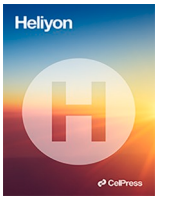
Heliyon 10 (2024) e23237

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Wavelets in the analysis of local time series of the Earth's surface air

Alexandr Volvach *, Galina Kurbasova, Larisa Volvach

Radio Astronomy and Geodynamics Department of Crimean Astrophysical Observatory RAS, Katsively, RT-22, Crimea, Ukraine

ARTICLE INFO

Keywords:
Earth's surface air
Ground-based measurements
Average annual temperature
Forecast
Numerical models

ABSTRACT

The practical application of local smoothing and wavelet analysis methods for studying the spectral composition and coherent relationships of local average annual surface air temperatures with solar activity and the displacement of the Earth's North Pole is presented. A preliminary analysis of local time series of surface temperatures revealed the presence of emissions and their localization. It is shown that to eliminate the influence of outliers (short-term events) on the reliability of identifying a long-term nonlinear trend, the wavelet decomposition method, which filters high frequencies, is most suitable. Functional approximation models are constructed and compared at different levels of wavelet decomposition of the data. Time or scale smoothing is used to improve the reliability of the wavelet spectrum. Based on data on average annual surface air temperatures in Yalta (44.48°, 34.17°, = 72.0 m) for the time interval from 1869 to 2022, functional models of long-term trends were built and used to obtain short-term forecasts. Information about the linear relationship of events in the compared time series is obtained and discussed in the analysis of wavelet cross-correlation, wavelet coherence and phase coherence. Local similarities were discovered between data on surface air temperature and solar activity data (Wolf numbers) for a period of ~(30–70) years, as well as oscillations with period of 11 years, manifested in the constancy of the phase difference and an increase in the modulus of wavelet coherence power over time. Localized similarities were also found in data on surface air temperature in Yalta and in data on displacements of the Earth's mean pole relative to the conventional beginning of EOP (IERS) CO1 in the interval of periods ~ (30–70) years.

1. Introduction

In the open system of interdependent spheres of Planet Earth, there are constant physical processes, which, due to the complexity of their manifestation in different time intervals, cannot be modeled with a sufficient degree of complexity. In this regard, it is extremely important to understand the instability of the geosphere, which is initiated by both external (cosmic) and distributed terrestrial influences. Many processes in the Earth system are monitored, and their spectral structure and relationships are analyzed using the time series they reproduce [1,2]. The origin and possibility of such changes in the communication time series reproduced by solar-terrestrial connections have not been sufficiently studied at present [3].

The analysis of time series generated by solar-terrestrial connections reveals a consistent change in the parameters of gravitational, electric and magnetic fields in different geospheres and allows, under certain conditions, to estimate the future value of the studied

* Corresponding author.
E-mail address: a.volvach@gmail.com (A. Volvach).

https://doi.org/10.1016/j.heliyon.2023.e23237
Received 10 August 2023; Received in revised form 28 November 2023; Accepted 29 November 2023
Available online 3 December 2023

connection parameter with the greatest possible accuracy. This estimate is a function only of the previous values of the analyzed time series [4–6].

In the process of analyzing the observed time series, an important point is the procedure for identifying deterministic behavior. A suitable mathematical description of the deterministic component allows not only describing the behavior of the studied time series but also predicting its dynamics [1].

The long-term processes discussed in the article have high energy intensity, which over a long period of time is often masked by rapid energy redistributions by high-frequency processes.

The localized spectral analysis methods we use are based on the mathematical apparatus of wavelet transform, which provides time-frequency scanning of a one-dimensional implementation and allows maintaining good resolution at different (high and low) frequencies [7]. Currently, this method is widely used in the field of analysis of the synchronization of oscillations of geocosmic time series [8–19].

At the same time, by synchronization we understand the property of time series generated by various dynamic systems, despite the difference in individual rhythms, to develop a single rhythm of coexistence even in the absence of cause-and-effect relationships. Time-stable synchronization of oscillations is observed in deterministic systems [20]. However, even in this case, changes in the oscillation parameters are possible under the influence of external factors and random disturbances. For example, the unpredictable effects of exogenous and endogenous factors on the kinematics and dynamics of the Earth create the need for constant support of the observational model of the movement of its pole (IERS). Due to the oscillation of the poles, the latitudes of all points on the earth's surface periodically change by several hundredths of an arc second, which in a linear measure corresponds to two to three tens of meters.

Latitude fluctuations affect the fluctuations of local and regional climatic indicators [21,22]. One such indicator, depending on the solar radiation incident on the Earth's surface, is the surface air temperature. Its value depends on the position of the Earth in orbit and the geographical latitude of the place. Establishing a link between the long-term change in the energy emitted by the Sun and the Earth's climate is the subject of active research at the present time. The results of these studies are closely related to the developing mathematical methods for analyzing time series of observations.

Wavelet methods, which have become popular since the 1990s, have made a significant contribution to the study of solar-terrestrial relationships and disturbances in the Earth system due to their unique capabilities in relation to data analysis. They are based on the decomposition of data into frequencies localized in time. Currently, there are a large number of publications demonstrating the versatile application of wavelets in the analysis of natural processes [23–27].

The results presented in this work allow us to conclude that it is possible to construct a short-term local forecast of the trend in surface temperature changes using wavelet decomposition of a priori data; update and refine the functional model coefficients as new observations become available. In contrast to smoothing methods (moving average, local regression), to eliminate the influence of short-term events on the long-term forecast, wavelet decomposition of oscillations into high-frequency (often short-term) and low-frequency ones is used.

Analysis of the results of wavelet coherence of surface air temperature in Yalta and data on changes in solar activity and the displacement of the Earth's North Pole reveals a general trend of a consistent increase in the power of local and global oscillations in the interval 1869–2022. However, a short-term forecast of surface air temperature in Yalta based on functional models indicates a decrease in the average annual surface temperature in the future. Of course, short-term energy events, since they are often unpredictable, locally influence temperature fluctuations.

2. Main stages and methods of analysis

- At the first stage of the analysis of the time series of observations of changes in surface air temperature in Yalta, the systematic and random components were separated. The calculations were performed using the wavelet data transformation method in the field of mathematical and software packages MATLAB. Statistical analysis of the data revealed gaps and outlier observations that could not be included in the random component. In order to preserve the information represented by outliers, a procedure was carried out to smooth the systematic component of the analyzed time series.
- The main purpose of the smoothing process was to calculate unbiased estimates of time series means to remove the influence of outliers on the regular pattern of surface air temperature. Based on this, we compared 3 smoothing methods: moving average, local smoothing using weighted linear regression, and wavelet decomposition of data at different levels of detail. A comparison of the smoothing results showed: with local smoothing using the weighted linear regression method, unlike the moving average smoothing process, the smoothing interval never changes, which preserves information at the edges of the time series. In addition, since outliers in the analyzed time series are localized in time, by calculating local regression we obtain unbiased means.

The wavelet decomposition method of data at various levels of detail made it possible to exclude high-frequency oscillations, which are not very informative for predicting a long-term trend.

To identify a nonlinear trend on a limited data interval, we used methods that maximally take into account the long-term systematic component, regardless of local fluctuations, which, as a result of statistical analysis (exclusion of noise), cannot be classified as random events.

- Using methods of calculation and analysis of wavelet cross-spectra and wavelet coherence [7], local similarities were established in data on surface air temperature and data on solar activity (Wolf numbers), as also in data on surface air temperature and data on the displacement of the Earth's North Pole.

The wavelet coherence results confirm the global conclusions about the synchronization of energy processes in the compared time series [2].

3. Outliers and local smoothing

3.1. Outlier detection

Outliers are data that are dramatically different from the rest of the time series. They may appear due to measurement errors or may represent significant features in the data caused by short-term external forces. In this paper, we used one of the common methods for detecting outliers - the analysis of time series members exceeding a certain number of standard deviations s from the mean m. The presence of outliers in the data is indicated by the histogram in Fig. 1.

Fig. 1 shows that some of the data are more than two standard deviations above the mean. Our testing of noise using the Kolmogorov-Smirnov test showed that at a significance level of 5%, the selected noise does not belong to the normal distribution (h = 1). However, the number of outliers and missing observations is small (see Fig. 1) and they are localized in time. Since the statistical properties of a non-stationary process do not change sharply over a short period of time, it can be assumed that the stationarity of the analyzed time series is justified locally [32].

To identify the main features of long-term changes in surface temperature, it is necessary to eliminate the influence of emissions on the regular regime of changes in surface air temperature. Our task is to isolate a long-term non-linear trend to analyze the long-term evolution of data. For this purpose, we use the most appropriate smoothing methods for the structure of the analyzed time series. Since the analyzed data contains outliers that are not white noise, we will perform the analysis using the smoothing method by calculating the local regression [26]. In our work, we used a robust version of this method [28,29]. It assigns less weight to outliers and defines 0 for data outside of 6 standard deviations. The essence of the method is as follows.

2.2. Local smoothing by weighted linear regression. Smoothing in this method is considered to be local, since, as in the moving average method [2], each smoothed value is determined by neighboring series members specified in the smoothing interval. The regression weight function is defined for all members of the series contained in the interval. The weights are determined by a cubic function of the form in Eq. (1) [27]:

$$w_i = \left(1 - \left|\frac{x - x_i}{d(x)}\right|^3\right)^3, \tag{1}$$

where x is the value of the smoothed member of the time series, x_i are the nearest neighbors x in the given interval, and dx is the distance along the abscissa from x to the farthest member of the time series in the smoothing interval. Weights have the following characteristics:

* the members of the series to be smoothed have the greatest weight and the greatest influence on the fit;
* row members outside the range have zero weight and do not affect the fit.

Next, a weighted linear regression using a first-degree polynomial is determined by the least squares method, which describes the smoothed values of the time series.

If the calculation includes the same number of adjacent series terms on either side of the smoothed series term, the weight function is symmetrical; if the number of neighboring samples is not symmetrical with respect to the smoothed term of the series, then the

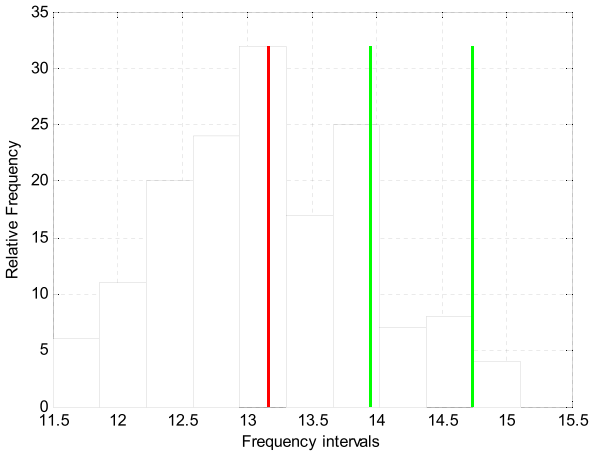


Fig. 1. Histogram of data on surface air temperature at the point with coordinates 44.48° N latitude, 34.17° E longitude, 72.0 altitude above sea level (Yalta). The figure shows: the mean (red line), standard deviations σ and 2 σ (green lines).

weight function is not symmetrical. Unlike the moving average smoothing process, the smoothing interval never changes, which preserves information at the edges of the time series.

If the data contains outliers, the smoothed values may be skewed and may not reflect the behavior of most adjacent series members. To overcome this problem, we used a robust procedure that is not affected by a small proportion of outliers. It consists of additional calculation for each member of the time series in the smoothing interval of reliable weights that are resistant to outliers (Eq. (2)). The weights are determined by the biquadratic function in the form [27]:

$$w_i = \begin{cases} \left(1 - (r_i/6MAD)^2\right), |r_i| < 6MAD \\ 0, \quad |r_i| \geq 6MAD \end{cases}, \tag{2}$$

where r_i are the remainders of the smoothing procedure described above; MAD is the median $|r_i|$.

The results of smoothing and restoring the continuity of data on surface air temperature in Yalta are presented in Fig. 2. The advantages of smoothing by the local regression method compared to the moving average method are manifested in unbiased local estimates of the smoothed term of the time series.

The data curve smoothed using the local regression method in Fig. 2 has a complex structure, including a linear trend and fluctuations around the trend, which can be modeled to predict their behavior in the future.

4. Modeling deviations from a linear trend

4.1. Linear trend model. The mathematical description of the linear trend was obtained by us in the form in Eq. (3)

$$f(x) = p_1 x + p_2, \tag{3}$$

where the coefficients (with 95% confidence limits) are:

$p_1 = 0.004568$ (0.002505, 0.006631) ℃/year,

$p_2 = 4.28$ (0.2662, 8.294) ℃.

4.2. Functional models of fluctuations relative to a linear trend

For the time series we are analyzing, a finite-dimensional description is allowed: a limited number of periodic oscillations describes the process of temperature change in the considered time interval [2]. Using mathematical functions often used for the analysis of natural processes, we will construct models of the best approximation. In the process of choosing the model of the best approximation, 5 functions were tested: polynomial, Fourier function, Gaussian, sum of sines, exponential. The selection criteria were not only estimates of the quality of the approximation, but also the values of the confidence intervals of the forecast. In particular, for the polynomial and Gaussian approximation models, an unacceptable extension of the confidence intervals for the forecast was found. Fig. 3 shows two best fit curves for locally smoothed weighted linear regression data with a removed linear trend, and the dotted lines show the confidence limits for the approximation and prediction of the analyzed data. The parameters that determine the type of functional models and estimates of the quality of the approximation are given in Table 1.

Table 1 shows in columns: 1 – model name; 2 - mathematical description of the approximation functions, 3 - colors of the graphs in Fig. 2; (4–7) are statistical estimates of the approximation quality, where.

* SSE is sum of squares due to approximation error. A value close to zero indicates fitness, which is more useful for prediction;
* R-sq. (R-square) is the square of the correlation between the original and simulated values. A value close to 1 indicates that most of the variance is accounted for by the model;
* Adj. R-square with adjustable degree of freedom. A value close to 1 indicates a better choice of the number of model parameters;

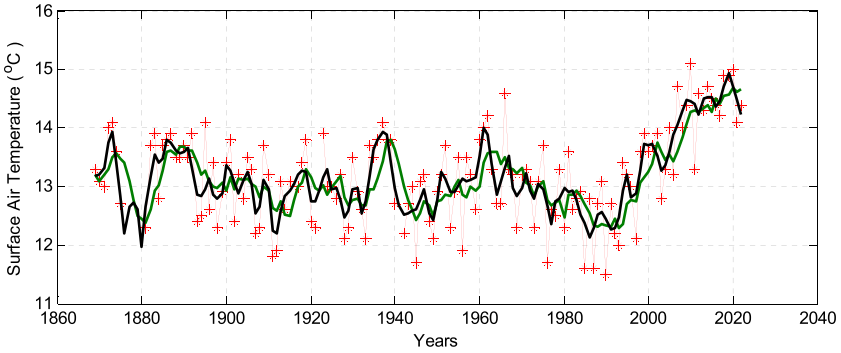


Fig. 2. Data graphs: original (red asterisks); smoothed by the 5-year moving average (green line) and local regression (black line).

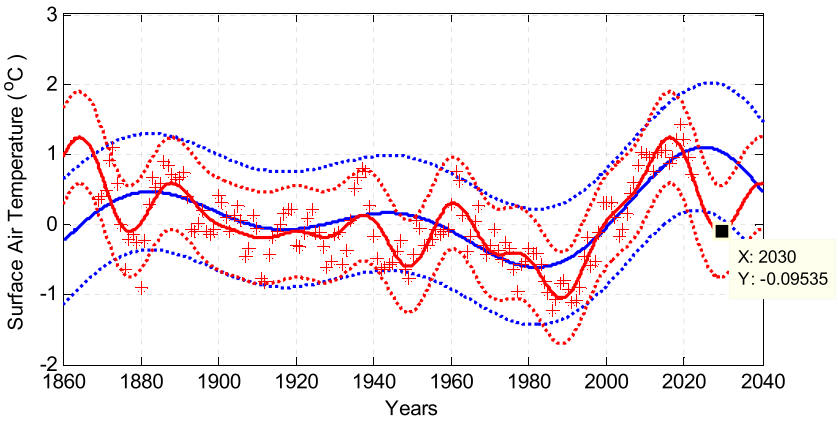


Fig. 3. Locally smoothed by weighted linear regression data on surface air temperature in Yalta and their functional models with 95% dashed confidence limits. Details of models in Table 1.

Table 1
Functional models of data on surface air temperature in Yalta.

Model name	Mathematical description	Color	SSE	R-sq.	Adj. R-sq	RMSE
"Fourier" n = 8	$a_0 + \sum_{i=1}^{n} \begin{pmatrix} a_i\, cos(n\omega x) \\ +b_i\, sin(n\omega x) \end{pmatrix}$	Red	13.4	0.73	0.70	0.31
"Sum sinuses" n = 2	$\sum_{i=1}^{n} a_i sin(b_i + c_i)$	Blue	25	0.50	0.48	0.41

* RMSE - root mean square error or standard error. A value close to 0 indicates the choice of the model that is most suitable for predicting the data.

Analysis of the modeling results shows that the Fourier model (n = 8) predicts a decrease in the average annual temperature in Yalta in 2030 by −0.095° C relative to the linear trend of the average annual temperature for the period 1869–2022 with a further oscillatory process of its change (Fig. 3).

5. Wavelet methods and wavelet models

We give a brief description of the wavelet methods used in this paper to detect the general behavior of frequency in jointly stationary time series.

5.1. Continuous wavelet transform

The continuous wavelet transform (CWT) computes the convolution of a signal using transformed and extended versions of the analysis wavelet. Comparing the signal with the wavelet at different scales and positions, we obtain a function of two variables. For the scale and position parameters, the wavelet transform is determined by the relation in Eq. (4) [22,29]:

$$C(a,b;f(t),\psi(t)) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)d\,t, \tag{4}$$

where the * sign denotes complex conjugation. By continuously changing the values of the scale parameter a and the position parameter b, we obtain the coefficients of the continuous wavelet transform $C(a,b)$.

The continuous wavelet transform (CWT) allows one to analyze the time evolution of the frequency content of a given signal or time series. Applying CWT to two time series and examining the two decompositions together can reveal local similarities in time and scale. Regions on the time-frequency plane where two time series exhibit common power or constant phase indicate local similarity between signals.

Continuous wavelet transform is comparable to bandpass filtering of the input signal. The CWT coefficients on lower scales represent the energy in the input signal at higher frequencies, while the CWT coefficients on higher scales represent the energy in the input signal at lower frequencies. However, unlike Fourier bandpass filtering, the width of a CWT bandpass filter is inversely proportional to scale. The width of the CWT filters decreases as you zoom in. This follows from the uncertainty relations between the time and frequency support of a signal: the wider the signal support bandwidth in time, the narrower its frequency support. The reverse

relationship also holds. More details about CWT can be found in [30,31].

In the wavelet transform, a scale or expansion operation is defined to conserve energy. Conserving energy while decreasing frequency support requires that the peak energy level increase. The quality factor or quality factor of a filter is the ratio of its peak energy to its bandwidth. Wavelets are often referred to as constant quality filters [7].

The reliability of the conclusions about the presence of one or another variation in the data depends on the wavelet power of this oscillation. Preliminary smoothing of the initial data highlights the long-term trends of the time series and contributes to the concentration of the wavelet power in the region of low-frequency oscillations. The CWT values in the figures (scalogram) are displayed in three dimensions - time on the abscissa axis, scale (converted to period or frequency) on the ordinate axis, and the power of the coefficients, estimated using a color gradient.

5.2. Cross-spectrum or otherwise joint wavelet spectrum of two time series

The wavelet spectrum defined for each signal is characterized by the magnitude and phase of the CWT obtained using the complex wavelet. Denote individual wavelet spectra as $C_x(a,b)$ and $C_y(a,b)$. The presence of a relationship between two signals in the time scale plane is indicated by the wavelet cross spectrum $C_{xy}(a,b)$ in Eq. (5), which is defined as [7,25]:

$$C_{xy}(a,b) = S\left(C_x^*(a,b)C_y(a,b)\right), \tag{5}$$

where $C_x(a,b)$ and $C_y(a,b)$ denote continuous wavelet transforms x and y in scales a and offsets b. The superscript * means the complex conjugate expression, and S the smoothing operator in time and scale. The magnitude of the wavelet joint spectrum can be interpreted as the absolute value of the local covariance between two time series on the time scale plane [7].

Wavelet coherence of two time sequences x and y is determined by the relation Eq. (6) [7]:

$$r_{x,y}(a,b) = \frac{S\left(C_x^*(a,b)C_y(a,b)\right)}{\sqrt{S\left(|C_x(a,b)|^2\right)}\sqrt{S\left(|C_y(a,b)|^2\right)}}, \tag{6}$$

where $C_x(a,b)$ and $C_y(a,b)$ denote continuous wavelet transforms x and y in scales a and positions b. The superscript * means the complex conjugate expression, and S the smoothing operator in time and scale. Wavelet coherence can be interpreted as the local square of the correlation coefficient on the time scale plane.

The coherence phase is calculated by the formula Eq. (7) [7]:

$$\theta = \tan^{-1}[\mathrm{I}\{W_n(s)\}/\mathrm{R}\{W_n(s)\}]\ . \tag{7}$$

These figures show information about the phase obtained when we used the complex Meyer wavelet ("Discrete" Meyer wavelet, DMeyer). The detected phase information is a measure of the phase difference between the corresponding frequency components in each signal [27].

The phase angles (7) in the signal have a uniform distribution and can vary widely between the values [-π, π]. Therefore, significance levels for the phase distribution cannot be established, and thus conclusions about the phase difference should be reduced to the interpretation of the phase distribution pattern [32].

The phase information is encoded by an arrow or an orientation vector. Phase information can be interpreted by finding some coherent manifestations in different parts of the time scale. In the case of incoherent behavior, an inconstancy of the phase difference is observed (Table 2) [33].

5.3. Wavelets used in this work

The choice of a wavelet is determined by the formulation of the problem of detecting the characteristics of a time series. We are interested in a smooth change in the characteristics of the time series, which boils down to searching for fluctuations with a smooth start and a shift. So far, there are no strict criteria for choosing a wavelet function. Recommendations are given in the literature for determining the best choice of wavelet based on a set of data for analysis, since the results of the transformation are very dependent on the wavelet used. Nevertheless, some wavelets have been repeatedly tested in the analysis of geophysical and climatic time series. Based on this, we used the following wavelets for wavelet analysis:

Table 2
Interpretations of the main phase arrows.

Arrow directions	Interpretation
Right →	Signals in phase
Left ←	Signals out of phase
Up ↑	y(t) leads x(t) by π/2 radians (or x(t) leads y(t) by 3π/2 radians)
Down ↓	x(t) leads y(t) by π/2 radians (or y(t) leads x(t) by 3π/2 radians)

* Wavelet from the Family Daubechies. The short name is 'db'N. Filter length 2 N. Refers to the minimum-phase filters. N is a strictly positive integer. Details about the characteristics of this wavelet are contained in [30].
* The complex Morlet wavelet (short name 'cmor') is defined as Eq. (8)

$$cmor(x) = (\pi f_b)^{-0.5} exp(2i\pi f_c x) exp\left(-x^2 / f_b\right), \tag{8}$$

where f_b is the bandwidth parameter and f_c is the center frequency of the wavelet. The values of the parameters f_b and f_c were established by us experimentally.

More detailed information is contained in [34].

* A wavelet from the family of discrete Meyer wavelets ("Discrete" Meyer wavelet, DMeyer) was used to smooth the data sequence. Short name 'dmey'. This wavelet is defined as an approximation of discrete Meyer wavelets based on a filter with a finite impulse response, FIR, (FIR, finite impulse response) - one of the types of linear digital filters. More details about this are given in [35].

6. Wavelet of the approximation model

Along with regular fluctuations in the data we analyze, there are local changes that are not identified as noise. Local smoothing by the method of weighted linear regression allows us to reduce their influence (but not eliminate) on the long-term process of temperature change that we establish (see Fig. 2). But even so, the functional data approximation model is quite complex and does not contain information about short-term transients and short extreme events. The wavelet analysis method makes it possible to distinguish the long-term part of the oscillatory process in data from short-term oscillations and determine their temporal localization.

Consider the successive stages of wavelet modeling.

6.1. Detecting gaps in data

In Fig. 4 shows graphs: wavelet decomposition of data into two levels $\mathbf{d_1}$, $\mathbf{d_2}$ (detail); approximation $\mathbf{a_2}$; analyzed data **s** in °C. We performed a wavelet decomposition using a wavelet from the Debuchy family (Daubechies wavelet family, short name 'db'2).

Having performed a continuous wavelet transform of the analyzed temperature time series, we were convinced of the presence and localization of discontinuities, as well as their influence on the wavelet power at different frequencies: from smaller scales (high frequencies) to large scales (low frequencies). The continuous wavelet transform graphs are shown in Fig. 5: the top one is the analyzed data, the bottom one is the distribution of wavelet power over frequencies. Below is a color indicator of the wavelet power, the numbers indicate the color with a confidence level value from 0 to 1 (one corresponds to the 95% confidence level). Judging by the color scheme, after ~2000 there is an increase in the wavelet power of oscillations (see Fig. 5). Moreover, the localization interval of

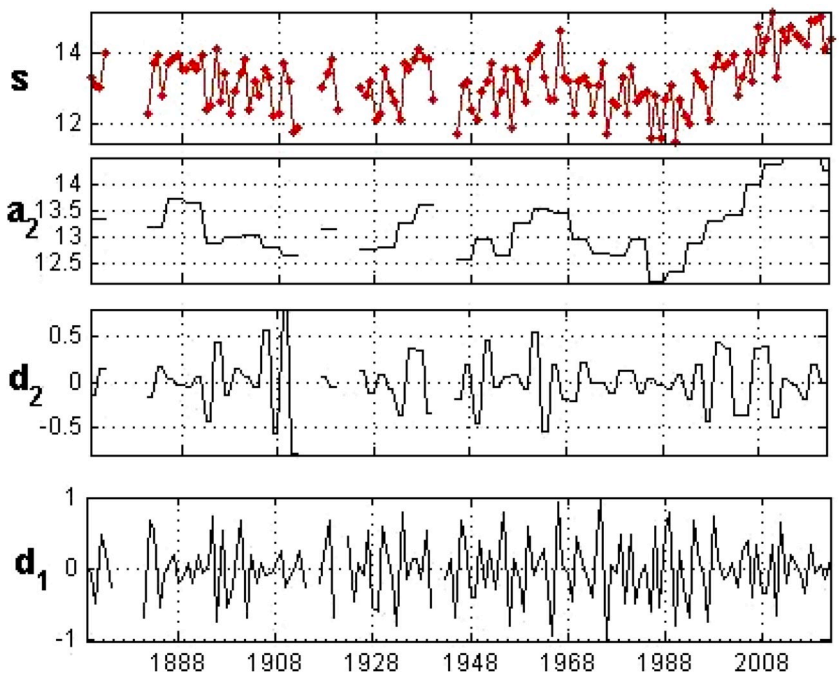


Fig. 4. Wavelet decomposition of the time series of surface air temperature in Yalta in order to detect missing observations and time intervals of their localization.

this phenomenon increases with a shift to the region of large scales (low frequencies). The pattern of continuous wavelet decomposition demonstrates the presence of a long-term trend and the expansion of the influence of missing observations on the wavelet power of low-frequency oscillations.

6.2. Long-term trend detection, functional models

Long-term evolution or trend is the slow part of the signal. Detecting evolution by a linear trend does not provide convincing conclusions for assessing the behavior of the temperature change process in the future, since with a limited duration of the time series, the linear trend may be part of a long cycle. Therefore, for forecasting, a functional model of long-term changes that is adequate to the available data is more preferable.

From the point of view of wavelet analysis, this corresponds to the value of the largest scale. As the scale increases, the resolution decreases, giving a more accurate estimate of the unknown long-term trend.

To highlight the long-term trend, we compare wavelet decompositions at different levels of detail. At the same time, a higher level of detail provides a smoother approximation curve that delineates high and low frequencies. This statement is demonstrated by the graphs in Fig. 6 and statistical estimates in Table 3.

The resulting wavelet models (see Fig. 6) can be represented as the most suitable functional Fourier model, as was done above (see Table 1). Information characterizing the form of the Fourier functional model at different levels of detail L and assessment of the quality of approximation is given in Table 3.

Table 3 contains information on columns: 1 – level of detail of the time series, 2 – order of the Fourier functional model, 3–6 – statistical estimates of the reliability of the models, 7 – "noise" excluded at various levels of decomposition L.

From the data presented in Table 3, it follows: functional models of wavelet decomposition at levels of detail L = 4 and L = 5 approximate the results of wavelet decomposition at these levels with high accuracy (SSE = 0; R-square end Adj. R- Square = 1; RMSE = 0), and the standard error of wavelet noise at these levels (see column 7) does not exceed the error of the original data). The smooth models we constructed (see Fig. 6) are resistant to the influence of local outliers compared to models obtained by smoothing using the moving average method and local regression (see Fig. 7). As the level of detail increases, the influence of the high-frequency component on the model disappears. Since the resolution is reduced in this case, the choice of the best level of detail depends on the predictive qualities of the model. For example, the Fourier model (L = 4, n = 6) allows for the possibility of a decrease in local temperature by −0.95 °C over the period from 2025 to 2040 (see Fig. 6).

7. Identification of localized similarities in data on surface air temperature and data on solar activity (wolf numbers)

The most commonly used index of solar activity is the relative number of sunspots or the Wolf number, determined for a given day by the formula Eq. (9):

$$W = k\,(10g + s), \tag{9}$$

where $\boldsymbol{W}$ is the Wolf number; $\boldsymbol{s}$ is the total number of sunspots regardless of their size on the visible hemisphere, $\boldsymbol{g}$ is the number of observed sunspot groups, $\boldsymbol{k}$ is a normalization coefficient that brings the values observed by different observers and telescopes to the standard chosen by Wolf.

Since 1981, a summary of all sunspot observations and the determination of Wolf numbers has been carried out at the World Data Center for the Observation, Preservation and Propagation of International Relative Sunspot Numbers (WDC-SILSO) of the Belgian Royal Observatory in Brussels. Data source used by us, Royal Belgian Observatory, Brussels (Yearly mean total sunspot number (1700-now): SN_y_tot_V2.0. csv).

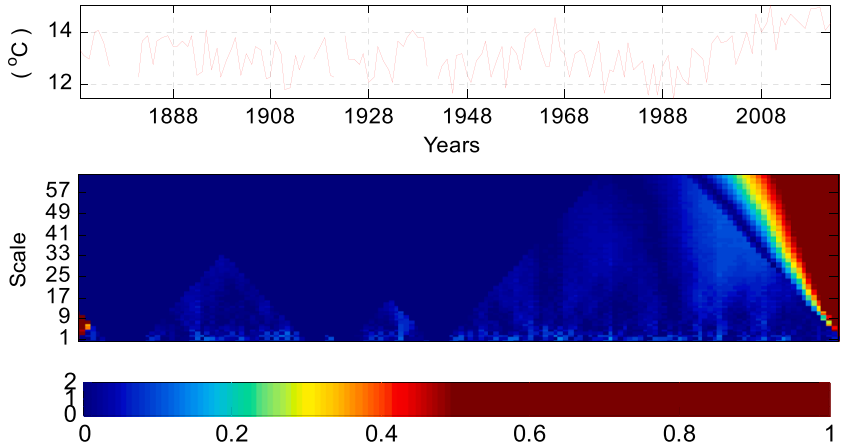


Fig. 5. Continuous wavelet transform of the average annual surface air temperature in Yalta.

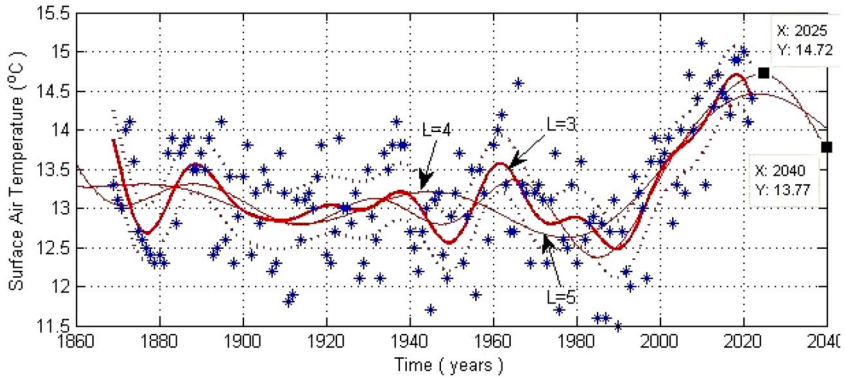


Fig. 6. Plots of average annual data on surface air temperature in Yalta (red asterisks) and their wavelet models for different levels of detail L. A wavelet from the Meyer family (waveletfamilies DMeyer, 'dmey') was used. Source of initial data: www.pogodaiklimat.ru.

Table 3
Functional models "Fourier" at different levels L of decomposition of the initial data.

L-level	Order of the Fourier model	SSE	R-sq.	Adj. R-sq	RMSE	Std "noise"
3	8	4.11	0.91	0.89	0.174	0.54
4	6	0.00	1	1	0.003	0.59
5	3	0.00	1	1	0.000	0.63

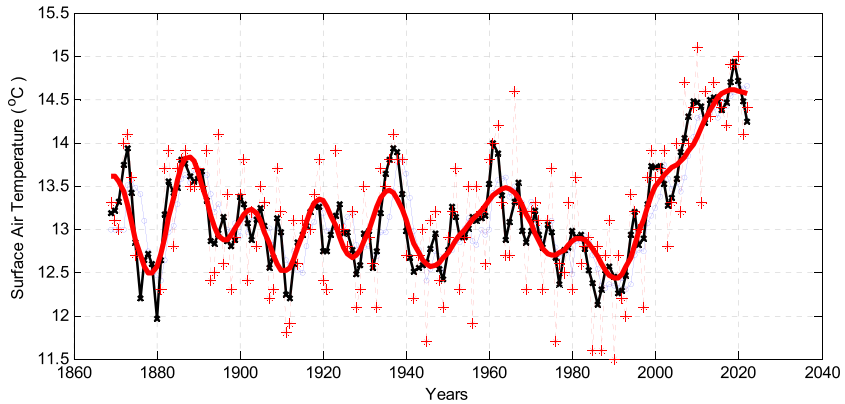


Fig. 7. Surface air temperature in Yalta: initial data on surface air temperature in Yalta (red asterisks) and their smoothed values using methods: 5-year moving average (blue circles), weighted linear regression (black crosses), Meyer family wavelets (waveletfamilies DMeyer, 'dmey'3, red curve). Source of initial data: www.pogodaiklimat.ru

Fig. 8 shows graphs of continuous wavelet transform (4) of initial data on surface air temperature and data on Wolf numbers with previously removed linear trends.

A clearly defined powerful cycle of 11 years is found in the decomposition of the solar data (right, lower figure). The wavelet power of this oscillation increases substantially after 1928. The power at the oscillation intervals of other periods is small, has no clear boundaries and increases with time. In the decomposition of the initial data on the surface air temperature in Yalta (Fig. 8, lower graph on the left), there are no clearly expressed fluctuations. However, diffuse energy bands are concentrated in the region of oscillations with periods of ≈40 years and ≈60 years. A characteristic feature of these oscillations is the continuous increase in power along the time axis and expansion along the period axis.

Despite the absence of a clear picture of the power distribution, the cross-spectrum in Fig. 9 contains information about a possible correlation between the two data series (5) in the interval of periods of 30.5–60.6 years and in the region of the 11-year cycle. The maximum wavelet power (95% confidence level) of the cross-spectrum of the 11-year oscillation falls on the interval (1928–1968) years.

More complete information about the linear relationship of events in the compared time series will be obtained by analyzing the wavelet coherence (6) and the coherence phase (7).

Fig. 10 reveals the constancy of the phase difference in the intervals of periods ≈ (30.5; 45.6) years and ≈ (53.1; 60.6) years and,

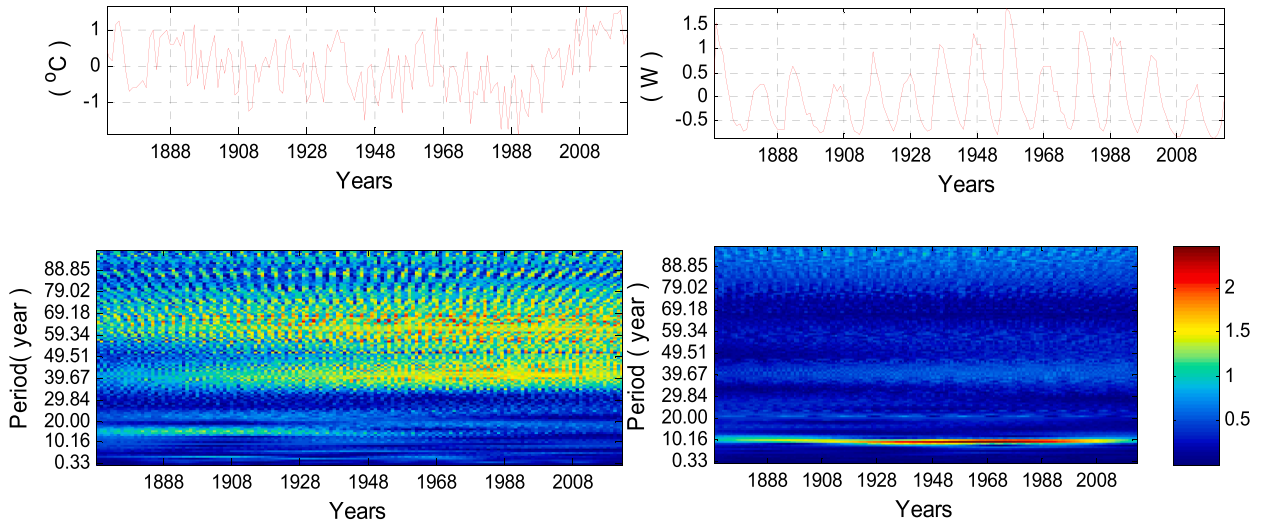


Fig. 8. Continuous wavelet transform (CWT, waveletfamilies Complex Morlet, 'cmor1-3′) of the initial data on the average annual surface air temperature in Yalta (a) and data on solar activity (b).

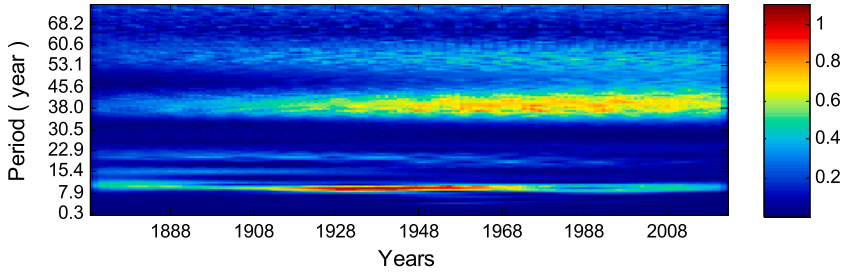


Fig. 9. Wavelet cross-spectrum (WCS, waveletfamilies Complex Morlet, 'cmor1-3') of the initial data on the average annual surface air temperature in Yalta and data on solar activity.

according to the color indication of the power module, a high power indicator with a confidence level of 95%. Fluctuations in these intervals of periods in the compared time series are consistent. A clearly pronounced coherence (the phase difference is stable, the power is at the level of 95% confidence) is observed for variations with a period of ≈22 years. However, after 2000 the phase difference of this oscillation is not constant.

Let us compare the obtained results with the wavelet decomposition pattern in the case of smoothed data on the average annual surface air temperature in Yalta.

Fig. 11 compare time series of data on the average annual surface air temperature in Yalta smoothed by the local regression method and data on solar activity (Wolff numbers). Comparison of the wavelet power distribution in Figs. 10 and 11 shows that the wavelet power of long-term (low-frequency) oscillations increases as a result of smoothing by the local regression method.

Comparison of the power of wavelet coherence in Figs. 11 and 12 shows the advantages of using wavelet smoothing compared to local regression (clearer selection of correlated frequencies and greater power). Fig. 12 clearly shows the suppression of high frequencies when smoothed by wavelets.

8. Identification of localized similarity in the data on the Earth's surface air temperature in yalta and in the data on the displacements of the Earth's mean pole relative to the conventional beginning EOP(IERS) CO1

For the analysis, we used data on the movement of the mean North Pole relative to the conditional start of CO1 for the period from 1900 to 2022 in the IERS definition.

EOP(IERS) C01 is a series of Earth orientation parameters given in the 1997 IERS system at 0.1 year intervals (1846–1889) and 0.05 year intervals (1890–present). Pole coordinates x end y are given in units seconds of arc.

They are regularly recalculated to take advantage of, on the one hand, improvements in the various individual contributions and, on the other hand, improvements in analysis procedures. This series is the basis of the IERS system.

Data plots are shown in Fig. 13.

The IERS models of the mean North Pole include the coefficients of the C21 and S21 tesseral harmonics in the definition of the

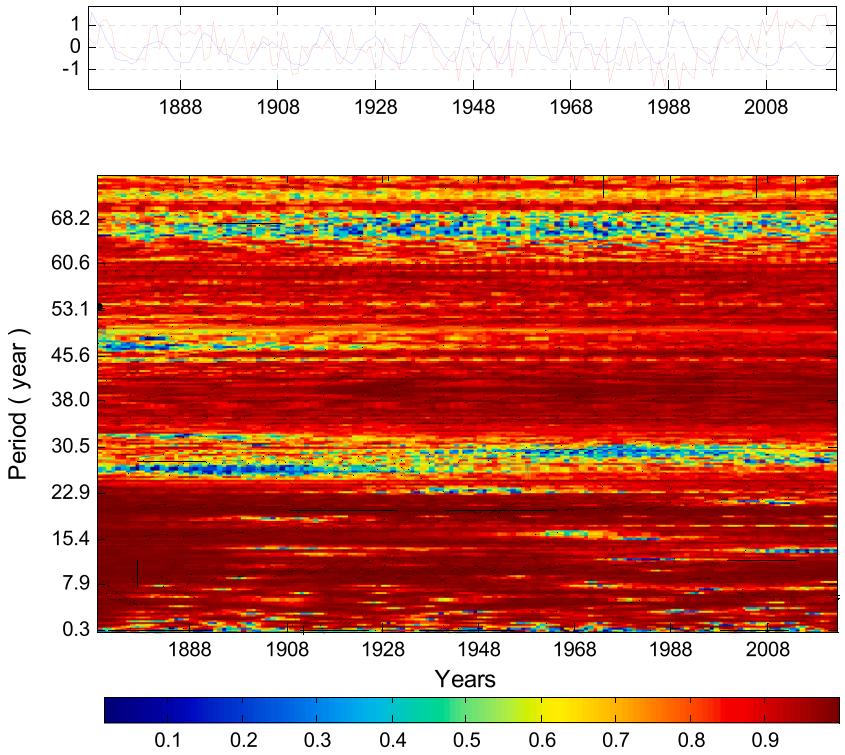


Fig. 10. Wavelet coherence of the initial data on the average annual surface air temperature in Yalta and data on solar activity. For calculations of wavelet coherence and phase difference, the complex Morlet wavelet (waveletfamilies Complex Morlet, 'cmor1-3') was used.

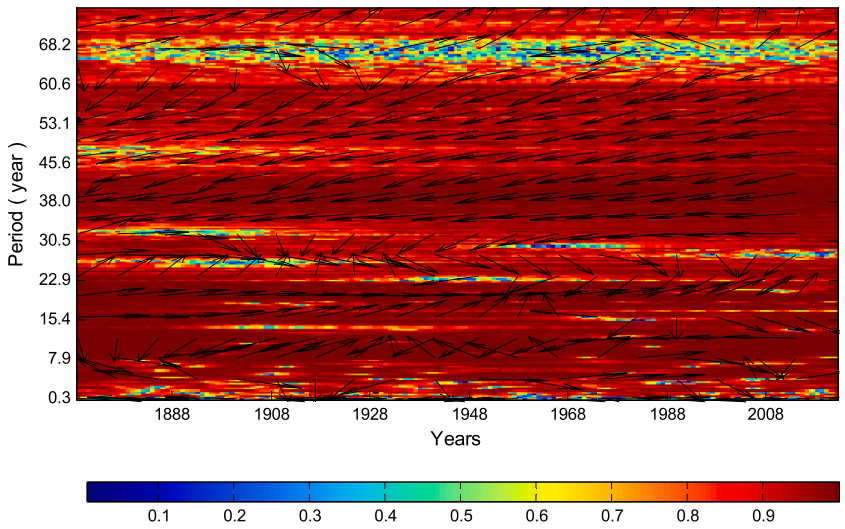


Fig. 11. Wavelet coherence of data on the average annual surface air temperature in Yalta smoothed by the method of local regression and data on solar activity. For calculations of wavelet coherence and phase difference, the complex Morlet wavelet (waveletfamilies Complex Morlet, 'cmor1-3') was used.

gravitational field. These parameters describe the position of the axis of the Earth's figure relative to the ground coordinate system: it must coincide with the observed position of the pole of rotation, averaged over the same period of time.

We use wavelet smoothing of the Earth's mean pole displacement data to eliminate the influence of high frequencies. Graphs of smoothed data are shown in Fig. 14.

Let us perform a continuous wavelet transformation of the initial data on surface air temperature in Yalta and data on the displacement of the average pole relative to the conditional origin CO1, smoothed by wavelets 'dmey'7.

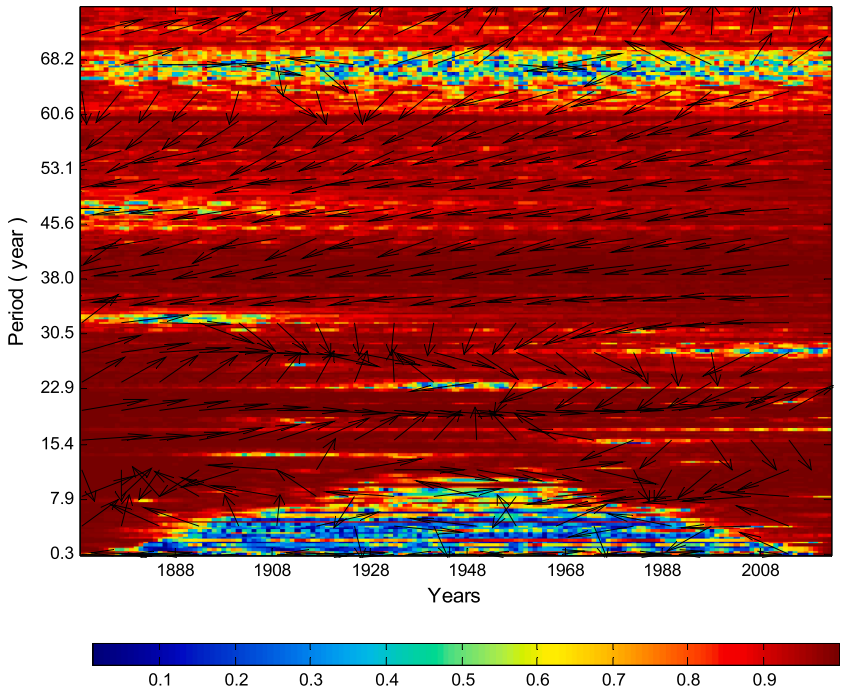


Fig. 12. Wavelet coherence of Yalta mean annual surface air temperature and solar activity data (Wolff numbers) smoothed by wavelets from the Meyer family (waveletfamilies DMeyer, 'dmey'3). For calculations of wavelet coherence and phase difference, the complex Morlet wavelet (waveletfamilies Complex Morlet, 'cmor1-3') was used.

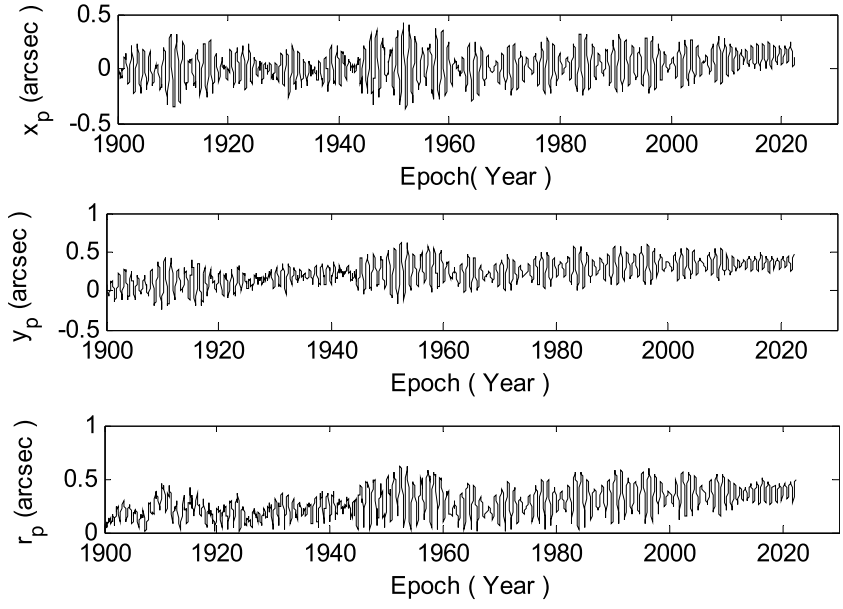


Fig. 13. Graphs of coordinates of the average pole of the Earth x_p, y_p and distances $r_p = \sqrt{x_p^2 + y_p^2}$ from the conditional origin CO1. Counting intervals are 0.05 years.

The wavelet decomposition picture of unsmoothed data on surface air temperature (Fig. 15) is not informative enough (the color indication does not correspond to the 95% confidence level). Therefore, we will continue the analysis using a model of smoothed data on surface air temperature.

The cross spectrum in Fig. 16 determines the region of spatiotemporal localization of the joint wavelet power over the interval of periods ≈ (34–53) years and the localization of the joint maximum (95% confidence probability) wavelet power over the time interval

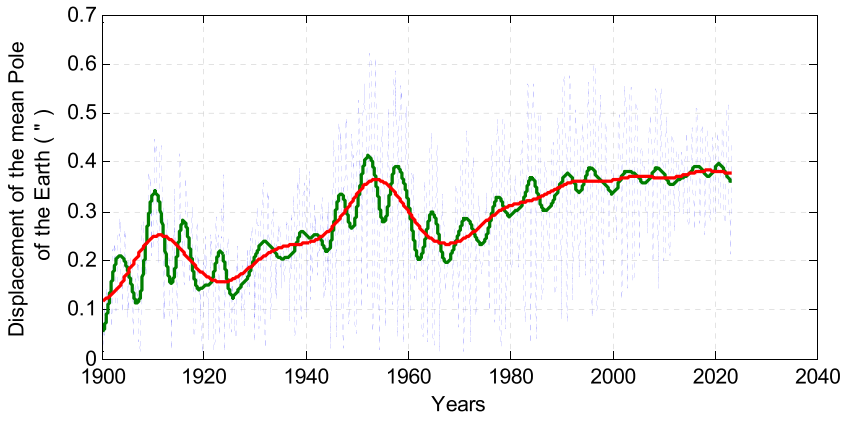


Fig. 14. Plots of distances r_p of the mean North Pole of IERS relative to the conditional origin (dashed line) and models of these data smoothed by wavelets from the Meyer family (waveletfamilies DMeyer, 'dmey'). Models were obtained at data decomposition levels L = 5 (green) and L = 7 (red).

from 1960 to 2000.

A more complete picture of the compared time series is represented by wavelet coherence (Fig. 17).

In Fig. 17 areas of high coherence with a constant phase difference are highlighted. The directions of the phase arrows indicate that the phases of oscillations with periods of ≈36, 64, 70 years coincide in the time interval for determining the local air temperature in Yalta and the displacement of the Earth's pole.

9. Summary

1. Statistical analysis of the time series of surface temperature data at Yalta reveals outliers and a difference in noise distribution from normal. The number of outliers and missing observations is insignificant (see Fig. 1) and they are localized in time, so it is customary to consider the analyzed time series to be locally stationary.
2. The advantages of smoothing data with outliers using the local regression method compared to the moving average method are manifested in unbiased local estimates of the smoothed members of the time series (see Fig. 2).
3. Analysis of the modeling results of data on surface air temperature in Yalta smoothed by local regression shows that the Fourier model (Table 1) predicts a decrease in the average annual temperature in 2030 by −0.095ºC relative to the linear trend of the average annual temperature for the period 1869–2022 with a further oscillatory process of its change (see Fig. 3).
4. Local smoothing by the method of weighted linear regression allows us to reduce the influence (but not eliminate) of emissions and short-term high-frequency fluctuations on the long-term process of temperature change that we establish. The wavelet analysis method makes it possible to distinguish the long-term part of the oscillatory process in data from short-term oscillations and

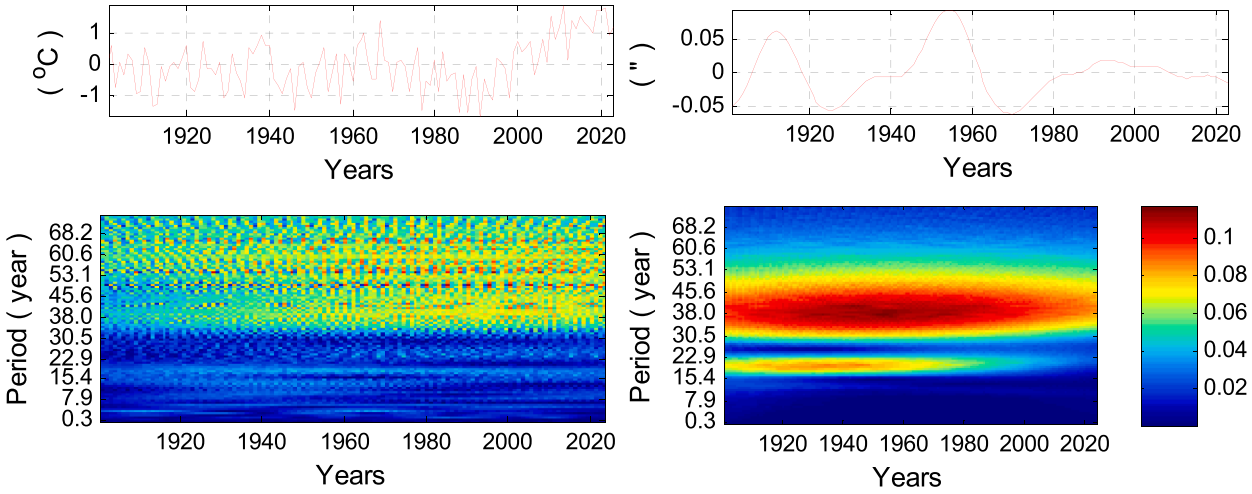


Fig. 15. Continuous wavelet transform (CWT, waveletfamilies Complex Morlet, 'cmor1-3') of two time series (linear trends removed): initial data on surface air temperature in Yalta (a) and data smoothed by wavelets from the Meyer family (waveletfamilies DMeyer, 'dmey'7) on the shift r_p of the mean pole relative to the conditional beginning (b).

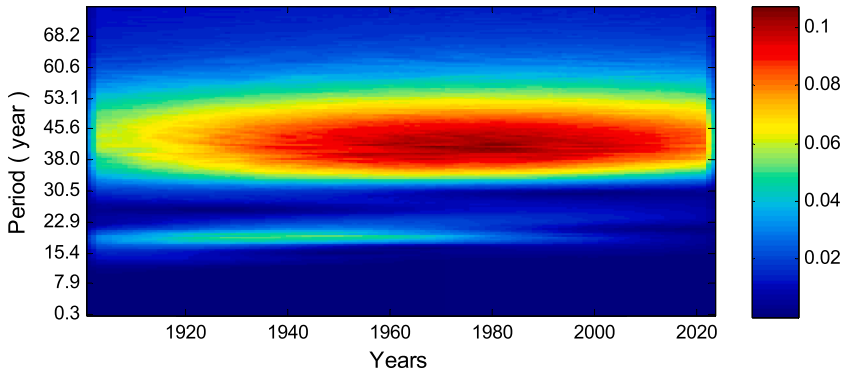


Fig. 16. Wavelet cross-spectrum of data on surface temperature in Yalta smoothed by wavelets from the Meyer family (waveletfamilies DMeyer, 'dmey'3) and data on the shift r_p of the mean pole relative to the conditional origin smoothed by wavelets from the Meyer family (waveletfamilies DMeyer, 'dmey'7). (WCS, waveletfamilies Complex Morlet, 'cmor1-3').

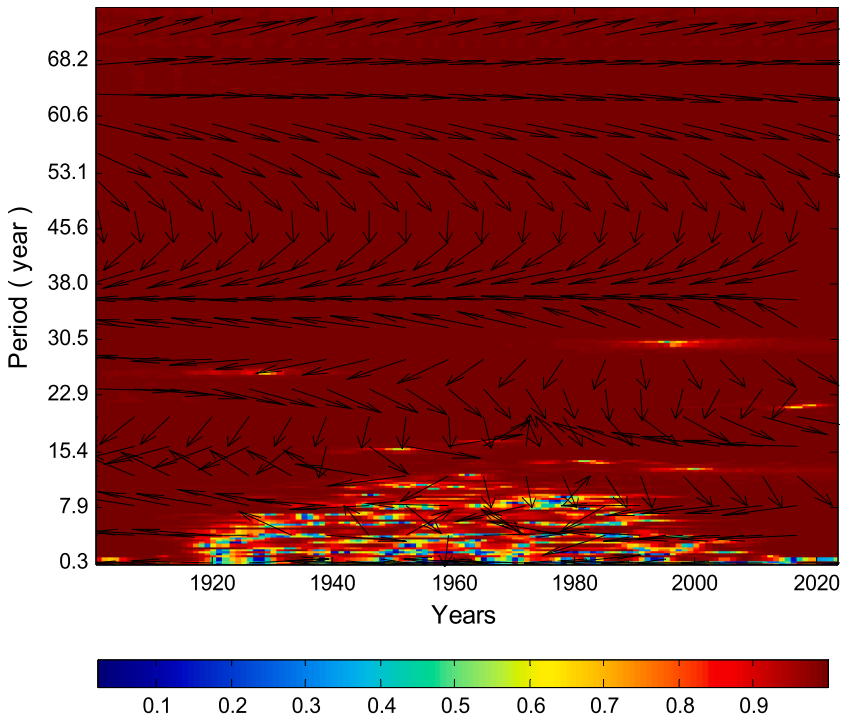


Fig. 17. Wavelet coherence of wavelet-smoothed data from the Meyer family (waveletfamilies DMeyer, 'dmey'3) on the surface temperature in Yalta and data smoothed by wavelets from the Meyer family (waveletfamilies DMeyer, 'dmey'7) on the shift of the mean pole relative to the conditional origin. (COHER, waveletfamilies Complex Morlet, 'cmor1-3').

determine their temporal localization. At the same time, a higher level of wavelet decomposition provides a smoother approximation curve, delimiting high and low frequencies (see Fig. 6; Table 2).

5. The wavelet coherence modulus (Fig. 12) of the data on the average annual surface air temperature in Yalta smoothed by wavelets from the Meyer family (L = 3) and data on solar activity allows for the possibility of local consistency of oscillations with a period of 11 years and in the interval of periods ≈ (30.5–60.6) years.
6. Localized similarity in the data on the surface air temperature in Yalta and in the data on the displacements of the Earth's mean pole relative to the conditional beginning is observed (see Figs. 16 and 17) in the interval of periods ≈ (30–70) years.

The formulation of the problem of our analysis of time series of surface temperatures in Yalta included the assumption that the analyzed time series has a long-term trend of change and can be represented as a sum of systematic, or trend, and random components [36]. Short-term periods, fluctuations and irregularities are included in the systematic component and thus cause local deviations of the long-term trend from the general trend of change.

To estimate the nonlinear long-term trend with the smallest error, a comparison was made of the data smoothing method with a 5-year moving average, local smoothing using the weighted linear regression method and wavelet decomposition.

Comparison of the smoothing results by various methods revealed the advantage of the wavelet data decomposition method, which made it possible to filter out emissions and high-frequency components.

The wavelet coherence modulus (Fig. 12) of the data on the average annual surface air temperature in Yalta smoothed by wavelets from the Meyer family (L = 3) and data on solar activity allow for the possibility of local consistency of oscillations with a period of ≈11 years and in the interval of periods ≈ (30.5–70) years.

The localized similarity of time series (see Figs. 12 and 17) in the interval of periods ≈ (30–70) years, manifested in a simultaneous increase in the power of wavelets and the coherence module, indicates the participation of the observed local process of increasing surface temperature in Yalta in the coordinated process of synchronization of natural cycles.

7. The work does not discuss the problem of global warming, therefore, the separation of the contribution of various external and internal influences on the temperature at a given local point is not considered. It is not possible to determine the influence of each external influence on a long-term nonlinear trend on such a short time interval [2].

10. Conclusion

The analysis of data on surface air temperature and the forecast of long-term trends in its change presented in the work are necessary for solving fundamental problems of climate study and planning of economic and environmental activities.

In local data on changes in surface air temperature, along with the general trend, deviations associated with external influences and local characteristics are observed.

The formulation of the problem of our analysis of time series of surface temperatures in Yalta included the assumption that the analyzed time series has a local trend of change and can be represented as the sum of systematic, or trend, and random components [36]. Short-term emissions, high-frequency fluctuations are included in the systematic component and, thus, cause local deviations in the long-term trend.

To estimate the nonlinear long-term trend with the smallest error, a comparison was made of smoothing methods: 5-year moving average, weighted linear regression and wavelet decomposition.

A comparison of the smoothing results revealed the advantage of the wavelet decomposition method, which made it possible to filter out outliers and high-frequency components.

The degree of consistency of fluctuations in data on surface temperature in Yalta in the interval 1869–2022 with changes in solar activity and displacements of the North Pole of the Earth demonstrates wavelet coherence in Figs. 12 and 17.

The problem of analyzing extreme events cannot be considered in this work, since its solution requires nonlinear methods of time series analysis, the use of which on such a relatively short time interval and the presence of a limited number of outliers (local stationarity) does not lead to a reliable result.

Due to local redistribution of energy in the surface air (resonances, various extreme events), the coefficients of the numerical model of the long-term trend must be updated taking into account observations obtained in the future.

Therefore, the goal of our further work is to clarify the functional models of long-term trends and substantiate the reliability of the forecast of long-term trends, taking into account future observations.

The conclusions about the coherence of oscillations presented in the work are more qualitative than quantitative. With new developments in wavelet coherence software, this shortcoming may be eliminated in the future.

Data availability statement

Data is available through the electronic resource: http://www.pogodaiklimat.ru/history/33990.htm. Location of the meteorological station in Yalta (Crimea): latitude 44.48°, longitude 34.17°, altitude 72 m.

CRediT authorship contribution statement

Alexandr Volvach: Writing - review & editing, Writing - original draft, Formal analysis, Data curation, Conceptualization. **Galina Kurbasova:** Writing - original draft, Software, Formal analysis, Data curation. **Larisa Volvach:** Writing - original draft, Visualization, Resources, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977, p. 689.
[2] M. Kendall, A. Stewart, Multivariate Statistical Analysis and Time Series, Nauka Publ., Moscow, 1976, p. 736.

[3] B.L. Bery, Synchronous Processes in the Shells of the Earth and Their Cosmic Causes, vol. 5, Bulletin of Moscow State University, 1991, pp. 20–27.
[4] A. Loskutov, I.A. Istomin, K.M. Kuzanyan, O.L. Kotlyarov, Testing and forecasting the time series of the solar activity by singular spectrum analysis, Nonlinear Phenom. Complex Syst. 4 (1) (2001) 47–57.
[5] I.A. Istomin, O.L. Kotlyarov, A.Yu Loskutov, The problem of processing time series: extending possibilities of the local approximation method using singular spectrum analysis, Theor. Math. Phys. 142 (1) (2005) 128–137.
[6] A.J. Loskutov, A.A. Kozlov, J.M. Hahanov, Entropy and forecasting of time series in the theory of dynamical systems, Izvestiya VUZ, Applied Nonlinear Dynamics 17 (4) (2009) 98–113, https://doi.org/10.18500/0869-6632-2009-17-4-98-113.
[7] C. Torrence, G.P. Compo, A practical guide to wavelet analysis, Bull. Am. Meteorol. Soc. 79 (1998) 61–78.
[8] D.V. Divine, F. Godtliebsen, Bayesian modeling and significant features exploration in wavelet power spectra, Nonlinear Process Geophys. 14 (2007) 79–88.
[9] M.A. El-Borie, A.M. El-Taher, A.A. Thabet, A.A. Bishara, The dependence of solar, plasma, and geomagnetic parameters' oscillations on the heliospheric magnetic field polarities: wavelet analysis, Astrophys. J. 880 (2019) 86, https://doi.org/10.3847/1538-4357/ab12d8.
[10] A.E. Volvach, G.S. Kurbasova, L.N. Volvach, The "atmosphere" model: analysis of the time series of updates to the deformations of the earth surface, Astrophysical Bulletin 73 (4) (2018) 487–493, https://doi.org/10.1134/S1990341318040120.
[11] A.E. Volvach, G.S. Kurbasova, Secular variations of geomagnetic declination in the Karadag point and the global helio-geodynamic processes, Geofizicheskiy Zhurnal-Geophysical Journal 41 (2019) 192–199.
[12] A.E. Volvach, G.S. Kurbasova, L.N. Volvach, Solar-terrestrial cycles in the climatic and geophysical properties of crimea, Astrophysical Bulletin 74 (3) (2019) 331–336, https://doi.org/10.1134/S1990341319030118.
[13] A. Volvach, G. Kurbasova, Model of Insolation of the Earth Surface in the Kara-Dag Locality According to SSE Data, vol. 85, Visnyk of Taras Shevchenko National University of Kyiv: Geology, 2019, pp. 51–58, https://doi.org/10.17721/1728-2713.85.07.
[14] A.E. Volvach, G.S. Kurbasova, L.N. Volvach, Analyis of periodical variability of insolation and soil temperature in the Crimea, Geofizicheskiy Zhurnal-Geophysical Journal 41 (2019) 195–202, https://doi.org/10.24028/gzh.0203-3100.v41i6.2019.190076.
[15] A.E. Volvach, G.S. Kurbasova, Seasonal oscillations in local deformations and insolation of Earth's surface, Cosmic Res. 58 (2) (2020) 79–85, https://doi.org/10.1134/S0010952520020094.
[16] A.E. Volvach, G.S. Kurbasova, L.N. Volvach, A.V. Ipatov, Features of the motion of the Earth's geographic North Pole and jumping in the geomagnetic field, Cosmic Res. 60 (4) (2022) 282–291, https://doi.org/10.1134/S0010952522040086.
[17] A. Volvach, G. Kurbasova, L. Volvach, Time series analysis of temperatures and insolation of the Earth's surface at Kara-Dag using satellite observation, Adv. Space Res. 69 (12) (2022) 4228–4239, https://doi.org/10.1016/j.asr.2022.04.016.
[18] A.E. Volvach, G.S. Kurbasova, L.N. Volvach, Analysis and numerical simulation of temperature measurements made on Earth and from Space, Heliyon 9 (2) (2023), e12999, https://doi.org/10.1016/j.heliyon.2023.e12999.
[19] C. Katsavrias, C. Papadimitriou, A. Hillaris, G. Balasis, Application of wavelet methods in the investigation of geospace disturbances, Atmosphere 13 (3) (2022) 499, https://doi.org/10.3390/atmos13030499.
[20] V. Migulin, V. Medvedev, E. Mustel, V. Parygin, Basic Theory of Oscillations, first ed., MIR, 1983, p. 400.
[21] Ebrahim Ghaderpour, Paolo Mazzanti, Gabriele Scarascia Mugnozza, Francesca Bozzano, Coherency and phase delay analyses between land cover and climate across Italy via the least-squares wavelet software, Int. J. Appl. Earth Obs. Geoinf. 118 (2023), 103241, https://doi.org/10.1016/j.jag.2023.103241. ISSN 1569-8432.
[22] J. Yan, L. Wang, H. He, D. Liang, W. Song, W. Han, Large-area land-cover changes monitoring with time-series remote sensing images using transferable deep models, IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–17, https://doi.org/10.1109/TGRS.2022.3160617. Art no. 4409917.
[23] I. Daubechies, Ten Lectures on Wavelets, CSBM-NSF Series Appl. Math., vol. 61, SIAM Publication, 1992, p. 357.
[24] N.M. Astaf'eva, Wavelet analysis: basic theory and some applications, Phys. Usp. 39 (1996) 1085–1108.
[25] M. Farge, Wavelet transforms and their applications to turbulence, Annu. Rev. Fluid Mech. 24 (1992) 395–457.
[26] A.C. Furon, C. Wagner-Riddle, C.R. Smith, J.S. Warland, Wavelet analysis of wintertime and spring thaw CO2 and N2O fluxes from agricultural fields, Agric. For. Meteorol. 148 (2008) 1305–1317.
[27] A. Grinsted, J.C. Moore, S. Jevrejeva, Application of the cross wavelet transform and wavelet coherence to geophysical timeseries, Nonlinear Process Geophys. 11 (2004) 561–566.
[28] W.S. Cleveland, S.J. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, J. Am. Stat. Assoc. 83 (1988) 596–610.
[29] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Stat. Assoc. 74 (1979) 829–836.
[30] I. Daubechies, Ten lectures on wavelets, CSBM–NSF Series Application Mathematics 61 (1994) 194–202.
[31] S.A. Mallat, Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.
[32] Z. Ge, Significance tests for the wavelet cross spectrum and wavelet linear coherence, Ann. Geophys. 26 (2008) 3819–3829.
[33] C. Fu, A. James, M.P. Wachowiak, Analyzing the combined influence of solar activity and El Nino on streamflow across southern Canada, Water Resour. Res. 48 (2012), W05507, https://doi.org/10.1029/2011WR011507.
[34] A. Teolis, Computational Signal Processing with Wavelets, Birkhauser, 1998, p. 65.
[35] P. Abry, Ondelettes et turbulence, Diderot ed., 1997, p. 268. Paris.
[36] T.W. Anderson, The Statistical Analysis of Time Series, John Wiley & Sons, Inc., 1994, p. 704, https://doi.org/10.1002/9781118186428.