Analysis of Caspian Sea Coastal Observations by Wavelet-Based Robust Coherence Measures

Alexey A. Lyubushin

Russian Academy of Sciences, Institute of the Physics of the Earth,
123995, Moscow, Russia, Bolshaya Gruzinskaya, 10;
fax: +007-095-255-60-40
e-mail: lubushin@mtu-net.ru

ABSTRACT

Multidimensional time series of Caspian Sea level and wind speed measured in 15 coastal stations during time interval 1977–1991 were analyzed using wavelet-based approach with a purpose detect time and scale-dependent effects of collective behavior. The results present a sequence of sharp peaks of the coherence measure at the 1st detail level with duration near 1.5 year. These bursts of coherence are connected to intensive slow movements of sea bottom during aftershocks seismic activity after the strong earthquake 06.03.1986, M=6.6. Another detected effect consists in decreasing of the "strength" of coherence bursts during seasonal peaks (which correspond to each autumn-winter period) with approaching to the end of observation interval. For wind speed observations the main result consists in similar decreasing the strength of collective behavior. This decreasing could be connected with regional change of atmospheric circulation within Caspian Sea region.

1 INTRODUCTION

Here we present results obtained by the method, which was elaborated in [*Lyubushin, 2000-2002*]. These methods are intended for detecting collective effects within scalar components of multi-dimensional time series of monitoring and are based on wavelet decomposition [*Chui, 1992; Daubechies, 1992; Mallat, 1998*]. The method is a wavelet modification of previously elaborated methods of collective effects extracting, which were realized for Fourier decomposition and using of classic multi-dimensional parametric models of multiple time series [*Lyubushin, 1998*]. In papers [*Lyubushin et al., 2003, 2004*] the Fourier-based method was used for statistical analysis of rivers' runoff and Caspian Sea level multiple time series. We preferred wavelet-based approach because it is the most suitable for investigating transient effects within signals. We used a robust modification of the wavelet-based coherence measure which was proposed in [*Lyubushin, 2002*].

2 METHOD

The method constructs an estimate of scale-dependent measure of coherence behavior in a moving time window. The scale-dependent coherence measure on the given detail level within given time window is the product of absolute values of canonical correlation coefficients [*Hotelling, 1936; Rao, 1965*] of wavelet coefficients of each scalar signal with respect to wavelet coefficients of all other signals. Thus, if we consider $\mathbf{q}, \mathbf{q} \geq 3$ time series then the wavelet-based coherence measure equals to the absolute value of the product of $\mathbf{q}$ canonical correlations. Absolute value of each canonical correlation describes "the strength" of connection of the considered scalar time series with the set of all other time series on the given detail level. It means that the product of $\mathbf{q}$ such values describes the strength of summary effect of collective behavior of the multiple time series. The main details of computing the used coherence measure are described in [*Lyubushin, 2002*]. It is essential to underline that the estimates of canonical correlations are constructed as a robust method [*Huber, 1981*], i.e. they are stable to the presence of outliers within data or within values of wavelet coefficients.

Let $\mathbf{q} \geq 3$ be a general number of scalar time series to be analyzed simultaneously, τ be time index corresponding to the right-hand end of the moving time window of the length $\mathbf{N}$ samples. Let $\mathbf{M} = \mathbf{min}\{\mathbf{2^m} : \mathbf{2^m} \geq \mathbf{N}\}$ be the minimum integer value of the type 2 in the integer power degree which is equal or greater than $\mathbf{N}$.

For each fragment of initial scalar time series corresponding to the current position of time window the following sequence of preprocessing operations are performed:

(i) the fragment is clarified from general linear trend;
(ii) optional: coming to time increments for the fragment;
(iii) the fragment is exposed to the procedure of *tapering* by cosine time window;
(iv) the fragment is renormalized to have a unit sample standard deviation, i.e. each value of the fragment is divided by the sample estimate of the standard deviation;
(v) the fragment is appended by zero values till the general length $\mathbf{M = 2^m}$.

Removing general linear trend for the fragment is a standard operation for suppressing those low-frequency components which could not be estimated statistically significant due to the finite length of the window. Optional operation (ii) of coming to time increments is an additional tool for making the sample within current time window to be more stationary. The tapering procedure [*Jenkins, Watts, 1968*] consists in multiplying the fragments of time series by the value which is equal to 1 for all inner samples of the time window and is going to zero when the time index is going to the left-hand or to the right-hand ends of the window. The law of going to zero of the tapering function is cosine. Afterwards (i.e. after general trends removing, tapering, renormalizing for unit standard deviation and appending by zero values) we perform discrete fast wavelet transforms [*Press et al., 1996*] of all these fragments using some orthogonal wavelet (Haar's wavelet for instance). Appending by zero values till the length $\mathbf{M = 2^m}$ is necessary for applying the fast wavelet transform.

As the result of the operations described above we will have a set of $\mathbf{q}$ wavelet coefficients for detail levels $\alpha = \mathbf{1,...,m}$:

$$\begin{aligned} &\mathbf{c_j^{(\alpha,\tau)}(k), \quad j = 1,...,q;} \\ &\alpha = \mathbf{1,...,m;} \\ &\mathbf{k = 1,...,M_\alpha = 2^{(m-\alpha)}} \end{aligned} \tag{1}$$

The wavelet coefficients $\mathbf{c_j^{(\alpha,\tau)}(k)}$ within the detail level α describe variations of the signals with scales from $\Delta\mathbf{t} \cdot \mathbf{2}^\alpha$ till $\Delta\mathbf{t} \cdot \mathbf{2^{(\alpha+1)}}$, i.e. the 1st detail level is the most high-frequency. The number of wavelet coefficients on the detail level α is equal to $\mathbf{2^{(m-\alpha)}}$. But only the first part of $\mathbf{L_\alpha = 2^{(m-\alpha)} \cdot (N/M) = N \cdot 2^{-\alpha}}$ wavelet coefficients corresponds to the initial, non-appended by zero values parts of time series. Index $\mathbf{k}$ within formula (1) defines the position of scale-dependent vicinity inside the current moving time window which has the influence on the value of wavelet coefficient due to applying orthogonal finite support wavelet transform of the fragment.

Let $\mathbf{j_0}$ be a number of some time series, $\mathbf{1 \le j_0 \le q}$, and let us try to construct a measure describing the connection of the selected time series $\mathbf{j_0}$ with all other scalar time series within current time window. This measure should be scale-dependent, of course. For this purpose let us consider a linear combination:

$$\sum_{\mathbf{j=1,\, j\neq j_0}}^{\mathbf{q}} \mathbf{c_j^{(\alpha,\tau)}(k)} \cdot \gamma_{\mathbf{j}} \tag{2}$$

with unknown coefficients $\gamma_{\mathbf{j}}$. It should be underlined that the index $\mathbf{j_0}$ is omit within the sum in the formula (2). The values of coefficients $\gamma_{\mathbf{j}}$ are found from solution the following minimization problem:

$$\sum_{\mathbf{k=1}}^{\mathbf{L_\alpha}} |\, \mathbf{c_{j_0}^{(\alpha,\tau)}(k)} - \sum_{\mathbf{j=1,\, j\neq j_0}}^{\mathbf{q}} \mathbf{c_j^{(\alpha,\tau)}(k)} \cdot \gamma_{\mathbf{j}} \,| \rightarrow \min_{\gamma_{\mathbf{j}}} \tag{3}$$

The robustness [*Huber, 1981*] of the procedure consists in minimizing the sum of absolute values in the formula (3) instead of the least squares approach, which follows to classical canonical correlation scheme of Hotelling [*Hotelling, 1936; Rao, 1965*]. For least squares approach the unknown coefficients $\gamma_{\mathbf{j}}$ in linear combination (2) are found by linear operation with sub-matrices of covariance matrix of wavelet coefficients whereas the robust minimization problem (3) must be solved numerically by the method of generalized gradient [*Clarke, 1975*] within each position of the moving time window.

Let $\gamma_{\mathbf{j, j_0}}^{(\alpha,\tau)}$ be a solution of the problem (3). Let us define a sequence of scalar values:

$$\mathbf{d_{j_0}^{(\alpha,\tau)}(k)} = \sum_{\mathbf{j=1,\, j\neq j_0}}^{\mathbf{q}} \mathbf{c_j^{(\alpha,\tau)}(k)} \cdot \gamma_{\mathbf{j, j_0}}^{(\alpha,\tau)} \tag{4}$$

which could be called a *robust canonical wavelet coefficients* for the selected time series $\mathbf{j_0}$. Now let us calculate correlation coefficient between the values of canonical coefficients $\mathbf{d_{j_0}^{(\alpha,\tau)}(k)}$ and initial wavelet coefficients $\mathbf{c_{j_0}^{(\alpha,\tau)}(k)}$ for $\mathbf{k = 1,...,L_\alpha}$. In order to provide robustness in all stages of final estimate we will use the robust formula for correlation coefficient [*Huber, 1981*] $\rho(\mathbf{x},\mathbf{y})$ for the pair of samples $\mathbf{x(k), y(k), k = 1,...,n}$:

$$\rho(\mathbf{x},\mathbf{y}) = \frac{\mathbf{S}(\hat{\mathbf{z}}^2) - \mathbf{S}(\breve{\mathbf{z}}^2)}{\mathbf{S}(\hat{\mathbf{z}}^2) + \mathbf{S}(\breve{\mathbf{z}}^2)} \tag{5}$$

where

$$\begin{aligned} &\hat{\mathbf{z}}(\mathbf{k}) = \mathbf{a} \cdot \mathbf{x}(\mathbf{k}) + \mathbf{b} \cdot \mathbf{y}(\mathbf{k}), \\ &\breve{\mathbf{z}}(\mathbf{k}) = \mathbf{a} \cdot \mathbf{x}(\mathbf{k}) - \mathbf{b} \cdot \mathbf{y}(\mathbf{k}), \\ &\mathbf{a} = \mathbf{1}/\mathbf{S}(\mathbf{x}), \mathbf{b} = \mathbf{1}/\mathbf{S}(\mathbf{y}), \\ &\mathbf{S}(\mathbf{x}) = \mathbf{med}\,|\,\mathbf{x} - \mathbf{med}(\mathbf{x})\,| \end{aligned} \tag{6}$$

where $\mathbf{med(x)}$ means median value of the sample $\mathbf{x}$ and $\mathbf{S(x)}$ means absolute median deviation of the $\mathbf{x}$.

Substituting $\mathbf{x(k)}$ for $\mathbf{c}_{\mathbf{j_0}}^{(\alpha,\tau)}(\mathbf{k})$, $\mathbf{y(k)}$ for $\mathbf{d}_{\mathbf{j_0}}^{(\alpha,\tau)}(\mathbf{k})$ and $\mathbf{n}$ for $\mathbf{L}_\alpha$ within formulas (5) and (6) we will obtain the values of robust correlation coefficient $\nu_{\mathbf{j_0}}(\alpha,\tau)$ describing "the strength" of connection of the selected time series $\mathbf{j_0}$ with all other time series. Let us call the value $\nu_{\mathbf{j_0}}(\alpha,\tau)$ as *robust wavelet-based canonical correlation of the series* $\mathbf{j_0}$ (which is scale-dependent obviously).

The necessity for using robust estimates is following from the strong instability of least squares calculations to the presence of outliers in initial time series or in the values of wavelet coefficients. Another reason is the ability of wavelet expansion accumulate the main information about the signal within a few number of wavelet coefficients – that is why the population of wavelet coefficients usually has a probability distribution with "heavy tails", i.e. has outlier values. Thus, using robust estimates for joint analysis of wavelet coefficients from different signals is necessary without connection to probable existence of outliers within initial data. Changing sequentially the number $\mathbf{j_0}$ of selected time series we will obtain the values of robust canonical correlations $\nu_\mathbf{k}(\alpha,\tau)$ for all time series.

Statistical significance of the $\nu_\mathbf{k}(\alpha,\tau)$-estimates depends on the number $\mathbf{L}_\alpha$ with the sum (3). But the number of wavelet coefficients $\mathbf{L}_\alpha$ rapidly decreases with increasing of α whereas for statistically significant estimates we must have some minimum value $\mathbf{L_{min}}$. The value of significance threshold $\mathbf{L_{min}}$ is the parameter of the method and it defines the maximum possible value of the detail level index $\alpha_{\mathbf{max}}$ which could be analyzed by the method: $\alpha_{\mathbf{max}} = \mathbf{max}\{\alpha : \mathbf{L}_\alpha \geq \mathbf{L_{min}}\}$.

After definition of the $\alpha_{\mathbf{max}}$-value the tapering preprocessing operation could be described in more details. The tapering operation is necessary for avoiding circular effects of discrete wavelet transform of the finite length sample [*Press et al., 1996*]. These effects are depending on the considered scales: the more is the scale, the longer are circular effects at the end of the transformed sample. Thus, the length of the end parts of tapering operations must be dependent on the maximum scale which is analyzed. That is why we have taken the length of end parts of the time window where the tapering operation is in action to be equal to $\mathbf{2}^{(\alpha_{\mathbf{max}}-\mathbf{1})}$.

The number of wavelet coefficients taking part in the estimate of the $\nu_\mathbf{k}(\alpha,\tau)$-values rapidly decreases with increasing of the detail level number α. That is why standard deviations of statistical fluctuations of $\nu_\mathbf{k}(\alpha,\tau)$-estimates are not uniform. In order to make statistical fluctuations of estimates dependent on the length of moving time window $\mathbf{N}$ only (and independent on the detail level number) let us introduce additional smoothing operation:

$$\bar{\nu}_\mathbf{k}(\tau,\alpha) = \sum_{\mathbf{s}=\mathbf{1}}^{\mathbf{m}_\alpha} \nu_\mathbf{k}(\tau - \mathbf{s} + \mathbf{1}, \alpha)/\mathbf{m}_\alpha, \mathbf{m}_\alpha = \mathbf{2}^\alpha \tag{7}$$

The more is the number of detail level the deeper is averaging operation (7) using the values of estimates obtained by adjacent previous time windows. According to the formula (7) the efficient length of moving time window became scale-dependent and equals to $\mathbf{N}_\mathbf{e}^{(\alpha)} = \mathbf{N} + \mathbf{2}^\alpha - \mathbf{1}$ samples.

A wavelet-based robust coherence measure (***WBRCM***) is defined by the formula:

$$\kappa(\tau,\alpha) = \prod_{\mathbf{k}=\mathbf{1}}^{\mathbf{q}} |\,\bar{\nu}_\mathbf{k}(\tau,\alpha)\,| \tag{8}$$

The measure (8) has possible values within boundaries from 0 up to 1. The more is the value of (8) the stronger is the collective effect of cooperation between analyzed signals on the time scales corresponding to the detail level number α. It should be noticed that the value (8) is the product of be noticed that the value (8) is the product of $\mathbf{q}$ values from the interval [0,1]. That is why the absolute values of $\kappa(\tau,\alpha)$ depend on the number $\mathbf{q}$ of analyzed time series and the comparison of $\kappa(\tau,\alpha)$-values could be possible for the same values of $\mathbf{q}$. The most interest lays in relative peak values of the statistics (8) for different τ-values (i.e. for different positions of time window). For convenience of simultaneous comparison of $\kappa(\tau,\alpha)$-values for different detail levels let us start the index τ with initial value $\mathbf{N}_e^{(\alpha_{\max})} = \mathbf{N} + 2^{\alpha_{\max}} - 1$. The value of $\kappa(\tau,\alpha)$ is based on the information about the signals for time indexes $\mathbf{t}: \ \tau - \mathbf{N}_e^{(\alpha)} \leq \mathbf{t} \leq \tau$ strictly. Thus, the method has 3 free parameters: type of orthogonal wavelet, the length $\mathbf{N}$ of moving time window and the significance threshold $\mathbf{L}_{\min}$. The purpose of the analysis consists in obtaining and interpretation of extreme $\kappa(\tau,\alpha)$-values.

Thus, free parameters of multiple WBRCM (8) are the length of moving time window $\mathbf{N}$ type of the wavelet and significance threshold $\mathbf{L}_{\min}$.

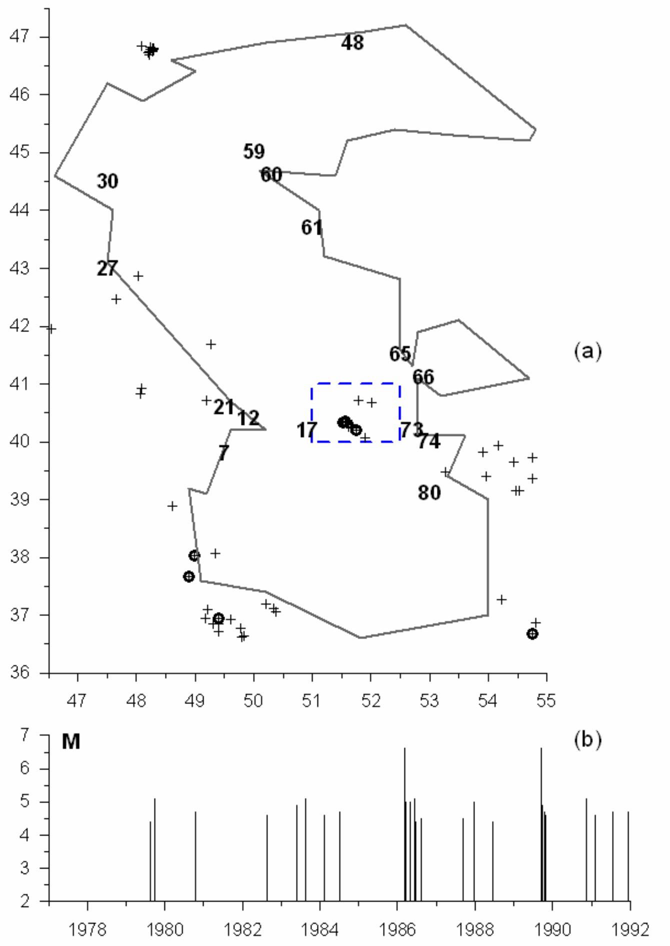

Fig.1. (a) - Positions of 15 coastal stations around Caspian Sea and epicenters of earthquakes with magnitude not less than 4.5 during 1977-1991;
(b) - the sequence of events within blue square at the bottom of the sea.
Crosses - epicenters of M 5 earthquakes, circles - epicenters of M 6 events.

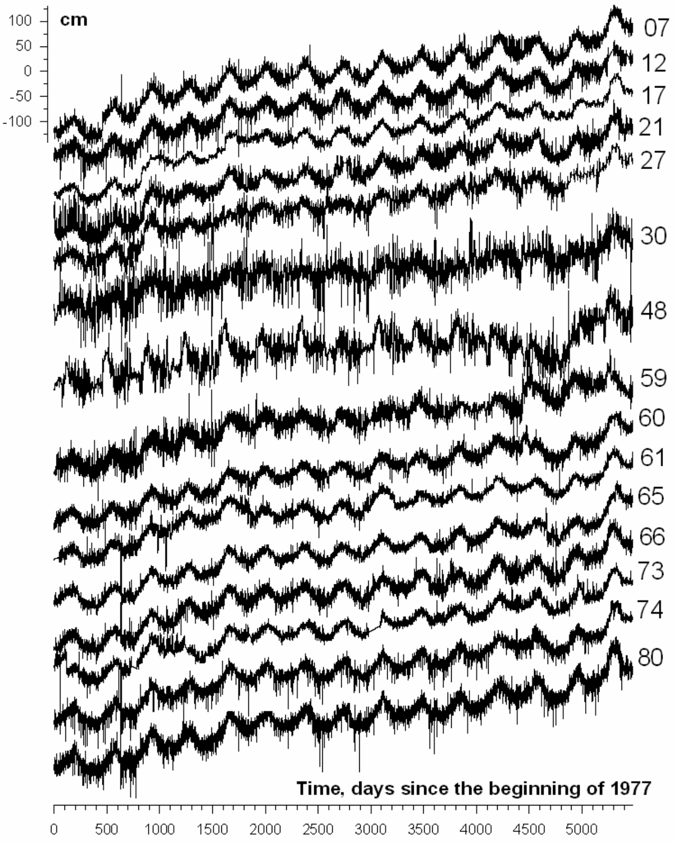

Fig. 2. Graphics of sea level variations at coastal stations with 6-hours sample time interval.

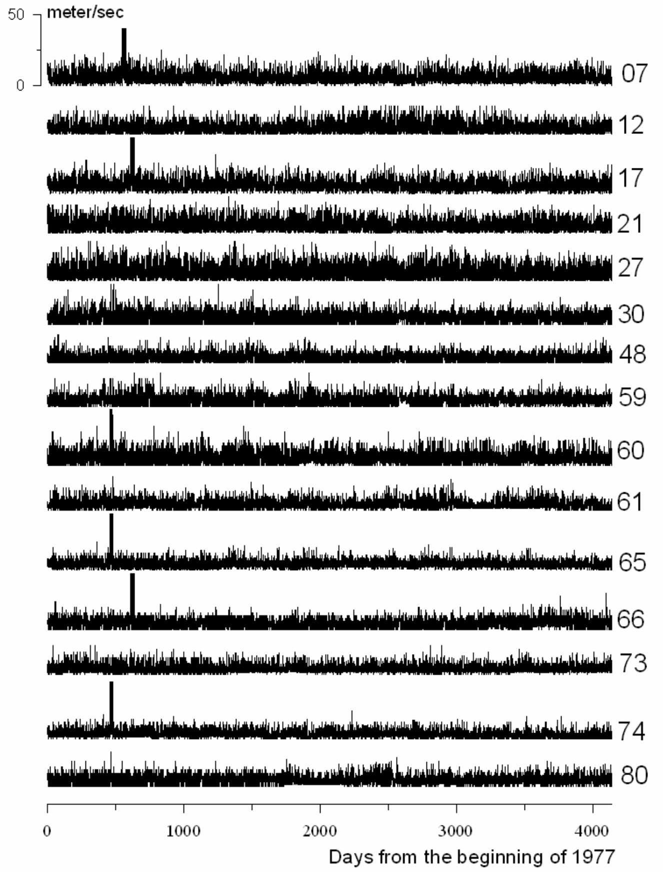

Fig.3. Graphics of initial time series of wind speed, 6-hours samples, from the beginning of 1977 till 30.04.1988.
The scales of Y-axes for all graphics are the same (meter/sec).

3 DATA

The arrangement of 15 coastal stations whose observations were used to solve the problem stated is schematically shown in Fig.1. The numbers of stations are coded; however, for simplification, the standard combination of numerals at the beginning of the codes (970) is omitted. For example, the actual number of station 48 is 97048. The initial time series represent the sequences of synchronous measurements at a time step of 6 h beginning on January 1, 1977, at 09:00 and continuing to the end of 1991. The total duration of each analyzed series is equal to 21908 counts.

In the paper [*Lyubushin et al., 2004*] the data of Caspian Sea level variations from these stations were analyzed with a purpose detecting common harmonic variations for the period 1977-1991. Unfortunately the wind speed data have less duration of synchronous recording.

The Fig.3 contains graphics of all wind speed time series. The initial time series represent the sequences of synchronous measurements at a time step of 6 h beginning on January 1, 1977, at 09:00 and continuing to the end of April, 1988. The total duration of each analyzed series is equal to 16556 counts. Some stations have more long wind records but we have taken this duration to have a possibility for simultaneous processing data from all stations.

It should be noticed that the first third part of observational time interval evidently is characterized by large values for some series (07, 17, 60, 61, 65, 74). The stronger is the wind the more area is touched by its influence. Thus, this is the 1st data peculiarity which can lead to increasing of collective effect covering all stations.

4 RESULTS

The length of moving time window was taken from the reason that it equals the length of climatic season (3 months) and can catch seasonal variations of common effects within wind speed data. The 1st

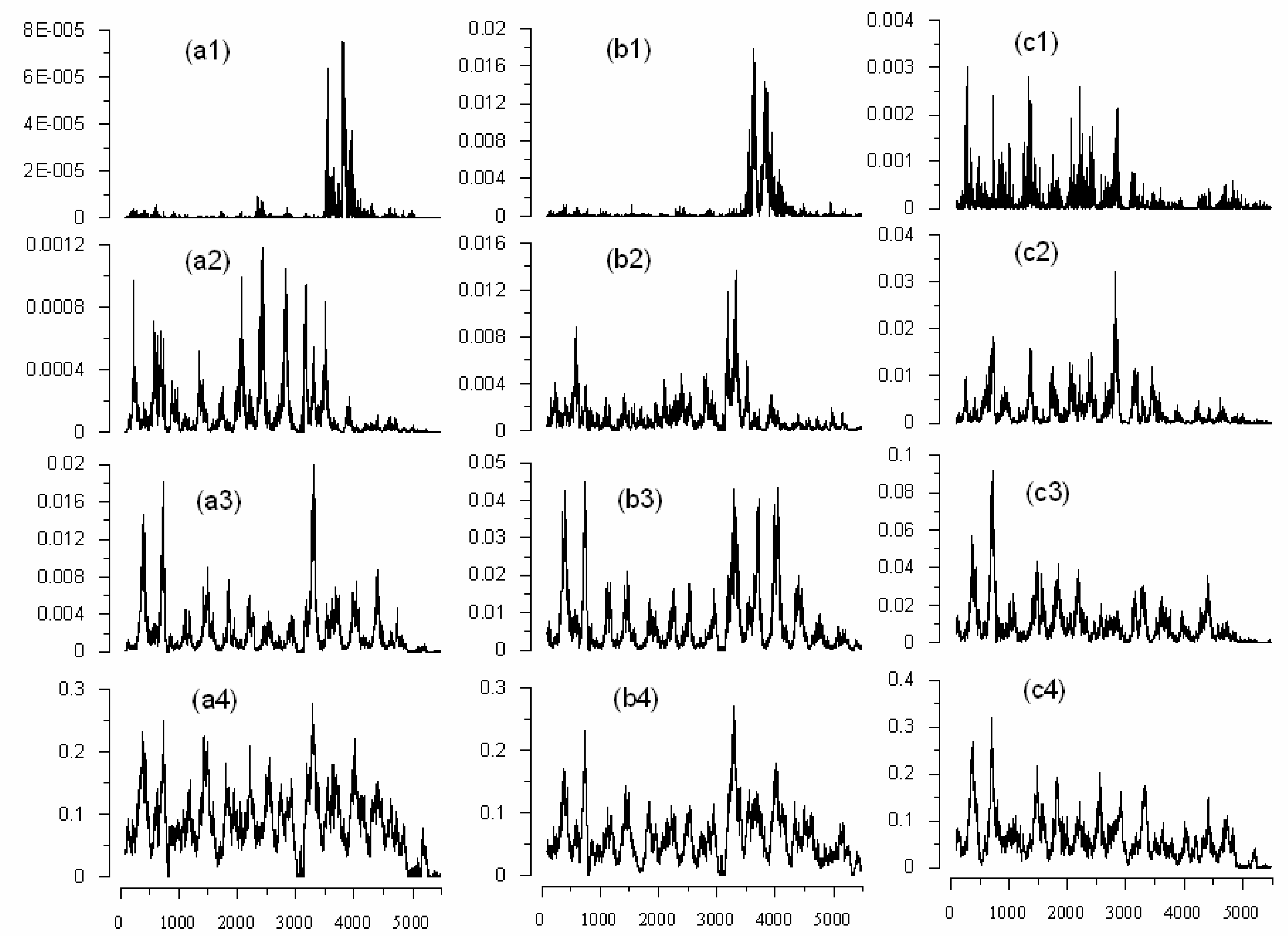

Fig.4. Evolution of wavelet-based robust coherence measures for sea level variations for the first 4 detail levels of wavelet decomposition ((1)-(4)), estimated within moving time window of the length 360 samples (90 days), Haar wavelets, =16. Variant (a) - all 15 stations, variant (b) - 9 stations at the middle of the Caspian Sea (points 7, 12, 17, 21, 65, 66, 73, 74 and 80); variant (c) - 6 stations at the North of the Caspian Sea (points 27, 30, 49, 59, 60 and 61). Time axes correspond to number of days since the beginning of 1977.

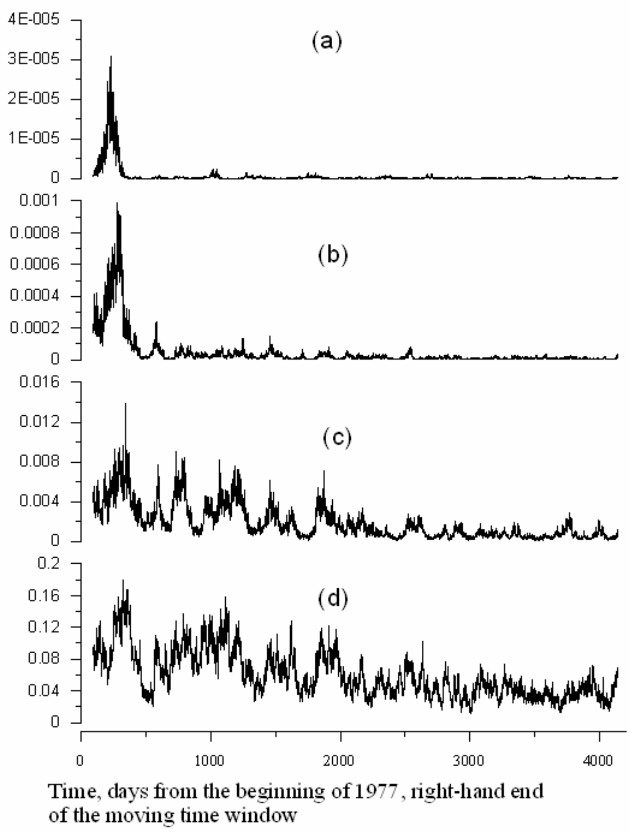

Fig.5. Graphics of the statistics (8) for wind observations for the length of moving time window =360 samples (90 days), significance threshold L_{min}=16 and Haar wavelets.

detail level corresponds to scale range from 12 to 24 hours, the 2nd – from 24 to 48 hours, the 3rd – from 2 till 4 days and the 4th – from 4 up to 8 days variations. The sharp peaks at the 1st detail levels is connected to intensive slow movements of sea bottom during aftershocks seismic activity after the strong earthquake 06.03.1986, M=6.6. Other detail levels have seasonal peaks of collective behavior with decreasing strength from with approaching to the end of observation interval. The more senior detail levels are not possible for the analysis because of the finite length of moving time window ($\mathbf{N}$=360 counts) and the chosen threshold $\mathbf{L_{min}}$=16 for minimum possible number of wavelet coefficients within detail level.

The results of joint analysis of 15-dimensioanl sea level variations time series present a sequence of sharp peaks of the coherence measure at the 1st detail level with duration near 1.5 year (Fig.1(a1)). These bursts of coherence are connected to intensive slow movements of sea bottom during aftershocks seismic activity after the strong earthquake 06.03.1986, M=6.6.

Another detected effect consists in decreasing of the "strength" of coherence bursts during seasonal peaks (which correspond to each autumn-winter period) with approaching to the end of observation interval.

For wind speed data processing the first peculiarity (Fig.5) is the burst of collective effects for detail levels 1 and 2 (12-24 hours and 24-48 hours variations, Fig.5(a) and 5(b)) for time interval 100-400 days (the 2nd half of 1977 – the beginning of 1978). Another interesting effect is gradual decreasing of the strength of collective effects for detail levels 3 and 4 (2-4 days and 4-8 days variations, Fig.5(c) and 5(d)) during all period taking for the analysis. Thus, we have obvious negative trend in common atmospheric circulation with the Caspian region.

5 CONCLUSIONS

Results of multidimensional wavelet-based coastal observations of Caspian Sea level and wind speed during 1977-1991 interval give evidence to the hypothesis about decreasing of the cooperative behavior "strength". This decreasing could be connected with regional change of atmospheric circulation within Caspian Sea region: with approaching to the end of observational time interval 1977–1991 the strong autumn-winter storm winds directions migrate from mostly North-South (along the Sea) more to West-East (across the Sea) what has reflection in decreasing of cooperative effects.

6 REFERENCES

Chui, C.K. (1992) An Introduction to Wavelets, Academic Press, San Diego, CA

Daubechies I. (1992) Ten Lectures on Wavelets. No.61 in CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia

Jenkins G., Watts D.G. (1968) Spectrum Analysis and its Applications. Holden Day. San Francisco

Hotelling H. (1936). Relations between two sets of variates. – Biometrika. Vol.28, pp.321-377.

Huber P.J. (1981) Robust statistics. John Wiley and Sons. New York, Chichester, Brisbane, Toronto.

Lyubushun A.A., G.N.Kopylova and Yu.M.Khatkevich. (1997) Analysis of the Spectral Matrices of Hydrogeological Observations at the Petropavlovsk Geodynamic Research Site, Kamchatka, and Their Comparison with the Seismic Regime – Izvestiya, Physics of the Solid Earth, vol. 33, pp. 497-507.

Lyubushin A.A. (1998(a)) Analysis of Canonical Coherences in the Problems of Geophysical Monitoring – Izvestiya, Physics of the Solid Earth, vol.34, pp. 52-58.

Lyubushin A.A. (1998(b)) An Aggregated Signal of Low-Frequency Geophysical Monitoring Systems. – Izvestiya, Physics of the Solid Earth, vol. 34, pp. 238-243.

Lyubushin A.A. (1999) Analysis of Multidimensional Geophysical Monitoring Time Series for

Earthquake Prediction – Annali di Geofisica, Vol.42, No.5, pp.927-937.

Lyubushin A.A. (2000) Wavelet-Aggregated Signal and Synchronous Peaked Fluctuations in Problems of Geophysical Monitoring and Earthquake Prediction. – Izvestiya, Physics of the Solid Earth, vol.36, pp. 204-213.

Lyubushin A.A. (2002) Robust wavelet-aggregated signal for geophysical monitoring problems – Izvestiya, Physics of the Solid Earth, vol.38, pp. 745-755.

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.

Lyubushin A.A., Pisarenko V.F., Bolgov M.V., Rodkin M.V., and Rukavishnikova T.A. (2004) Synchronous Variations in the Caspian Sea Level from Coastal Observations in 1977–1991. – Atmospheric and Oceanic Physics, 2004, Vol. 40, No.6, pp. 737-746.

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

Press W.H., Flannery B.P., Teukolsky S.A. and Vettering W.T. (1996) Numerical Recipes, 2-nd edition, Chapter 13, Wavelet Transforms, Cambridge Univ. Press, Cambridge.

Rao C.R. (1965) Linear statistical inference and its applications. John Wiley & Sons, Inc. N.Y., London, Sydney.