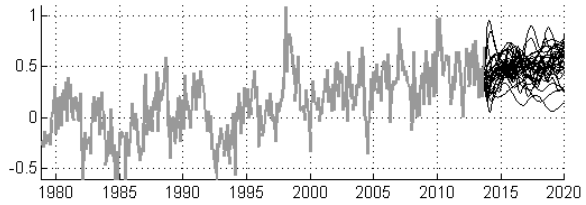

Fig. 20. Earth land temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013 for $L = 20, \ldots, 50$ and $r = 5, 7$.

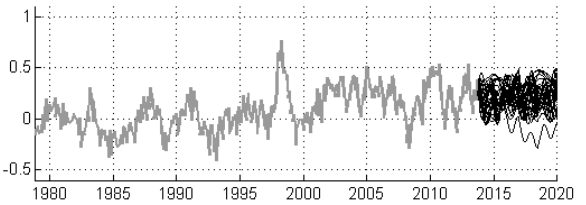

Fig. 21. Earth ocean temperature. The time series (gray) until Sep 2013, the family of forecasts (black) from Oct 2013 for $L = 20, \ldots, 50$ and $r = 5, 7$.

References

1. *Supplementary* materials http://earth-temperature.com/stability/ (дата обращения: 12.09.2013).

2. *Ghil M., Vautard R.* (1991) Interdecadal oscillations and the warming trend in global temperature time series. Nature, 350, 324–327.

3. *Golyandina N., Nekrutkin V., Zhigljavsky A.* (2001). Analysis of Time Series Structure: SSA and related techniques. Chapman & Hall/CRC.

4. *Golyandin N., Zhigljavsky A.* (2013). Singular Spectrum Analysis for Time Series. Springer briefs in statistics, Springer-Verlag.

5. *Pepelyshev A., Zhigljavsky A.* (2010). Assessing the stability of long-horizon SSA forecasting. Statistics and Its Interface 3, 321–327.

6. *Plimer I.* (2009). Heaven and Earth: Global Warming - The Missing Science. Quartet books, London.

7. *Schlesinger M., Ramankutty N.* (1994). An oscillation in the global climate system of period 65–70 years. Nature, 367, 723–726.

8. *Vautard R., Yiou P., Ghil M.* (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signal. Physica D 58, 95–126.

СВЕДЕНИЯ ОБ АВТОРАХ:

Корников Владимир Васильевич
доцент, к. ф.-м. н.
Санкт-Петербургский государственный университет
Университетская наб., 28, Санкт-Петербург,
Россия, 199034
эл. почта: vkornikov@mail.ru

Пепелышев Андрей Николаевич
научный сотрудник, к. ф.-м. н.
Школа математики, Университет Кардиффа
Кардифф, Великобритания
эл. почта: pepelyshevan@cardiff.ac.uk

Жиглявский Анатолий Александрович
профессор, д. ф.-м. н.
Школа математики, Университет Кардиффа
Кардифф, Великобритания
эл. почта: zhigljavskyaa@cardiff.ac.uk
тел.: +44 (0) 29 208 7507 6

Kornikov, Vladimir
Saint-Petersburg State University
28 Universitetskiy pr.
198504 Saint-Petersburg, Russia
e-mail: vkornikov@mail.ru

Pepelyshev, Andrey
Cardiff University
Senghennydd road
Cardiff, UK, CF24 4AG
e-mail: pepelyshevan@cardiff.ac.uk

Zhiglajvsky, Anatoly
Cardiff University
Senghennydd road
Cardiff, UK, CF24 4AG
e-mail: zhigljavskyaa@cardiff.ac.uk
tel.: +44 (0) 29 208 7507 6