П. А. Вайновский, В. Н. Малинин, Ю. В. Митина

СТАТИСТИЧЕСКИЙ АНАЛИЗ ТЕМПЕРАТУРЫ ВОЗДУХА СЕВЕРНОГО ПОЛУШАРИЯ ЗА ПОСЛЕДНИЕ ДВЕ ТЫСЯЧИ ЛЕТ

P. A. Vaynovsky, V. N. Malinin, I. V. Mitina

STATISTICAL ANALYSIS OF AIR TEMPERATURE IN THE EXTRA-TROPICAL NORTHERN HEMISPHERE FOR THE LAST TWO MILLENNIA

Длительные изменения температуры воздуха за последние 2000 лет обсуждаются на основе данных Лундквиста по кольцам деревьев. С помощью вейвлет-анализа выделены нестационарные по частоте и амплитуде 4 цикла, причем вклад 1000-летнего цикла в дисперсию исходного ряда достигает почти 58 %. Предложен метод аппроксимации и долгосрочного прогноза временного ряда температуры воздуха с помощью комплекса сдвиговых авторегрессионных моделей вейвлет-коэффициентов. Обсуждаются возможные изменения климата до 2100 г.

***Ключевые слова**: изменения климата, температура воздуха, вейвлет-анализ, авторегрессионные модели, циклы.*

We consider long changes in the air temperature during the last two millennia, based on Lundqvist data. Using wavelet analysis we distinguished 4 cycles transient in frequency and amplitude, the contribution of the 1,000th cycle in the dispersion of the initial series being nearly 58 %. The method of approximation and the long-term forecast of the air temperature time series using complex shear autoregressive models of wavelet coefficients are proposed. Possible climate change until 2100 is discussed.

***Keywords**: climate change, air temperature, wavelet analysis, autoregressive models, cycles.*

Introduction

The main indicator of climate is the air temperature in the near-surface layer of the atmosphere. Relatively reliable estimates of its variations can be obtained only on the basis of instrumental measurements, the further they receding into the past, the less being their accuracy and the more approximate being the conclusions. Nowadays, there are several time series of reconstruction of the global air temperature (GAT), but since they are based on almost the same time series of temperature and on similar quality models, by which the retrospective calculation of GAT has been carried out, it is essential that they all give similar results of global climate change over the past century and a half. It is considered significant that since the 20th century the GAT has been growing, the linear trend describing more than 60 % of GAT dispersion [4]. In fact, the linear trend is the global warming itself.

What are the causes of today's climate change? According to IPCC experts the dominant cause of modern climate change is anthropogenic emission of CO_2 into the atmosphere, since the mid-20th century its probability reaching 98 % [8]. However, the climate had changed before intensive industrialization of human society, i.e. until the 20th century. Obviously, one can understand its changes only via paleoclimatic reconstructions, their reliability being determined by the quantity and quality of the materials used.

In 2010, Ljungqvist [9] proposed a reconstruction of temperature variation in the extra-tropical Northern Hemisphere (90–60° N) for the last two millennia based on a thorough analysis of 30 tree rings (fig. 1), the data of 16 rings being used since AD 1. Fig. 1 shows that the Medieval Warm Period (10–11 centuries), the Little Ice Age (16–17 centuries) and the current warming of the 20th century are very clearly visible. The values of temperature have decadal resolution. Soon this multi-proxy reconstruction was updated to 91 tree rings, 26 of them dating back to the 1st year of AD [7]. Compared with the previous reconstruction, these temperature data already have annual resolution. First of all, we should note that common patterns of temperature fluctuations for both reconstructions remain on the qualitative level: the Medieval Warm period (10–11 centuries), the Little Ice Age (16–17 centuries) and the current warming in the 20th century are distinguished. However, in quantitative terms, there are significant differences for the Little Ice Age and the 13th century, when the temperature of the latter reconstruction is more than 0.6 °C lower than the temperature of the former. In the 17th century the minimum temperature, being 1 °C lower than the temperature of the base 1880–1960 period, was distinguished. The maximum temperature is more clearly defined as well, being observed in 950–1050 and exceeding 0.6 °C relative to the base period. Moreover, the temperature variability (dispersion) of the new reconstruction more than 1.5 times exceeds the previous one.

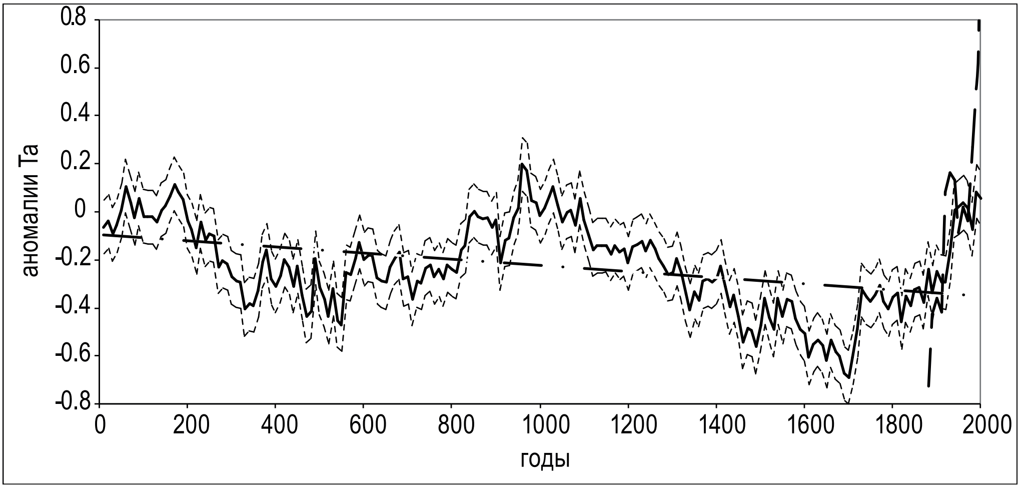

Fig. 1. The time variation of the average decadal air temperature anomalies in the extra-tropical Northern Hemisphere (90–30° N) relative to the base 1961–1990 period according to Ljungqvist [9]. Thin dotted lines are the RMS temperature errors. The dashed line — the observed decadal air temperature anomalies in the extra-tropical Northern Hemisphere according to the archive CRUTEM3 + HadSST2 data

Naturally, the fundamental question is the accuracy of these multi-proxy reconstructions. Unfortunately, accuracy assessment is possible only in an indirect way, namely, by comparing them with the air temperature data over the period of instrumental observations or with results of similar reconstructions. Fig. 2 shows the distribution of time series [9] and [7], as well as the land surface air temperature anomalies in the extra-tropical Northern Hemisphere and the globe as a whole based on instrumental data of NOAA GHCN (https://www.ncdc.noaa.gov/cag/time-series/global) for the period 1880–2000. More detailed documentation of the reanalysis archive GHCN (Global Historical Climate Network) could be found in [10].

Attention is drawn to the sharp contrast of Christiansen and Ljungqvist series [7] to other time series. The maximum discrepancy is noted in the 1940–1950 and reaches 1 °C. At the same time, the series by Ljungqvist [9] corresponds well with the time series of land surface air temperature anomalies of the extra-tropical Northern Hemisphere and the globe as a whole, except for the last 20 years of the 20th century. It is likely that with a sharp increase in air temperature within the specified period the trees are simply unable to adapt quickly to new warming conditions and apparently reflect the natural climate change to a greater extent. These time series discrepancies are caused by features of air temperature recovery method, using the data on the tree rings (LOC reconstruction). According to the authors [7], this method significantly exaggerates and distorts the high-frequency temperature variability with periods less than 10 years, therefore, the initial series with decadal averaging or even 50-year sliding averaging should be used.

Thus, despite the significantly fewer trees the time series of Ljungqvist more accurately reflects the temperature variation in the 20th century, and possibly in the previous centuries. Therefore, it is this series that we use in the statistical analysis of the temperature variation over the past two millennia.

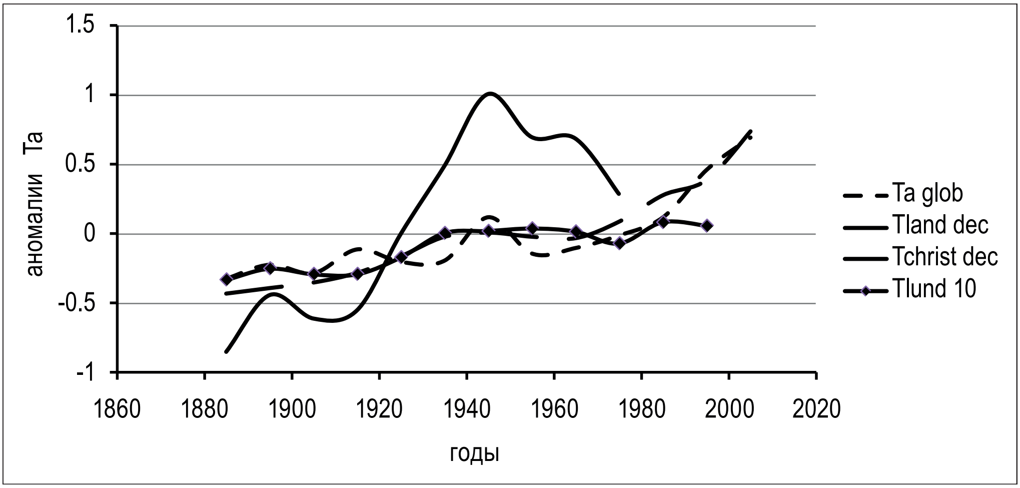

Fig. 2. Interannual variations of air temperature anomalies time series in the extra-tropical Northern Hemisphere by Ljungqvist (solid line) and Christiansen, by Ljungqvist (solid line with a marker), obtained from the data on tree rings, the land surface air temperature anomalies in the extra-tropical Northern Hemisphere (dot-dashed line) and the globe in whole (dashed line) according to NOAA data [https://www.ncdc.noaa.gov/cag/time-series/global] for the period 1880–2015

Air temperature variation during the last two millennia

We consider the temporal variability of time series ($n = 200$) of the Northern Hemisphere air temperature anomalies by Ljungqvist (TA_{NH}) based of the following additive model [5]:

$$X(t) = Tr(t) + C(t) + P(t), \tag{1}$$

where $Tr(t)$ is the trend component; $C(t)$ — cyclic component, characterizing regular (cyclic) interannual variations; $P(t)$ — remainder, characterizing random variations. Fig. 1 shows that this series has a negative trend, i.e. the air temperature has been slowly decreasing during two millennia with the rate −0.0013 °C per decade. This trend describes 19 % of the initial series dispersion. The trend is likely to be the part of the longer cycle than the considered time period, or to characterize the general direction of climatic system towards next ice age.

To analyze cyclic component after removing the linear trend in decomposition (1) we have used a set of statistical methods: harmonic, spectral and wavelet-analysis. The temperature time series has appeared to have a rather complex poly-cycle structure. The spectrum of the time series has showed the dominance of 1075-year harmonics, amplitude of which almost 5 times exceeds the next harmonics with the period of 444 years. Moreover, the harmonics with the lengths of 227 and 112 years are distinguished in the temperature values.

Table 1 gives statistical estimates of these harmonics. The contribution of the first harmonica to the dispersion of the temperature series after trend removal is 70.8 %, to the dispersion of the initial series — 57.8 %. It is essential that reliability of this harmonica is of no doubt. The contribution of other harmonics varies within 2–4 %. Significance of harmonics can be checked by the Student's criteria. In [5] it is shown that the critical contribution to the dispersion of the initial series is determined as $k_{icr} \approx 4/(n+2)$ for long enough time series, the significance level being α=0.05. If $k_i > k_{icr}$, harmonica is significant. As $n = 200$, then $k_{icr} \approx 0.02$, or 2.0 %. Table 1 shows that even last harmonica equal to 112 years is on the edge of significance despite being small. As a result, random variations amount only to 14.3 %. As a whole, harmonics estimates correspond to the well-known fact that increase in harmonics periods in air temperature leads to increase in their amplitudes.

Table 1

Statistical characteristics of different components of the equation (1) to [2]

Component	Dispersion, (D °C)²	Contribution to dispersion of the initial series, %
Initial series	0.0337	100
Linear trend	0.0063	18.7
1075-year harmonics	0.0194	57.8
444-year harmonics	0.0011	3.3
227-year harmonics	0.0013	3.9
112-year harmonics	0.0007	2.0
Random component	0.0049	14.3

Thus, on the basis of cycle estimates the equation (1) in the dispersion form may be written as follows:

$$D(\Delta TAsp) = D(Tr) + D(C_{1075}) + D(C_{444}) + D(C_{227}) + D(C_{112}) + D(P) = \\ = 0.19 + 0.58 + 0.03 + 0.04 + 0.2 + 0.14. \quad (2)$$

We should note that in [3] given is the air temperature spectrum of the northeastern Europe for two millennia obtained by the method of maximum entropy. A quasi-millennia cycle is absent there, although the temperature maximum is shown clearly at the beginning of the 11th century. Obviously, it has gone into the trend. The main is the half-century cycle (499 years). Moreover, expressed are the cycles of 195, 73 and 48 years. According to the authors [3] half a century cycle and two hundred-year cycles are caused by solar activity. In principle, these cycles are similar to the ones shown in Table. 1.

Fig. 3 shows the time variation of the initial series [9], calculated as the sum of the trend components and 4 harmonics and the remainder series as the difference of the initial and calculated values. Remainder series contains some oscillations with a period considerably less than 100 years, their analysis showing to have random character. But if we consider specific, shorter periods of time, there could arise a false idea of their importance. In particular, this applies to 60-year cycle, which is registered twice on the instrumental data in the interannual variations of the global temperature, but no longer in multi-proxy reconstructions.

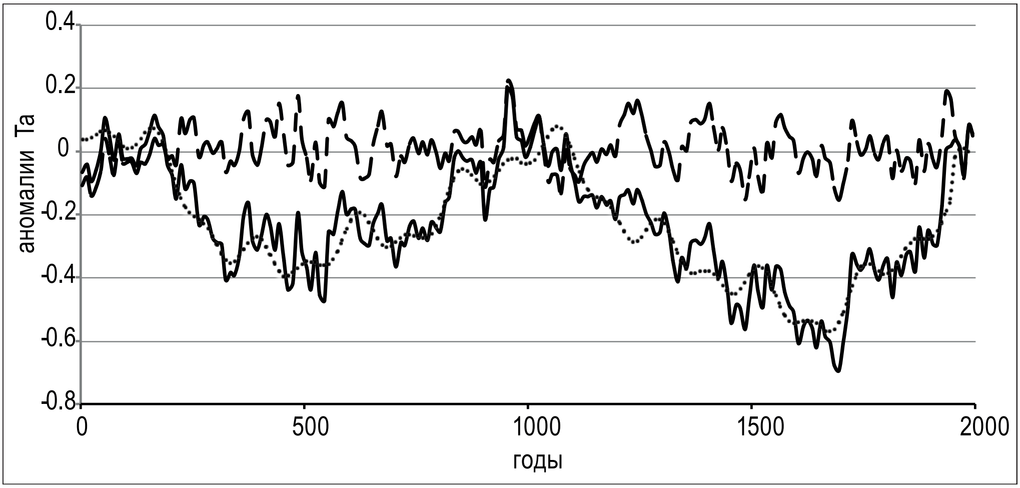

Fig. 3. The time variation of decadal values of temperature anomalies of the time series [9] (solid curve), its recovery values as the sum of the trend component and 4 Fourier harmonics (curve in the form of dots) and the remainder series (dashed curve)

It is known that a fundamental disadvantage of the spectral analysis is that it allows to distinguish harmonic fluctuations only in stable conditions, i.e. with strictly time-constant parameters: amplitude, period and phase. However, in fact, the natural processes are usually cyclical, whose parameters experience irregular variations over time within a certain range. The most appropriate method of distinguishing cyclical fluctuations is wavelet analysis. After removing temperature trend and 1075-year harmonic from the initial series the wavelet decomposition by Morlaix method has been carried out (fig. 4).

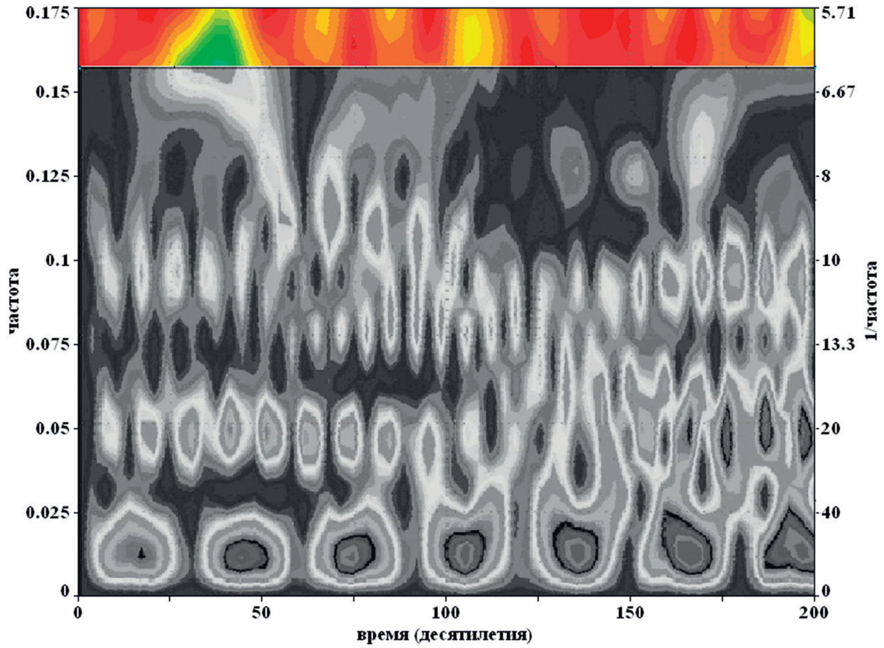

Fig. 4. The wavelet decomposition of decadal average values of air temperature anomalies in the extra-tropical Northern Hemisphere over the period 0–2000 using the Morlaix method according to [9]. Years and periods are given in decades

Fig. 4 shows that the cycles in Table 1 are clearly seen in the wavelet decomposition as well. But thereat, they are non-stationary in amplitude and frequency (period). Especially significant period variations are specific for quasi-hundred-year cycle, which varies in the range of 90–130 years. Less pronounced are the variations in frequency of quasi-two-hundred-year and quasi-three-hundred-year cycles. At first glance, the variations in the amplitude are insignificant. However, in fact it is not so. This is easily seen, if we refer to fig. 5, which shows time variation of the wavelet coefficients corresponding to the cycles shown in table 1 quasi-four-hundred-year cycle is most clearly expressed in the first millennium. Then, until the 16^{th} century, practically it is not observed. The maximum intensity of quasi-two-hundred-year cycle is observed, on the contrary, since the 16^{th} century, and its minimum intensity — in the 11–15 centuries. The maximum amplitudes of the centennial cycle are confined to the 9–10 and 19–20 centuries. We should note that the cycle periods do not remain constant. For example, in the first millennium the period of quasi-two-hundred-year cycle is slightly higher than 200 years (~240–250 years), and when it appears again in the 16th century, its period is close to 200 years. So, it is quite obvious that the variability of cycle period is significantly lower than variability of their amplitudes.

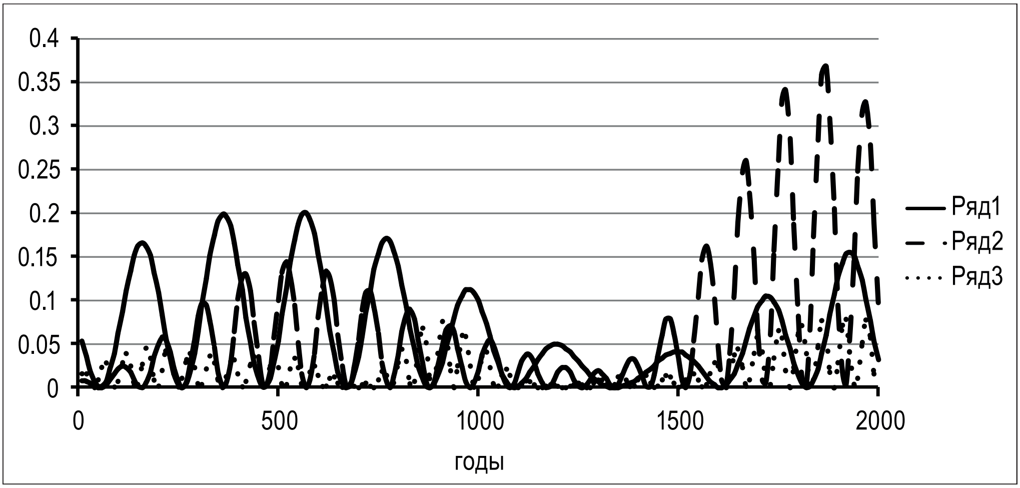

Fig. 5. Time variation of normalized wavelet-coefficients of air temperature anomalies time series for 0–2000 for the frequencies corresponding to the periods of 444 years (solid line), 227 years (dashed line) and 112 years (dotted line)

On the possibilityof air temperature variation in the 21st sentury

The problem of estimating possible air temperature variation during this century is of considerable scientific and practical interest. Nowadays, it can be solved primarily by using general circulation models of the atmosphere and the ocean, and by setting different scenarios of climate change, based mainly on estimates of the potential growth of carbon dioxide emissions into the atmosphere. In this case the natural climate variability is completely rejected [8]. However, as we have seen above, the pre-industrial temperature variability was formed mainly by 4 cycles. This means the possibility of using multi-proxy data for super long-term temperature forecast resulting in the necessity of approximating the Ljungqvist series. For this purpose, harmonic analysis has been used in the stationary approximation by 4 harmonics and wavelet decomposition in accordance with their pronounced unsteadiness. If the calculation of harmonics at any given time is a trivial and easy task, the calculation of cycles by the wavelet coefficients is not self-explanatory.

Consideration of the basic cycles amplitude variability has been carried out at the first phase, consideration of their frequency (period) variability — at the next stage. Fig. 5 shows that the variability of the cycle amplitudes is very complex and clear increase in their amplitude is noted only in recent quincentenary. For each of the 4 time series of wavelet coefficients for the 0–1500 period (dependent sample is $n = 150$ values) the classical autoregressive model of p-order has been built in the form of:

$$X^o(t) = \alpha_1 X^o(t-1) + \alpha_2 X^o(t-2) + \ldots + \alpha_p X^o(t-p) + Z(t), \qquad (3)$$

where $X^\circ(t) = X(t) - \overline{X}$-centered random process, $Z(t)$ — white noise.

The main aim of decomposition (3) is to find the optimal order of p-model. For this purpose, the method of optimal linear extrapolation of Yu. Alekhin [1] has been used according to which the first 15 autoregression models have been calculated and standard error and coefficient of determination R^2 between calculated and initial time series of wavelet coefficients have been determined for each of the models. Autoregression coefficients have been calculated by solving the system of normal equations of Yule-Walker. It has been found that for the considered cycles for the dependent sample 0–1500 years it is reasonable to consider $p=6$ as an optimal order, wherein R^2 value amounted to $R^2 = 0.6$–0.7. Hereafter, for the independent sample of 1500–2000 period we have calculated wavelet-coefficients using shear autoregression model with a step $n=10$. The main point of the shear autoregression has been as follows. Extrapolation of wavelet coefficients 100 years ahead (1500–1600) has been carried out at the first step. Then, the dependent sample has been shifted by 10 values (100–1600), for which the procedure of determining optimal autoregression models has been repeated. Using the optimum model obtained we have made extrapolation of wavelet coefficients for the next hundred years. Hereafter, we have performed another shift of dependent sample (200–1700 period) and the whole procedure has been repeated. Using of this procedure has been dictated by the aim to reduce the computational errors of autoregression models. It is worth noting that the amount of these calculations carried out is 20, each calculation requiring determination of the optimal order of p-model. As a rule, it has been in the range of $p=6$–9.

Calculation has allowed to make a matrix of wavelet-coefficients sized 4×200. Further, we have used the reverse wavelet-decomposition:

$$X(t) = \frac{1}{K_\psi} \sum \sum C(a,b) \psi\left(\frac{t-b}{a}\right) \frac{1}{a^{1/2}} \frac{\Delta a \Delta b}{a^2}, \qquad (4)$$

where $X(t)$ is the reconstructed series at t time; $\psi(t-b)/a$ — Morlaix wavelet function; $C(a,b)$ — wavelet coefficients being time-shift function b and wavelet scale a; K_ψ — normalizing coefficient dependent on the choice of the wavelet function form. Due to transformation (4) it is easy to restore the original time series. The determination coefficient between the restored and original TA_{NH} series for 1500–2000 makes $R^2=0.61$, the RMS error of TA_{NH} estimates being equal to 0.08 °C (table 2). As expected, the accuracy of this method of calculating TA_{NH} is significantly higher compared with the harmonic analysis (table 2).

However, the accuracy of TA_{NH} series approximation by wavelet decomposition considering only cycle nonstationarity in amplitude is still not sufficiently high. Therefore, there is a need for additional consideration of variability of the cycle frequency. Based on the results of wavelet decomposition and spectral analysis of the Ljungqvist series, the frequencies, which carry information about the cycles shown in Table 1, have been distinguished. If the main 1075-year cycle is best seen at two frequencies (0.0097 and 0.0116), the most variable quasi-hundred-year cycle — at 5 frequencies. Therefore, in addition to the 4 basic time series of wavelet coefficients we have distinguished another 10. For each of them we have used a shear autoregression procedure. The resulting matrix of wavelet coefficients of 14´200 size has been obtained. After that, using the inverse wavelet transformation (4) we have calculated the air temperature time series of Ljungqvist. Table 2 shows thus calculated estimates of determination coefficient and RMS error for independent samples for 1500-2000. It is easy to see that the approximation accuracy of TA_{NH} series has grown substantially, its RMS error 2 times decreasing.

Table 2

Statistical estimates of comparison of initial and calculated by different methods air temperature series in the North Hemisphere [9] for the 1500–2000 independent sample

Parameter	Harmonic analysis by 4 harmonics	Wavelet-decomposition	
		Taking into account amplitude nonstationarity	Taking into account amplitude and frequency nonstationarity
Determination coefficient	0.53	0.61	0.79
root-mean-square of calculated series, °C	0.10	0.08	0.04

Discussion of the results

The results obtained indicate the possibility of an approximate estimation of prognostic values of air temperature until the end of this century. Using the procedure described above (dependent sample 500–2000, step n = 10), we have carried out the prognostic TA_{NH} calculation for 100 years, i.e. till 2100, taking into account non-stationarity of cycles in amplitude and frequency (fig. 6). Fig. 6 shows that the harmonic "curve" of TA_{NH} sharply goes down. This behavior is due to its thousand-year cycle, whose contribution to the variance of the original series exceeds 50 %. Obviously, this cycle represents mainly natural variations of climate acquiring the tendency of cooling after 2000.

Fig. 6. Comparison of multi-proxy (solid line) and calculated by different methods data until 2000: harmonic analysis — dots; wavelet analysis, taking into account nonstationarity cycles in amplitude — dot-dash; wavelet analysis, taking into account nonstationarity cycles in amplitude and frequency — dotted line. Air temperature anomalies forecast by Ljungqvist up to 2100

More interesting are the results obtained by the wavelet analysis. Until mid-century the TA_{NH} values almost match the air temperature over the period 1961–1990. However, then wavelet-based curve taking into account nonstationarity in amplitude goes down, and taking into account the nonstationarity in amplitude and frequency, on the contrary, increases. The question is how to interpret the results obtained, i.e. how can they be correlated with instrumental data on air temperature? Let us leave this question open. Most importantly, the prognostic estimates of TA_{NH}, reflecting mainly the natural climate variability, do not show the monotonic increase in temperature as it is accepted by the IPCC experts. One more important fact should be noted. The fact that the trend value in the air temperature for the period 1880–2000 in Ljungqvist data is 2–2.5 times underestimated compared with the reanalysis archive. On the one hand, it is obvious that the values of the air temperature do not adequately reflect the current global warming, but on the other hand, there may be reasonable suspicion in artificial trend overestimation by the reanalysis archives developers. For example, a detailed comparison of data on air temperature in Russia used in HadCRUT archive with all the present data in the Roshydromet data-base without conducting any meaningful selection has showed that «*overestimating the extent of warming, carried out by HadCRUT staff for the territory of Russia from 1870s till 1990s can be valued at least 0.64 °C*» [6].

Conclusion

The statistical analysis of temperature variations in the extra-tropical Northern Hemisphere (90–30°) calculated by Ljungqvist for the last two millennia using data of tree rings has been carried out. The trend and 4 cyclic variations, which contribution to dispersion of the initial series makes 19 % and 67 % respectively, have been distinguished. Thereat, the main is the quasi-thousand-year cycle which contribution reaches nearly 58 %. By means of the wavelet-analysis it has been established that the specified cycles are nonstationary in frequency and amplitude. The method of approximation of air temperature time series has been proposed on the basis of a complex of shift autoregression models of the wavelet-coefficients considering nonstationarity of cycles both in frequency and amplitude. For independent sample of 1500–2000 it has been shown that the coefficient of determination of the initial series with the one calculated by models makes 79 %, and the standard error of estimate of air temperature is equal to 0.04 °C. This has served as the basis for prognostic calculation of air temperature variations till the end of the 21st century. It has been shown that the prognostic estimates of temperature reflecting mainly natural variability of climate do not show its monotonic increase, as it is accepted by IPCC experts.

References

1. *Алехин Ю.М.* Статистические прогнозы в геофизике. — Л.: Изд-во ЛГУ, 1963. — 82 с.
2. *Вайновский П.А., Малинин В.Н.* Об изменениях температуры воздуха северного полушария за последние 2000 лет // Общество. Среда. Развитие. 2015.
3. *Клименко В.В., Мацковский В.В., Дальманн Д.* Комплексная реконструкция температуры российской Арктики за последние два тысячелетия // Арктика: экология и экономика. 2013. № 4(12). — С. 84–95.
4. *Малинин В.Н.* Уровень океана: настоящее и будущее. — СПб.: РГГМУ, 2012. — 259 с.

5. *Малинин В.Н.* Статистические методы анализа гидрометеорологической информации. — СПб.: РГГМУ, 2008. — 407 с.
6. *Пивоварова Н.А.* Как делается потепление. Случай России. — М.: Институт экономического анализа, 2009. — 21 с.
7. *Christiansen B., Ljungqvist F.C.* The extra-tropical Northern Hemisphere temperature in the last two millennia: Reconstructions of low-frequency variability // Clim. Past. 2012. Vol. 8. — P. 765–786.
8. *IPCC.* Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change / Eds. Stocker T.F. at al. — Cambridge (UK); N.Y.: Cambridge University Press, 2013. — 1535 p.
9. *Ljungqvist F.C.* A new reconstruction of temperature variability in the extratropical northern hemisphere during the last two millennia // Geograf. Annal. Ser. A. 2010. Vol. 92. — P. 339–351.
10. *Smith T.M., Reynolds R.W., Peterson T.C., Lawrimore J.* Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006) // J. Climate. 2008. Vol. 21. — P. 2283–2296. — DOI:10.1175/2007JCLI2100.1.