Forecasting of climatic characteristics by the wavelet transform method

S.Yu. Zolotov, I.I. Ippolitov, M.V. Kabanov, and S.V. Loginov

Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch of the Russian Academy of Sciences, Tomsk

Received October 5, 2004

A possibility to forecast the climatic characteristics with the use of wavelet transform is described. The wavelet transform of the time series of climatic characteristics reveals the quasi-periodicity for different time scales. The forecast of the sought time series is constructed based on the forecast of much simpler quasi-periodic functions of wavelet coefficients.

The analysis of accumulated observational series of some parameters of the regional climatic system shows that annually mean values of these characteristics demonstrate some regular changes.[1,2] Along with conclusions following herefrom on the complex monitoring of the observed climatic changes, the obtained results are also useful for predicting these changes. Such a forecast based on the analysis of empirical evidences can be referred to as formal and not claiming for consideration of physical mechanisms of the observed changes. Nevertheless, with a properly chosen method of statistical analysis, we may expect a reliable prediction of evolution of the observed changes for the nearest years.

In the time series, from the viewpoint of forecast, latent quasiperiodic variations on different scales, as well as a long-term trend are informative. If after removal of the trend from the original series, the remainder meets stationarity, then in the forecast (forward extrapolation) we may use a mixed model of autoregression and moving average (ARIMA).[3] In this model, the time sequence element $X(t)$ is expressed through the previously known elements of the same sequence:

$$X(t) = \beta_1 X(t-1) + \ldots$$

$$\ldots + \beta_p X(t-p) + \varepsilon(t) + \alpha_1 \varepsilon(t-1) + \ldots + \alpha_q \varepsilon(t-q), \quad (1)$$

where p and q refer to the model order; $\beta_1, \ldots, \beta_p$, $\alpha_1, \ldots, \alpha_q$ are the ARIMA model parameters; $\varepsilon(t)$ is the white noise. The forecast peculiarity here is the following: in practice, the model order is not set large (not exceeding ten) and therefore, the information on the variation structure of the whole series, especially the series, which is a system of nonlinear periodicities, is poorly taken into account.

The studies of a variation structure often invoke the Fourier transform, the mathematical tool useful for the frequency signal analysis, but ineffective in the processing of complex signals. In particular, the Fourier transform fails to analyze local properties of signals, because the basic functions of the Fourier transform are defined throughout the all time axis. Nevertheless, there exist procedures that allow the forecast by calculating the sum of the harmonic constituent series of individual components of narrow peaks in the Fourier spectrum.[4]

Here we consider a new approach to the forecasting connected with the use of wavelet transform of observational time series. The wavelet transform of a signal[5] is the signal resolution in the basis formed by a special function (wavelet) with certain properties by means of scale variations and transfers. Thus, one-dimensional signal is projected onto a time–frequency plane in the form of a two-dimensional distribution of coefficients of a continuous wavelet transform

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi_{ab}^{*}(t) \mathrm{d}t, \quad (2)$$

where $W(a, b)$ are the wavelet transform coefficients, $f(t)$ is the function of interest, t is the time, $\psi_{ab}(t)$ is the wavelet, a stands for the scale magnitude; b is the shift on the time axis; and asterisk denotes the complex conjugation.

The transform inverse to Eq. (2) looks like

$$f(t) = C_{\psi}^{-1} \int_{-\infty}^{\infty} \frac{\mathrm{d}a}{a^2} \int_{-\infty}^{\infty} W(a,b) \psi_{ab}(t) \mathrm{d}b, \quad (3)$$

where C_{ψ} is the normalizing coefficient.

Equations (2) and (3) for a discrete case are given in Ref. 6. Wavelet transforms of temperature series[7] and some geophysical indices[8] have shown that the distribution $W(a, b)$ is characterized by quasi-periodical structures of different time scales. The idea of the suggested approach consists in the following. Prediction is made for simple smooth quasi-periodical functions $\varphi_a(b) = W(a, b)$ on different scales a, rather than for the original function $f(t)$, and then the latter is restored using Eq. (3). For each function $\varphi_a(b)$, we fix local minima and maxima. The extrapolation procedure for $\varphi_a(b)$ consists in determining both the time position of a next maximum/minimum (as an

0235-6880/05/04 314-03 $02.00

average interval for a given series) and its numerical value (by using the linear polynomial). The values of the wavelet coefficients for the range between the maximum and minimum are filled in with the help of a cubic spline. Confidence intervals are evaluated by calculating the upper and lower quantiles of the empirical function of remainder series distribution.

Values for the scale a were chosen according to Ref. 6, that is, on the basis of the quantization procedure by the power of two:

$$a_j = a_0 2^{j\delta j}, \quad j = 0,1,\ldots,L,$$

where a_0 is the smallest possible scale ($a_0 = 2\delta t$, δt is the interval between two neighbor time points in the original time series), δj stands for the accuracy of discrete realization of the continuous wavelet transform (in these examples it was 0.1).

The first step in implementing this approach is a choice of the wavelet type. The analysis of temperature series in Ref. 7 and 8 shows that to educe the scales of oscillations and temporal localization of extremes in the series, it is reasonable to choose the Morlet wavelet.

In choosing the wavelet form, we used $f(t) = \sin(t)$, $t \in [0; 0.1\pi; \ldots; 18\pi]$ as a test function. For this function, the wavelet coefficients of the Morlet wavelet have a clear quasi-periodical structure with some distortions at the beginning and the end of the considered time interval. These distortions are connected with the time boundedness of the original series. In case a biorthogonal wavelet[9] is applied, the distortions at the ends are reduced significantly, however, some diffusiveness on the scales greater than 2π increases. Further, using the values of the wavelet transform coefficients, we made a forecast into the region $t \in [18.1\pi; 18.2\pi; \ldots; 20\pi]$, and reconstruct the original function by Eq. (3). Analysis of the forecast, performed with the considered wavelets, shows that at comparable estimates of reconstructing of the test harmonic function on the interval $t \in [0; 0.1\pi; \ldots; 18\pi]$, the biorthogonal wavelet gives a lower prediction error.

Taking into account this fact, we have forecasted two actual observational series. The first series included the annual mean Wolf numbers for the period from 1700 to 2003. This series characterizes solar activity and, to some extent, solar–terrestrial relationship, and shows on average an 11-year regular alteration of maxima (minima). As the initial series, we have chosen a Wolf number series for the period from 1700 to 1978 and forecasted it for 1979 to 2003. The result is shown in Fig. 1*a*, where a solid curve refers to observations, a dashed curve is the forecast made using a biorthogonal wavelet, and a dotted curve is the ARIMA-model forecast. Vertical straight lines refer to 90% confidence intervals for the forecast by the wavelet transform. It follows from Fig. 1*a* that the forecast made by the wavelet transform gives much more accurate results than the ARIMA forecast.

Figure 1*b* illustrates the forecast of the Wolf numbers for the period 2004–2015 using the biorthogonal wavelet (dashed curve). The 90% confidence interval does not exceed 15 Wolf units. For comparison, we present a forecast[10] for 2004–2007 as a dotted curve, obtained by McNish and Lincoln method[11] with allowance for regression coefficients and average Wolf numbers for one cycle. The original data for the forecast were the Wolf numbers for cycles from 8th to 20th. The 90% confidence interval did not exceed 20 Wolf units.

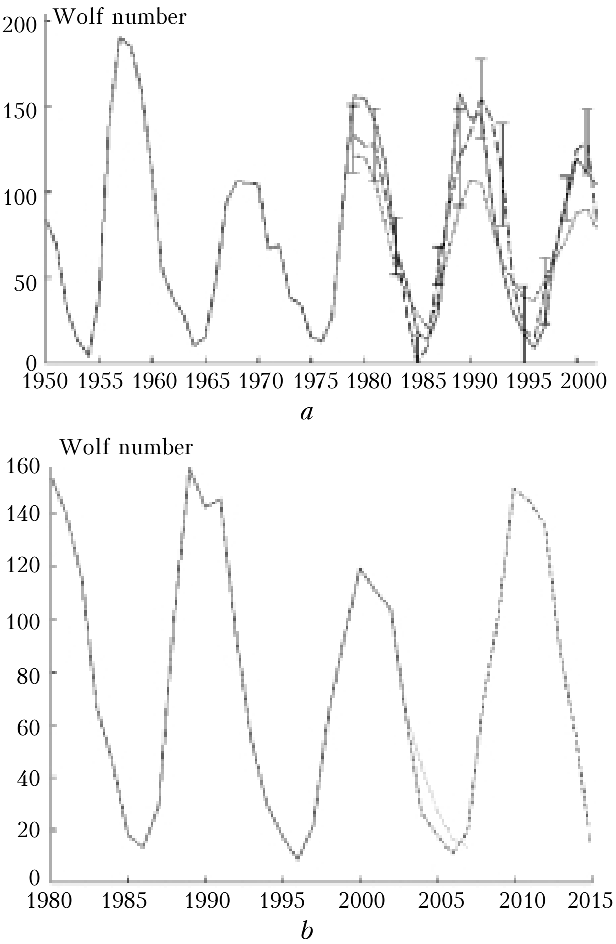

Fig. 1. Test of the Wolf number forecast (1979–2003).

Figure 2 shows the forecast of monthly mean values of the ground pressure for the Omsk meteostation for the period from January 2002 to December 2004 with a 90% confidence interval not exceeding 5 millibar. This forecast interval was based on the available data.

Another object of investigation was a series of annual mean temperatures for Omsk for the period 1916–2002. Taking into account that the last decades in Siberia were characterized by a significant positive temperature trend, we discriminated from the original series the trend represented by a cubic polynomial. We applied the above forecasting procedure to the modified series and added the results to the trend values forecasted by the cubic polynomial. The result is shown in Fig. 3, where the solid line refers to actual observations, and the dashed line is the forecast for 1995–2005. The 90% confidence interval does not exceed 0.5°C.

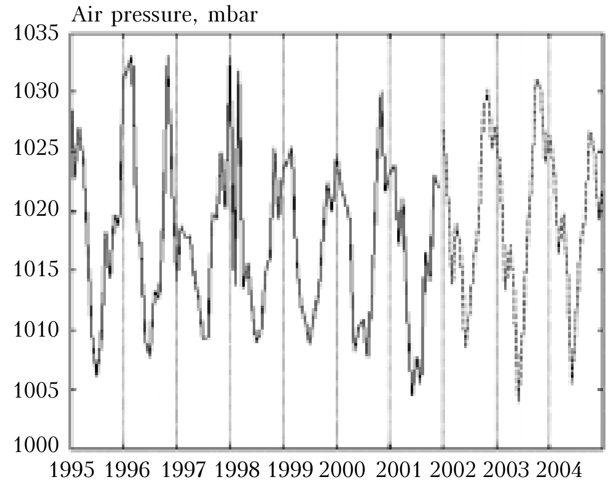

Fig. 2. Forecast (dashed line) of atmospheric pressure for Omsk using the wavelet transform.

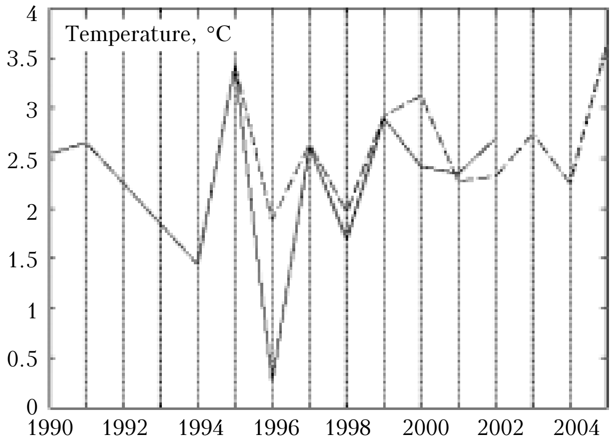

Fig. 3. Actual observations of the annual mean temperatures in Omsk (solid line) and their forecast for 1995–2005 (dashed line).

Figure 3 allows a number of conclusions. The interval 1995–2002 shows a good agreement between the forecast and the observations. There is almost complete quantitative agreement for the odd years, and there is some discrepancy for the even years most strongly expressed in 1996. We believe that the reason for these discrepancies is in the fact that besides quasi-periodic variations and the trend, the temperature series demonstrates some variability with characteristics approaching a random process. Such variability in the distribution of the wavelet transform coefficients definitely contributes to the region of small wavelet coefficient values (compared to numerical values of peaks), which adversely affects the prediction of the wavelet coefficients.

References

1. M.V. Kabanov, Atmos. Oceanic Opt. **15**, No. 1, 95–99 (2002).
2. I.I. Ippolitov, M.V. Kabanov, and S.V. Loginov, Geografiya i Prirodnye Resursy, No. 1 (2005) [to be published].
3. I.G. Zhurbenko and I.A. Kozhevnikova, *Stochastic Process Simulation* (Moscow State University, Moscow, 1990), 150 pp.
4. V.V. Ivanov, Usp. Fiz. Nauk **172**, No. 7, 777–811 (2002).
5. N.M. Astaf'eva, Usp. Fiz. Nauk **166**, No. 11, 1145–1170 (1996).
6. C. Torrence and G.P. Compo, Bull. Am. Meteorol. Soc. **79**, No.1, 61–78 (1998).
7. I.I. Ippolitov, M.V. Kabanov, and S.V. Loginov, Atmos. Oceanic Opt. **15**, No. 1, 16–23 (2002).
8. I.I. Ippolitov, M.V. Kabanov, and S.V. Loginov, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 11, 49–55 (2002).
9. M. Misiti, Y. Misiti, G. Oppenheim, J.M. Poggi, *Wavelet toolbox* (Mathworks Inc., 2001), 891 pp.
10. ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SUNSPOT_NUMBERS/pred0407.prn.
11. ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SUNSPOT_NUMBERS/sunspot.txt.