LONG WTMM-CHAINS OF RIVERS RUNOFF TIME SERIES

Lyubushin* A.A., Bolgov** M.V., Maxutova* M.V.

*Russian Academy of Sciences, Institute of the Physics of the Earth,
123995, Moscow, Russia, Bolshaya Gruzinskaya, 10;
e-mail: lubushin@mtu-net.ru

**Russian Academy of Sciences, Institute of Water Problems,
119991, Moscow, Russia, Gubkina, 3;
e-mail: caspian-sea@mtu-net.ru

ABSTRACT

The problem of extracting time intervals with trend or stationary behavior of hydrological regime of 16 annual river's runoff time series of Volga, Don and Dnepr basins is investigated using analysis of long continuous wavelet transform modulus maximum chains. Characteristic periods 4.5-7.5 and 12-13 years are detected and time intervals of transient hydrological regime behavior are extracted. Cyclic behavior with periods 12-13 years started and migration of 7.5-years periodicity to the 4.5-years took place at 1920-1940 whereas a chaotic regime without explicit periodicity has been occurred at 1950-1970. Time intervals 1903-1912 and 1923-1937 are characterized by the most intensive regime changes at time scales 3-18 years.

INTRODUCTION

One of the problems of time series analysis is detecting of change points, i.e. those time moments which correspond to rapid changing of the signal to be analyzed. Another problem is seeking for the stationary time points, i.e. time moments, corresponding to local extremes (minimums or maximums). It is obvious that the type of the behavior of the signal depends on the time scale at which the signal is analyzed. The last means that the signal must be averaged within certain moving time window before analyzing. Thus, the properties of the averaged signal depend on the radius of the averaging moving time window or when using other methods of smoothing – on the efficient radius of "influence" of smoothing kernel function. This radius of smoothing is nothing else as the time scale of the signal analysis. The time points of local minimums, maximums and minimum and maximum values of the 1st derivative of the smoothed signal provide natural fragmenting of the signal's behavior at the given time scale. Let us call these time points as the scale-dependent extreme points of the signal.

For the most of natural time series when the scale value is small the averaged signal possesses a lot of extreme points but with scale increasing this number decreases. When the scale value is increasing gradually we can perform the chaining of the extreme points of the smoothed signal of the same type i.e. separately points of local minimums, local maximums, points of maximums of the 1st derivative (maximum positive trends) and points of minimums the derivative (maximum negative trends). Thus, we have four types of chains of extreme points on the plane of time-scale values. The most of these chains abort with scale increasing rather rapidly but some of them have a very large length and propagate from minimum scale values up to maximum possible scale values which are admitted for the analysis taking into account the finite volume of the time series sample.

The procedure described above is known in the wavelet analysis as wavelet transform modulus maximums (WTMM) analysis and the set of chains of WTMM-points is called the WTMM-skeleton [*Mallat, 1998*]. Skeleton of WTMM-points are used for image analysis (detecting of boundaries and textures of the patterns) [*Hummel, Moniot, 1989; Yuille, Poggio, 1986*]. In the turbulence and financial researches multi-fractal spectrums of WTMM-points at the limit of scale values tending to zero (spectrum of singularities) are used [*Bacry, Muzy, Arneodo, 1993; Muzy, Delour, Bacry, 2000*].

At the same time an individual pattern of the signal is formed by longest chains of WTMM-skeleton, i.e. for scale values which are comparable with the length of time interval where the signal is defined. That is why let us leave for the analysis the longest chains of scale-dependent extreme points only. These longest chains form a characteristic pattern of the time series behavior which could be regarded as its "fingerprint". The long WTMM-chains present the evolution of scale-dependent extreme points in time and in scale simultaneously. The time moments corresponding to the beginnings of long chains (for the smallest scale value) indicate the most significant extreme points among all others within the smallest scale level. Another interesting class of points is formed by the final points of the long chains of the local extremes which became close to each other on some rather big scale level. Let us call these points as bifurcation points. At the vicinity of bifurcation time point the smoothed signal behaves approximately as a constant. From continuity of the smoothed curves it follows that long chains of local extreme points of these curves could either connect with opposite type of extremes (minimum with maximums) in the bifurcation points or "come to infinity", i.e. go to some upper limit of possible scale values.

The purpose of this paper is an effort to characterize quantitatively the general behavior of the group of annual rivers runoff time series using their long WTMM-chains. At this sense it is a continuation of the study [*Lyubushin et al., 2003*] on detecting collective effects in annual variations of monthly runoff time series using multidimensional spectral approach.

METHOD

Let $\mathbf{x(t)}$ be a signal for analysis. Note that at least we will subtract the mean value of the signal $\mathbf{x(t)}$ from its values before the analysis but for some cases we will subtract its linear trend or even fit the trend polinom of the higher order.

A scale-dependent kernel smoothed signal [*Hardle, 1989*]:

$$\mathbf{c_0(t,a)} = \frac{\int_{-\infty}^{+\infty} \mathbf{x(t+av)\cdot\psi_0(v)\,dv}}{\int_{-\infty}^{+\infty} \mathbf{\psi_0(v)\,dv}} \tag{1}$$

where $\mathbf{a>0}$ is a scale value and $\mathbf{\psi_0(t)}$ is some fast decay function. Further on we shall use Gaussian: $\mathbf{\psi_0(t)=\exp(-t^2)}$. Let us take the wavelet function:

$$\mathbf{\psi_n(t)=(-1)^n\cdot\frac{d^n\psi_0(t)}{dt^n}\equiv(-1)^n\cdot\psi_0^{(n)}(t)} \tag{2}$$

Using formula for by-part integrating and fast decay properties of the function $\mathbf{\psi_0(t)}$ we can obtain a formula for Taylor's coefficients (the $\mathbf{n}$-th derivative of the smoothed signal, divided by $\mathbf{n!}$) for the smoothed signal:

$$\mathbf{c_n(t,a)\equiv\frac{1}{n!}\frac{d^n c_0(t,a)}{dt^n}=\frac{\int_{-\infty}^{+\infty} x(t+av)\psi_n(v)\,dv}{a^n\int_{-\infty}^{+\infty} v^n\psi_n(v)\,dv}} \tag{3}$$

The formula (1) is a particular case of the formula (3) for $\mathbf{n=0}$.

The wavelet transform modulus maximum point (WTMM-point) $\mathbf{(t,a)}$ for $\mathbf{n\geq 1}$ is defined as the point of local maximum of the value $\mathbf{|c_n(t,a)|}$ with respect to time $\mathbf{t}$ for given scale $\mathbf{a}$ [*Mallat, 1998*]. For $\mathbf{n=0}$ WTMM-points are defined as points of local extremes (maximums or minimums) of the smoothed signal $\mathbf{c_0(t,a)}$. WTMM-points could by jointed into chains. The set of all chains forms a WTMM-skeleton [*Mallat, 1998*] of the signal. If $\mathbf{\psi_0(t)}$ is a Gaussian, then WTMM-skeleton chain could not be aborted when the scale $\mathbf{a}$ is decreased [*Hummel, Moniot, 1989; Yuille, Poggio, 1986; Mallat, 1998*]. The WTMM-points for the 1st derivative $\mathbf{c_1(t,a)}$ indicate time points of the maximum trend (positive or negative) of the smoothed signal $\mathbf{c_0(t,a)}$ for the given scale value $\mathbf{a}$.

Let the signal $\mathbf{x(t)}$ be given for $\mathbf{t\in[0,T]}$. When the time moment $\mathbf{t}$ is close to the beginning or to the end of the interval $\mathbf{[0,T]}$, then smoothing transform (3) is exposed to the absence of information about behavior of the signal $\mathbf{x(t)}$ for $\mathbf{t<0}$ or for $\mathbf{t>T}$. Usually this difficulty is overcome by regarding the signal $\mathbf{x(t)}$ as given not on time interval but on the circle, i.e. by extension the signal outside time interval $\mathbf{[0,T]}$ by the rule: if $\mathbf{t<0}$ then $\mathbf{x(t)=x(T+t)}$ and if $\mathbf{t>T}$ then $\mathbf{x(t)=x(t-T)}$. This circular extension of the signal provides the ability to compute smoothing transform (3) for all time moments and is useful from the points of view of applying fast Fourier transform for fast computing the values (3). Nevertheless the values (3) are garbled at the ends of time interval $\mathbf{[0,T]}$ and it would be better to introduce some "dead intervals" at the beginning and at the end of $\mathbf{[0,T]}$ such that for time moments within these dead intervals WTMM-points are excluded from the analysis and from chaining procedure. For Gaussian $\mathbf{\psi_n(t)\approx 0}$ for $\mathbf{|t|\geq 3}$. Thus, we can introduce the following rule for dead end interval:

$$\begin{aligned}&\textbf{if}\quad \mathbf{0\leq t\leq 3a}\quad\textbf{or}\quad \mathbf{T-3a\leq t\leq T}\\ &\textbf{then}\quad \mathbf{c_n(t,a)}\quad\textbf{are excluded}\\ &\textbf{from analysis}\end{aligned} \tag{4}$$

From the rule (4) it follows the value of maximum possible scale value $\mathbf{a_{max}}$ which is suitable for the analysis: $\mathbf{a_{max}=T/6}$ – for this value the only admitted time point is $\mathbf{t=T/2}$. The right-hand ends of the dead time intervals adjacent to $\mathbf{t=0}$ and the left-hand end of dead time intervals adjacent to $\mathbf{t = T}$ form a cupola-like area of permissible points $\mathbf{(t, a)}$ on the 2D time-scales diagrams on the plane of $\mathbf{(t, ln(a))}$-values. From the property of continuality of smoothed curves $\mathbf{c_0(t, a)}$ it follows that long chains of their local minimums or maximums could have their ends either at the bifurcation point or on the upper boundary of possible values of the scale which is followed from the condition (4).

The long chains of local extreme (minimum and maximum) values of $\mathbf{c_0(t, a)}$ and maximum absolute values of its 1st derivative $\mathbf{c_1(t, a)}$ present the most interest for characterizing the main features of the signal behavior for various scales because they give some kind of "fingerprint" of the signal. Characterizing chains of 1st derivative modulus maximums $\mathbf{|c_1(t, a)|}$ we must differ chains with negative from positive signs of $\mathbf{c_1(t, a)}$ as chains of maximum scale-dependent negative (decreasing) or positive (increasing) trends. For time series with sampling time interval $\mathbf{\Delta t}$ the minimum scale is equal to the period of Nyquist: $\mathbf{a_{min} = 2 \cdot \Delta t}$. Further on we must define the criteria of long chain: it must attain or exceed certain threshold level $\gamma \cdot \mathbf{a_{max}}$ where parameter of the method γ must

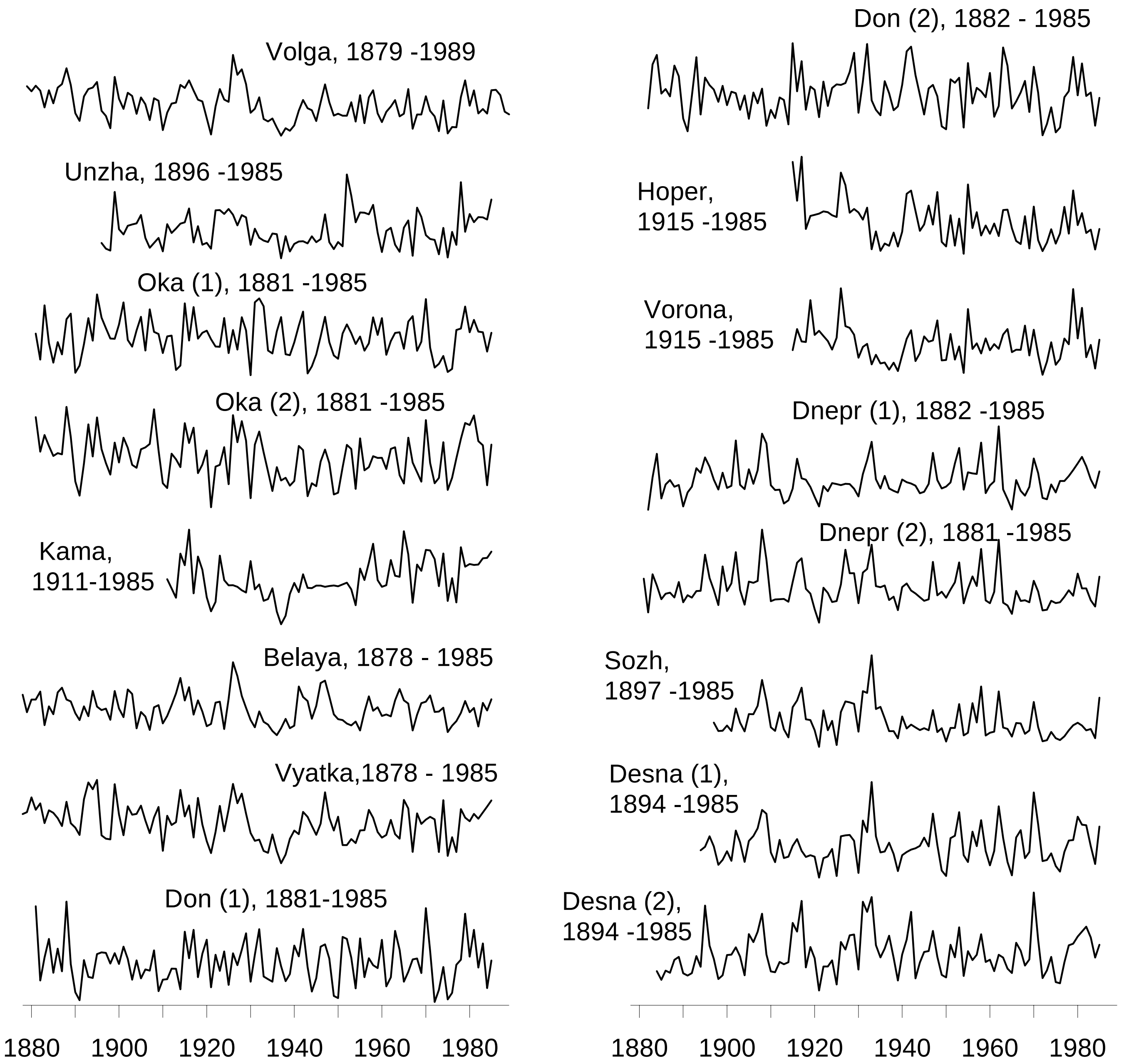


Fig.1. Time series of annual runoff of 16 rivers of the Volga, Don & Dnepr basins

satisfy the condition: $\frac{a_{min}}{a_{max}} \equiv \gamma_{min} \le \gamma < 1$. The closer is the value of γ to 1, the less is the number of "long chains". For $\gamma = \gamma_{min}$ all chains are considered to be "long".

If the value of the parameter γ is gradually increasing then a hierarchical set of scale-dependent change points of time moments of the beginning of "long chains" occur. The study of such hierarchical change points is an alternative to usual approach based on spectral characteristics of the signals [*Detection of abrupt changes*, *1986*] which is valid especially for the case of short time series.

CASE STUDY.

The Fig.1 presents graphics of all 16 annual runoff time series and the Fig.2 – graphics of all long WTMM-chains for Volga time series for γ =0.2. The same graphics could be plotted for each of the signals. Let us calculate histograms of time moments separately for beginning of different types of long WTMM-chains and bifurcation points. These histograms are plotted on the Fig.3.

It is necessary to emphasize an arrangement of the diagrams: at first goes histograms of the time moments of the beginnings of long local minima chains, then - chains of maximal positive derivative, further - local maxima and finally - maximal negative derivative chains. This order is chosen not casually, because it corresponds to the "natural cyclic order": after a minimum - growth up to a maximum, and then - recession up to a minimum etc. The visual analysis of the diagrams in a Fig.3 allows easily to notice, that

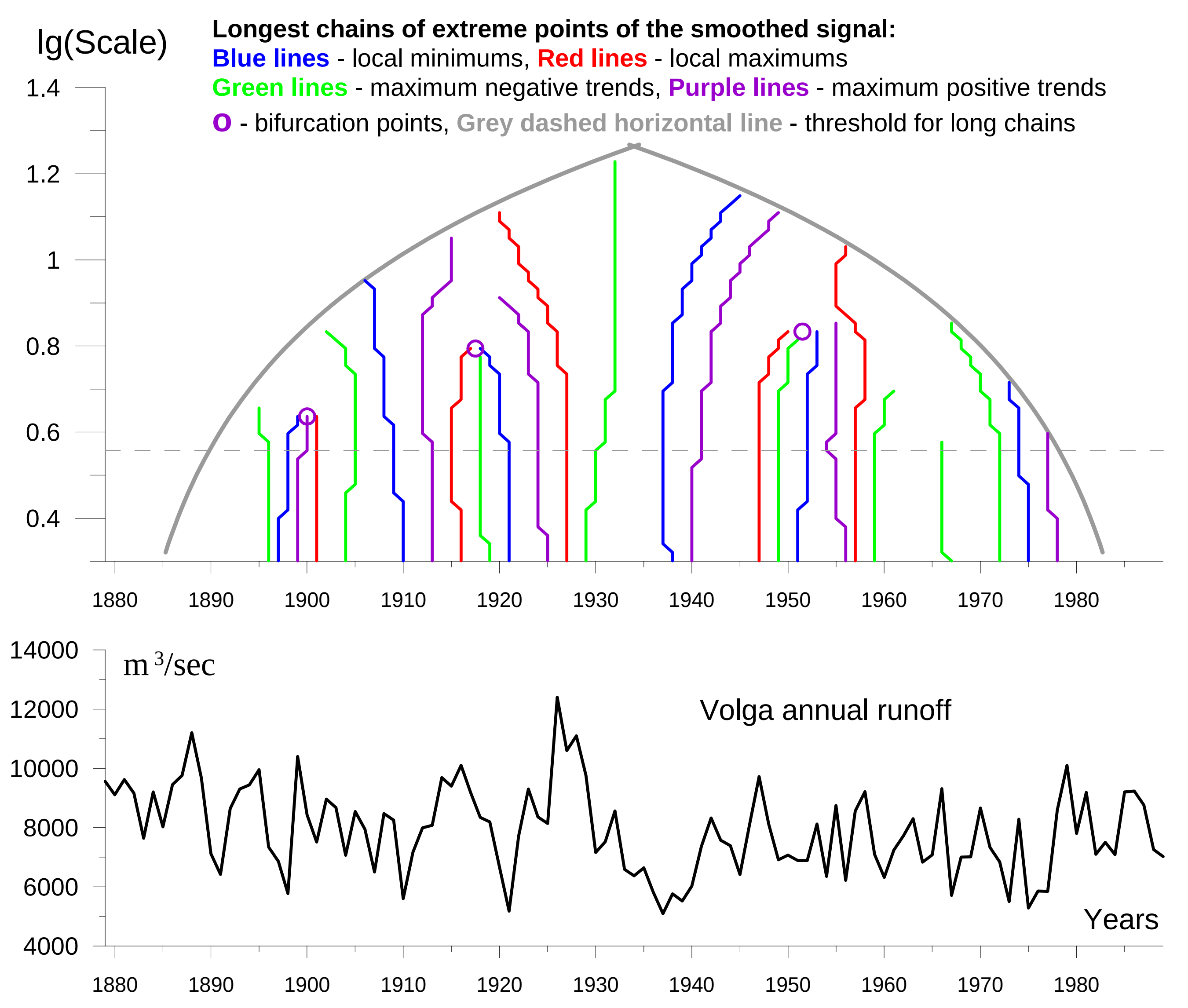


Fig.2. Long WTMM-chains for annual Volga runoff time series, γ = 0.2

approximately till 1950 the peak values of histograms are located in such the "natural cyclic order", but then this order is broken and at the end of an interval of supervision the histograms' peaks arise uncorrelated with histograms' peaks of the moments of time of other types, that is there comes chaos.

This quantitative conclusion should be supplied by some quantitative estimate. Let $\mathbf{H}_\alpha(\mathbf{t})$ be histogram on the Fig.2 at the time moment $\mathbf{t}$ (years), where index α takes values from 1 up to 4 and corresponds to one of the types of long WTMM-chains at the order from up downwards as presented at the Fig.3 for the first 4 graphics. Let $\bar{\mathbf{H}}_\alpha$ be the mean value of correspondent histogram. Let's consider a 2-parametric family of time moments: $\tau_\alpha(\mathbf{k},\eta) = \mathbf{k} + \eta\cdot(\alpha - 1), \alpha = 1,...,4$. Let us consider the function:

$$\Phi(\eta) = \frac{\sum_{k=t_0}^{t_1-3\eta} \left| \sum_{\alpha=1}^{4} (\mathbf{H}_\alpha(\tau_\alpha(\mathbf{k},\eta) - \bar{\mathbf{H}}_\alpha)/4 \right|}{(\mathbf{t}_1 - 3\eta - \mathbf{t}_0)} \qquad (5)$$

The inner sum in the formula (5) presents the mean

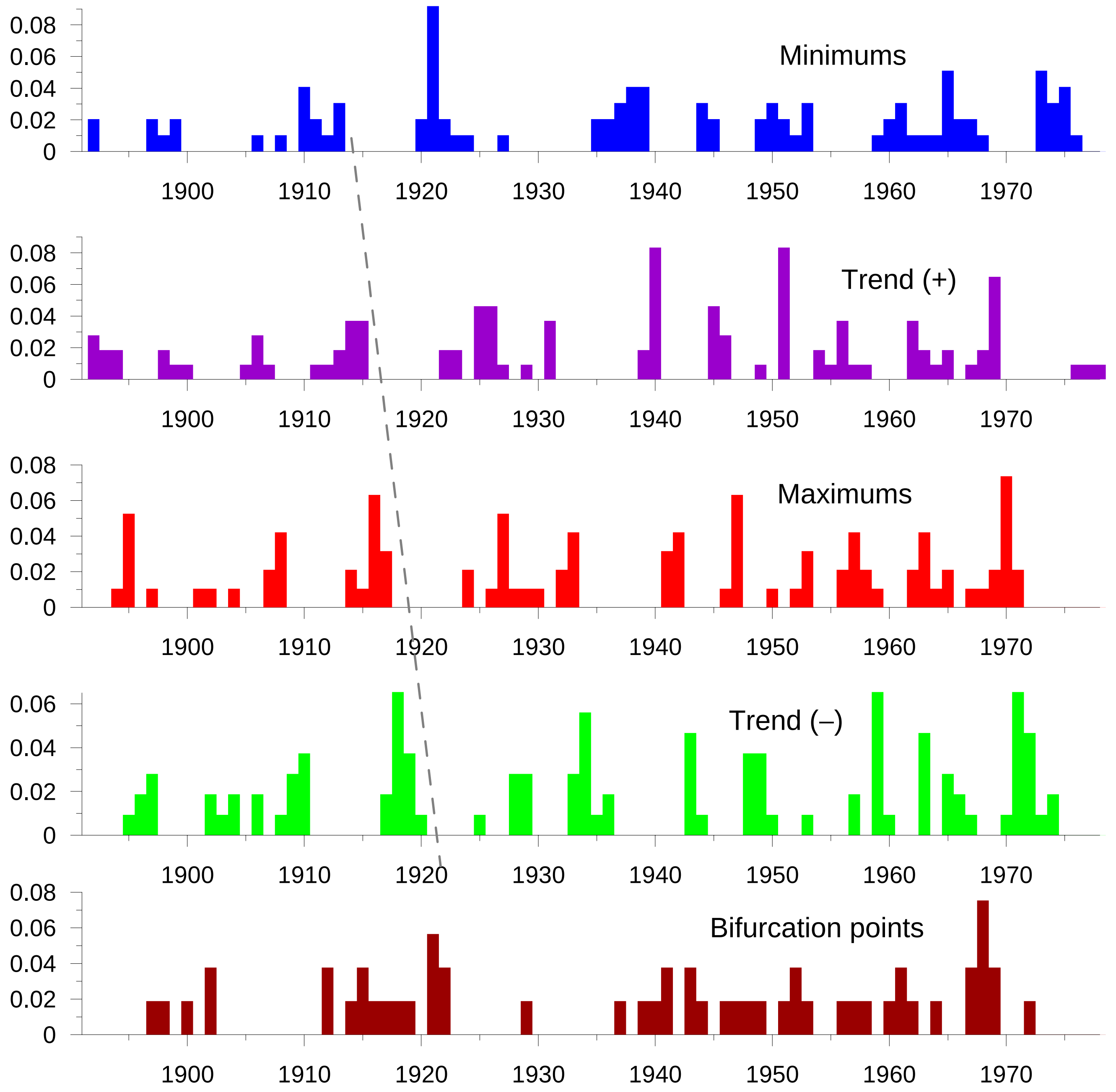


Fig.3. Histograms of time moments of long WTMM-chains beginnings and bifurcation points for all 16 time series in "natural cyclic order" for $\gamma = 0.2$.

value of deviations from mean histogram values along the rectilinear line of probable migration of peak histogram values with the slope η. Further on these mean values are taken by absolute values and are sum by all possible beginnings of rectilinear line (the point $\mathbf{k}$) at the upper graph. After that the resulting sum is divided by the general number of possible values of $\mathbf{k}$ under the condition that tested rectilinear lines begin not earlier than time moment $\mathbf{t_0}$ and are finished not later than time moment $\mathbf{t_1}$ (that is why the restriction $\mathbf{k \le t_1 - 3\eta}$ is imposed in the outer sum in the formula (5)). If histograms' peaks occur at the natural cyclic order with the period $\mathbf{P}$, then the function (5) will have a maximum for the slope $\eta = \mathbf{P/3}$ – just because the peaks and troughs of histograms must be correlated along the line with a slope η. The value of $\mathbf{\Phi(\eta)}$ is some kind of coherence criterion for the behavior of histograms after their simultaneous shift along the time axes on the value of η. Thus, the points of maximums of $\mathbf{\Phi(\eta)}$ distinguish one thirds of cycle period of hydrological regimes within considered time interval $\mathbf{[t_0, t_1]}$. Minimum tested value of η equals zero that corresponds to vertical line straight line, whereas maximum value is taken from condition that $\mathbf{3\eta_{max} \approx a_{max} = T/6}$. The most typical length of time series equals 100 years – that is why $\eta_{\mathbf{max}} = \mathbf{6}$.

The Fig.4 presents graphics of the function (5) for different variants of time intervals $\mathbf{[t_0, t_1]}$ for $\eta \in \mathbf{[0, 6]}$. For the whole time interval 1891-1978 of the histograms' values the most prominent peak equals approximately 2 years (period equals 6 years). The same peak value remains the dominant for the 1st half 1891-1935, but for the 2nd half (1935-1978) this peaks migrates towards less values and becomes near 4.5 periodicity. At the same time a new periodicity near 12-13 years occurs. For moving time window length 30 years (the 2nd row of graphics at the Fig.4) this pattern is preserved quantitatively, but time interval 1920-1950 has distinct features of transfer from one regime into another. The using of 20-years length moving time window (the 2 last rows of graphics at the Fig.4) helps to clear details of the regimes changes. There are 2 time intervals of regimes changes: 1920-1940 and 1950-1970. The interval 1920-1940 is characterized by occurrence of 12-13-years periodicity which exists together with the 6-7-years one. Besides that the 18-years periodicity has explicit signs for this time interval and this is the only interval it is observed. The time interval 1950-1970 has no signs of periodicities and have a large value of function (5) for $\eta = \mathbf{0}$, that is a sign of chaotic behavior of histograms' peaks. It should be noticed that a dashed line of histograms' peaks migration plotted at the Fig.4 corresponds to the value of the slope $\eta = \mathbf{2.5}$ what is typical for interval 1910-1930.

Finally, the bifurcation points histogram at the Fig.3 contains 2 prolonged time intervals with absence of bifurcations points or when these points occur very rarely: 1903-1912 and 1923-1937. It means that these intervals of time posses the most intensive changes of hydrological regimes for all time scales values exceeding some threshold $\mathbf{a_*}$ for all time series to be analyzed. Taking into account that the most typical length of time series $\mathbf{T}$=100 years, we can obtain that $\mathbf{a_* \approx 3}$ years. For these conditions the maximum scale value will be near 18 years.

Conclusions

The typical regimes of annual rivers runoff time series of the European part of the former USSR are distinguished which have periods 4.5-7.5 and 12-13 years. Time intervals 1920-1940 and 1950-1970 are detected as the change intervals. At the 1st case a new 12-13-years periodicity occurs and at the 2nd case a chaotic regime took place without any periodic features. Time intervals 1903-1912 and 1923-1937 have the most intensive changes for all scales from 3 up to 18 years.

References

Bacry E., J.F. Muzy, and A.Arneodo. (1993) Singularity spectrum of fractal signals: exact results. J. of Stat Phys., 70(3/4): 635-674.

Detection of abrupt changes in signals and dynamical systems (1986) – Lecture notes in control and informational sciences, Issue 77, Edited by M.Basseville and A.Banveniste *Hardle W.* (1989) Applied nonparametric regression. Cambridge University Press, Cambridge, New York, New Rochell, Melbourne, Sydney.

Hummel B., R. Moniot. (1989) Reconstruction from zero-crossings in scale-space. IEEE Trans. Acoust., Speech, and Signal Proc., 37(12), December 1989.

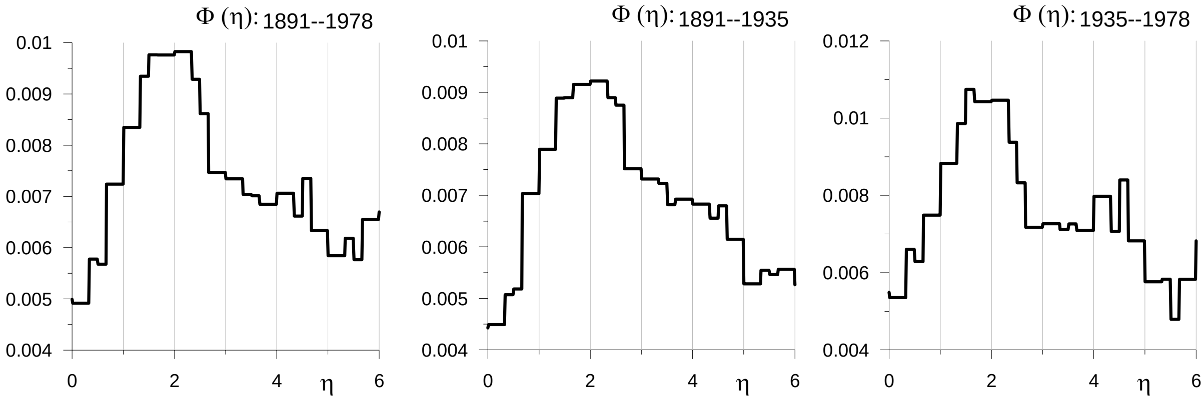
Lyubushin A.A., Pisarenko V.F., Bolgov M.V. and Rukavishnikova T.A. (2003) Study of General Effects of Rivers Runoff Variations – Russian Meteorology and Hydrology, 2003, No.7, pp. 59-68, http://lyubushin.hotbox.ru/Coherence_measures_of_rivers_runoff_time_series.pdf

Mallat S. (1998) A wavelet tour of signal processing. Academic Press. San Diego, London, Boston, N.Y., Sydney, Tokyo, Toronto. 577 p.

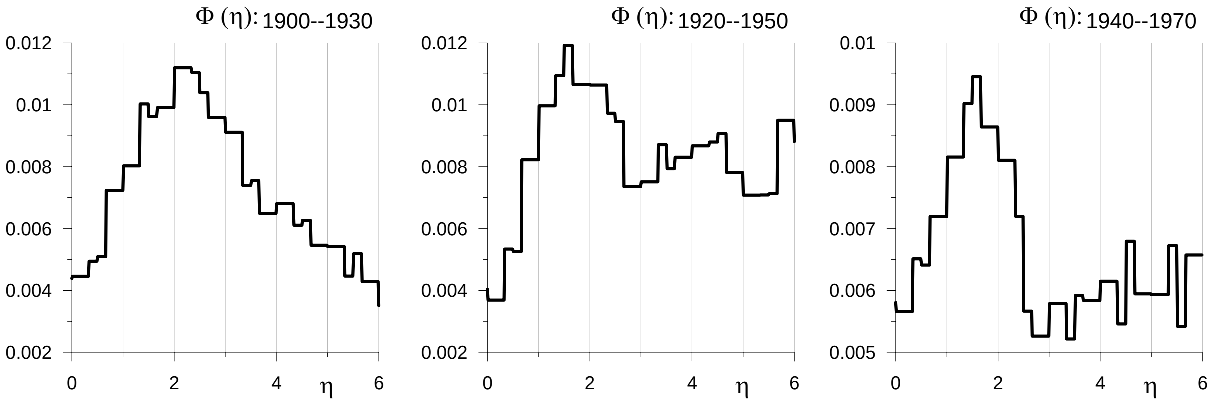
Muzy J.F., J. Delour, and E. Bacry (2000) Modeling fluctuations of financial time series: from cascade process to stochastic volatility model. Eur. Phys. J. B 17, 537-548 (2000)

Yuille A., T. Poggio. (1986) Scaling theorems for zero-crossings. IEEE Trans. Patt. Recog. And Mach. Intell., 8. January 1986.

The whole interval, the 1st and the 2nd half



Moving time window of the length 30 years



Moving time window of the length 20 years

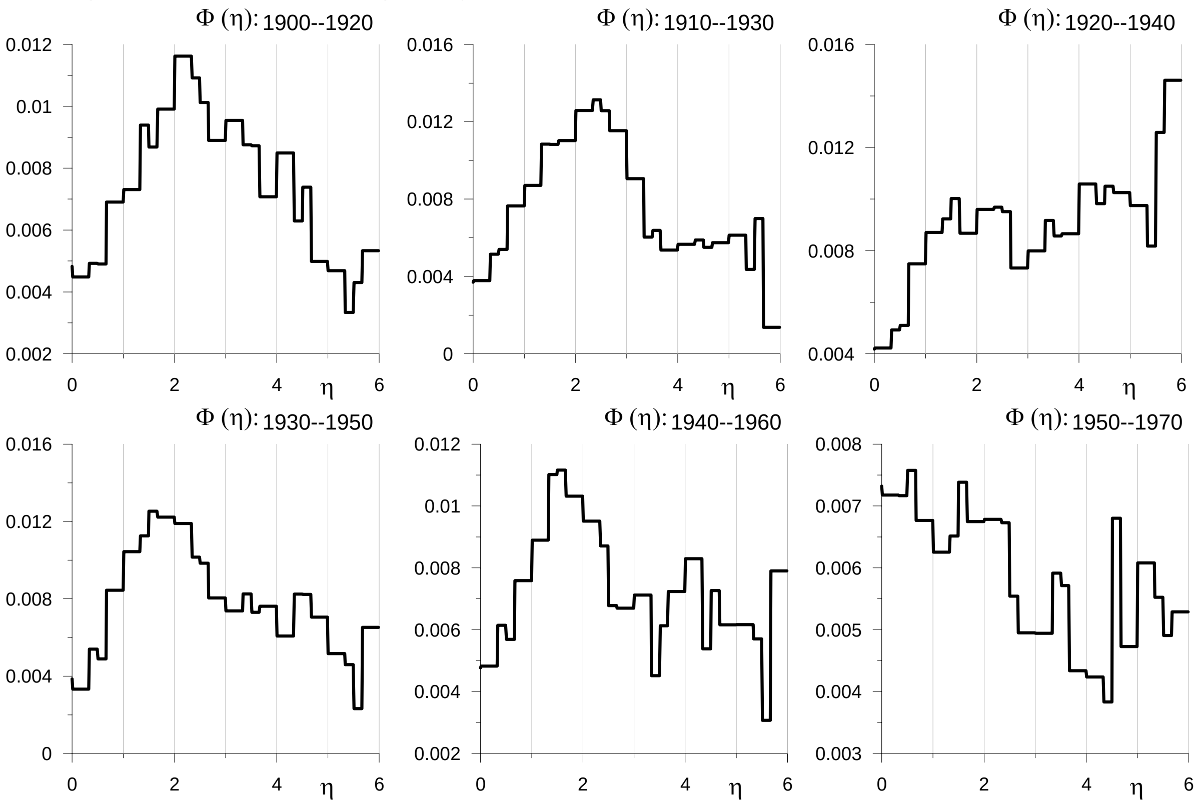


Fig.4.