HINDCASTING OF WIND AND WAVE CLIMATE OF SEAS AROUND RUSSIA

Leonid J. Lopatoukhin[1], Alexander V. Boukhanovsky[2],
Ekaterina S. Chernysheva[2], Sergey V. Ivanov[2]

[1] St. Petersburg State University, Dep. Oceanology. St. Petersburg, Russia
[2] Institute for High Performance Computing, St. Petersburg, Russia

1. INTRODUCTION

In previous decades wave climate data was mainly based on ship observations (e.g., "Ocean wave statistics", "Global wave statistics", etc), sometimes on the wave measurements in selected (mainly near shore) points. Generalization of such data is based on models of random values and time series. Presently the main source of wave climate information is based on the results of hydrodynamic simulation (in other words hindcasting). A lot of models are used. Continuous long-term hindcasting is performed. In this paper the NCEP/NCAR and Sweden reanalysed wind fields are used as preliminary input data. Nested models Wave Watch (WW, versions 1.18, 2.22) and SWAN (versions 40.11, 40.31) are applied. Barents, Caspian, Baltic, Okhotsk, North, Black, Azov seas and Ladoga Lake considered as the deep and shallow water basins. Hindcasting was performed for the grid points and years, shown in the table 1.

Table 1. Characteristics of continuous hindcasting has been made in a seas around Russia

Sea	Years	Lat., N.	Long., E.	Model	Grid step	
					Δx	Δy
Barents	1970–1999	60°-81°	30°W-60°N.	WW-III (1.18)	0.5^0	1.5^0
Okhotsk	1970–1995	35°-65°	135°-165°	"	1.6^0	0.7^0
Caspian	1990–1995	36.5°-47.2°	48°-55.6°	"	0.2^0	0.2^0
Caspian	1980-1999	"	"	SWAN C.III. v.40.11	9 nm	
Caspian (N. part)	1988, 1998	"	"	"	3 nm	
Azov	1989-1998	45°-47.3°	34.7°-39.4°	"	10 nm	
Azov	1979–1998	"	"	"	3 nm	
Baltic	1979-2000	53.8°-66.1°	9°-30°	"	10 nm	
North	1983–1998	50°-70°	5°W.-10°E.	WW-III (2.22)	15 nm	
Black	1974-2003	40.9^0-46.5^0	27.5^0-42.7^0	SWAN C.III. V.40.11	10 nm	
Ladoga lake	1994-2003	59.9^0-61.8^0	29.9^0-33.0^0	SWAN C.III. V.40.31	2 nm	

One of the main results of hindcasting is a set of wave fields (two-dimensional spectra $S(\omega,\theta)$ on selected grids and synoptic terms t, in particularly). Hence, for generalization of the hindcasted results the model of spatiotemporal random fields is needed. For statistical generalization as the tools multivariate statistical analysis are used. The objective of this paper to consider some approaches, results, and models, based on these data. Statistical analysis of spectra parameters, carried out in the terms of Markov processes, has allowed to construct the stochastic model of a spectral wave climate taking into account spatial heterogeneity and temporal variability of wave–making conditions. Special attention is drawn to point and field extremes. Specific and difference between these two statistics is crucial for solution of some applied problems.

2. WIND DATA AND MODEL INPUT

Wave climate investigations, based on hindcasting, needs input wind data $\vec{V}(\vec{r},t)$ on regular grid $\{\vec{r}_k\}$ for time t_i. Now the main source of such information is wind fields reanalysis. The most frequently used data is NCEP/NCAR reanalysis. It contains 6-hourly surface wind fields on the 1.875^0 by latitude Gaussian grid, covering the globe from 1948 up to now. During last years there arises reanalysis with more fine grids, but they are either limited by space (e.g., Swedish reanalysis) or by shorter time span or restricted access (e.g., ECMWF). Therefore one of the main practice problems is adopting of NCEP/NCAR reanalysis as the base for hydrodynamic simulations. Now it is known, that reproduction of wind field in reanalysis depends from specific of basin, i.e. it position, amounts of wind measurements, specific of shore line, etc. As an example three time series of wind speed absolute value are shown in the Fig. 1. They are for Barents, Okhotsk and Caspian seas.

It is seen, that the best correspondence is for Barents Sea (correlation is $\rho \cong 0,9$). A bit worse correspondence is for Okhotsk sea (correlation is $\rho \cong 0,7$). It is connected with specifics of wind data for this sea and their quality. However, for closed Caspian Sea correspondence between measurements and reanalysis is poor ($\rho \cong 0,5$). Moreover, the far to south of the sea the worse correspondence become. It seems, that the main reason is scarce net meteorological stations as compared with the north part, and poor quality of measurements (see, Graham, Cardone et al. (2002), there were forced to reject all the data from Lencoran).
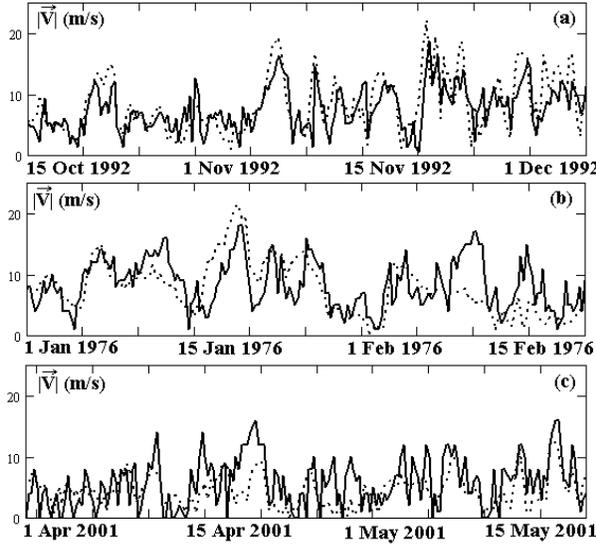
Fig. 1. Comparison of wind speed value time series.
Solid line – Measurements, *Dot line* – Reanalysis.
(a) – Barents Sea (Sental Banken, $74^0.5$N–$31^0.0$E), (b) – Okhotsk Sea (Odoptu, 58^006'N, 143^028' E.), (c) – Caspian Sea (Tuleny island, $44^0 30$N', $47^0 40$'E)

Hence, lack of significant correlation between wind measurements and reanalysis force to assimilate additional sources of data. Such source is ship observations. Naturally all the difference in time of averaging between reanalysis and ship data are taken into account. They are brought to one-to-one correspondence. Starting from 1948, there are about 280 thousands ship observations in Caspian Sea. In contrast to classical problem of assimilation Gill (1991), the reanalysis array is used as it is, and this in turn cause to use stochastic (not hydrodynamic) model of vector spatiotemporal field with the following features:

- Wind speed is the random geometric vector $\vec{V} = (u, v)$, with mean value vector $m_{\vec{V}}$ and tensor of variance $D_{\vec{V}} = E[\vec{V} \otimes \vec{V}]$. This property requires using the tensor algebra for developing of assimilation procedure.

- Vector field exhibit synoptic, annual, and year-to year variability. It demands the development of the multiscale stochastic model, where synoptic variability considered as the stationary, with the periodical modulation of annual rhythms. This modulation must take into account not only mean values, but also the variance as the measure of year-to-year variability. Hence, for the expression of the modulation process the model of periodically correlation stochastic process (PCSP) would be used, see Cyclostationarity, (1993). This model assume the $m_{\vec{V}}(t) = m_{\vec{V}}(t + T)$, $D_{\vec{V}}(t) = D_{\vec{V}}(t + T)$, where $T = 1$ year is the period of modulation.

- Assimilation performs in scattered time and at different points, consequently the approach of converting to the gaps is needed.

Geometric image of tensor $D_{\vec{V}}$ is ellipse with half axis λ_1, λ_2,

turned to angle α. Value $\chi = \lambda_1 / \lambda_2$ specify shape of ellipse. Mean vector s $m_{\vec{V}}$ and r.m.s. $S_{\vec{V}} = D_{\vec{V}}^{0.5}$ of wind speed at 10m level are shown on the fig. 2 (reanalysis grid) for Caspian sea. It is seen, that in summer prevail winds from north. In winter wind variability is more, than in summer (mean vector is noticeably less, than diameter of r.m.s ellipse). Mean vectors and r.m.s. ellipses are quite dissimilar in different points and months.

It follows, that both spatial inhomogenity and periodic nonstationatrity must be included to stochastic model. Method of conditional mathematical expectations (Ogorodnikov, Prigarin, 1996) may be used for description of joint variability of wind field velocity as generalized vector $V_t = \{\vec{V}(r_k, t)\}_{k=1}^n$, consisting from values in the points $\{r_i\}_{i=1}^n$. In the synoptic range for stationary case, this method reduces to vector autoregression:

$$V_t = \sum_{j=1}^{R} \Psi_{t,j} V_{t-j} + \Sigma_t \varepsilon_t . \qquad (1)$$

Where $\Psi_{t,j}$ - block (2x2) matrix of autoregression coefficients, Σ_t - matrix of linear transformation of $2 \cdot n$ - dimensional vector of white noise ε_t. For the case of large grid space n and close positions of points $\vec{r}_i, \vec{r}_j$ identification procedure for $\Psi_{t,j}$ became unreliable. Therefore for closed basins with homogeneous regions relation (1) better to replace on the factor model with, based on orthogonal expansion on basis $\vec{\Phi}_n$:

$$\vec{V}(\vec{r}, t) = m_{\vec{V}}(\vec{r}, t) + \sum_n a_n(t)\vec{\Phi}_n(\vec{r}, t) + \sigma_0(t)\vec{\varepsilon}(\vec{r}, t) . \quad (2)$$

Where $m_{\vec{V}}(t) = m_{\vec{V}}(t + T)$, $\sigma_0(t) = \sigma_0(t + T)$. Scalar coefficients of expansion $a_n(t) = \sigma_n(t)\xi_n(t)$ are general factors responsible for spatial variability of a field, and $\vec{\varepsilon}$ - specific factor responsible for location difference in the points. By this mean system of stationary processes $\xi_n(t)$ define specific of synoptic variability. Variability of r.m.s. $\sigma_n(t) = \sigma_n(t + T)$ define annual and year-to-year modulation of wind speed fields. Orthogonal basis $\Phi_n(\vec{r}, t) = \Phi_n(\vec{r}, t + T)$ are vector empirical orthogonal functions (EOF), which form optimal basis for tensor covariance function $K_{\vec{V}}(\vec{r}_1, \vec{r}_2)$ in the space of Euclidian vectors (see e.g., Boukhanovsky et al (2003a)). The estimates of the first and second EOFs for January and July are shown on the fig. 2(c-f).

It is seen, that the first EOF show intensity of wind field variability from N to SE. In the southern part of the Sea summer and winter variability are different. Second EOF has complicated form and reveals the rotating of wind field due to shore pornography. First five EOF reveals 75%-80% of variability. This defines number of members in expansion (2).

(a)

(b)

Solid line – reanalysis NCEP/NCAR, *dotted line* – correspondent ship observations.
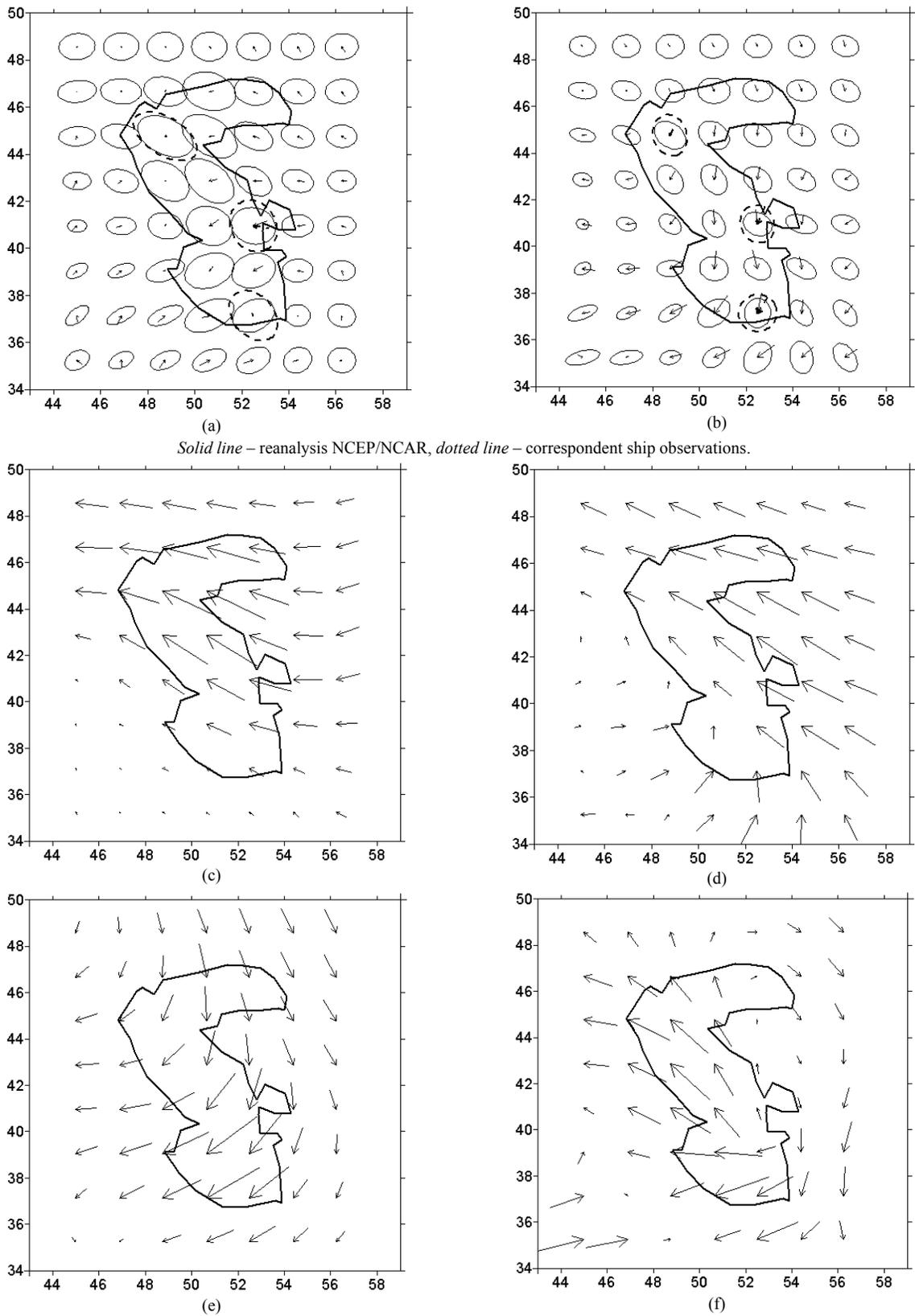
(c)

(d)

(e)

(f)

Fig. 2. Probability characteristics of wind fields over Caspian Sea (reanalysis data: vectors of mean wind speed and r.m.s. in January (a) and July (b), vector EOF in January (first – (c), second – (e)) and in July (first – (d), second– (f)).

Models (1) or (2) express the variability of the wind fields both in synoptic, annual and inter-annual scales. The association of these models with the real data $\vec{V}_s$, observed in point $(\vec{r}_k)$ at the moments t_0, may be presented by means of linear measurement equation:

$$\vec{V}_s = H\vec{V}(\vec{r}_k, t_0) + \vec{\delta}, \qquad (3)$$

where H is the measurement matrix (taking to account the systematic biasing, e.g. due to different averaging of the observations), and $\vec{\delta} = (\delta_u, \delta_v)$ is the vector of random error. The components δ_u, δ_v are correlated due to physical anisotropy of the wind fields.

The coupling of equations (1,2) and (3) allows to reconstruct the **new** ensemble of wind fields $\vec{V}^*(\vec{r}, t)$ by means of optimal assimilation of both the data sources (reanalysis and observations), Himmelblau (1970). If H in (3) is the identity matrix (e.g. the averaging for all the sources is the same), the value of $\vec{V}^*(\vec{r}, t)$ in each point is expressed by means Kalman filter:

$$\vec{V}^*(\vec{r}_k, t_0) = \vec{V}(\vec{r}_k, t_0) + \Re_{\vec{V}}(\vec{r}_k, t_0)(\vec{V}(\vec{r}_k, t_0) - \vec{V}_s). \qquad (4)$$

Here $\Re_{\vec{V}} = \left(\Gamma^{-1} + D_{\vec{V}}^{-1}\right)^{-1}\Gamma^{-1}$ - Kalman tensor coefficient of amplification, where tensor $\Gamma = E\left[\vec{\delta} \otimes \vec{\delta}\right]$ determine the covariance of the observation noise in (3). The principal purpose of the tensor $\Re_{\vec{V}}$ is the weighting of the impact of each data source in new ensemble. These weights depend on the statistical features of the data: e.g. for different points of Caspian Sea impact of the reanalysis is 0.45-0.73 (and for ship observations – 0.27-0.55 respectively). Tensor $\Re_{\vec{V}}(\bullet, t) = \Re_{\vec{V}}(\bullet, t+T)$ periodically varies from month to month due to annual variability of initial data.

Relation (4) is valid only for points and dates, with ship observations. Transfer of assimilated data from point to field is made by mean of model (2), this model consider both regression between adjacent points and their temporal variability. In such a manner model approach (1-4) create the array of reanalysis with assimilation. Data of wind measurements, which did not use for assimilation, are used for verification of this approach. Components (u, v) of wind in 108 severe storms (from 1954 to 1990) for point Krasnovodsk (now, Turkmen-Bashi, on the Eastern shore of the sea), and 3-hour observations in January-April 2001, on Tuleny island (North Part) are presented in the Fig. 3. It is seen, that assimilated data considerably better (correlation 0.7-0.8), than initial data.

3. HINDCASTING PROCEDURE

Difference between model simulations and measurements of wind waves may be due a lot of reasons. In particularly:
- The difference of input wind fields array, used for calculations, from the real. This reason may be partly eliminated by assimilations of additional data in he frames of model (1-4). Also the quality of these data depends on the technique for wind interpolation to the model grid points. We use the smoothing procedure of interpolation for spatiotemporal fields, see Akima (1978).
- Sample variability of wave measurements.
- Errors due to various choices of parameters in spectral wave models. The default values may be not valid for specific basin.

Results of hydrodynamic simulation highly depend from spatial grid and temporal step. Comparison of wave heights measured in the N. Caspian and calculated by SWAN with the step 1 hour and 15 minutes is presented at Fig 4. The directions of severe waves also shown in the same figure. It is seen, that for the long fetches (S, SE) different time steps leads to almost the same results. For short fetches (N, NW) 15-min. time step calculations shows better agreement with the measurements (both by peak wave height and time of it arising) than calculations with 1-hour time step. The reason is, that N and NW storms develops over shallow part with short fetch (this in turn is followed by "swift time" of wave arising). Q-Q plots of these calculations are on the Fig. 5b. Moreover, as example of one of the last (see above) reason we present results (Fig. 5a) of calculations with different values of coefficients $C_{вы}$ and $\widetilde{S}_{pm}$ (cds and stpm in SWAN definitions). Pointed values are shown in the table 2.

Table 2. Values of coefficients $C_{вы}$ and $\widetilde{S}_{pm}$, accepted for calculations

Variants	$C_{вы}$	$\widetilde{S}_{pm}$
By default	2.36	3.02
1	2.85	3.62
2	2.85	2.42
3	1.86	2.42
4	1.86	3.62

It is seen from Fig 5a, that variations in parameters lead to about linear changes of wave heights estimate. Additionally including of whitecapping leads to underestimation of wave height (approximately as twice).

Results of calculations for some severe storm cases are shown in the Fig.6: February 12-14 of 1952 (wind NW-N), March 1-2 of 1952 (wind NW-NNW), November 10-13 of 1952 (wind SE), November 20-21 of 1957 (wind NW). Values of measured highest waves also pointed in this fig. In accordance with Boukhanovsky et al (1998) the upper limit of extreme wave is $h_{max} \cong 4\bar{h} = 2.5h_s$, this is in agreement with values on the Fig. 6.

4. CLIMATIC WAVE SPECTRA

The size of the output data (arrays of the directional wave spectra) is really huge. E.g., for Barents sea (see Table 1) the directional spectra $S(\omega_i, \theta_j)$, calculated in more than 1200 marine grid points. Each spectrum has 24 values in direction θ_i (step 15^0) and 25 in frequency. Total number of spectra is 52 million, and bulk of data is $3.15 \cdot 10^{10}$ number.
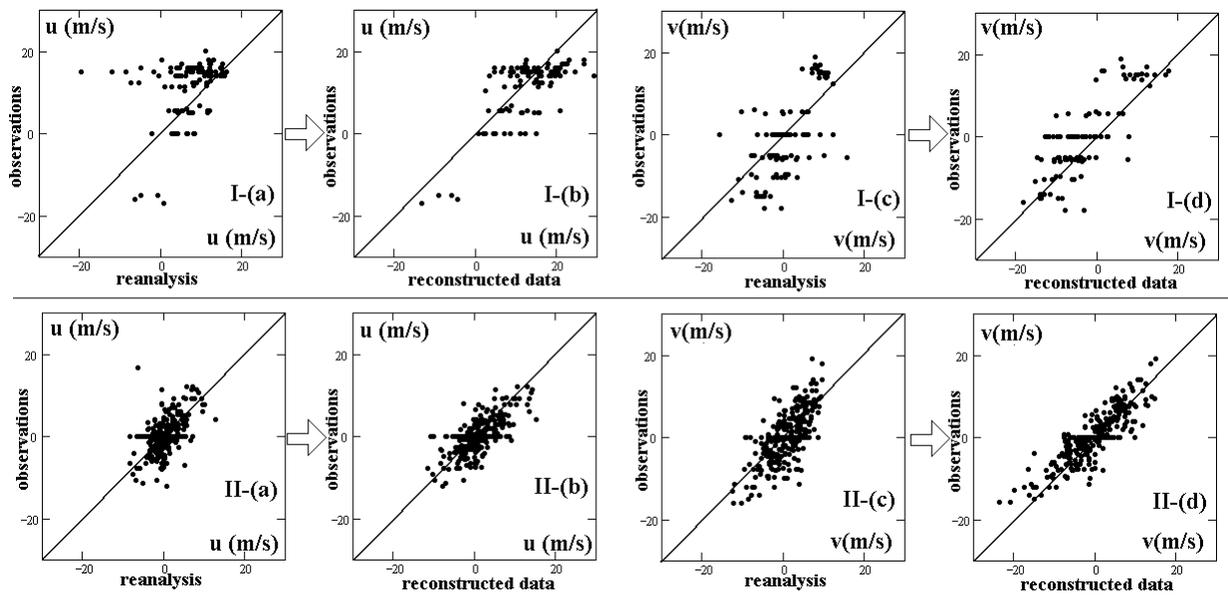
Fig. 3. Scatterplots of observations and reanalysis data for Krasnovodsk (I) and Tuleny (II).
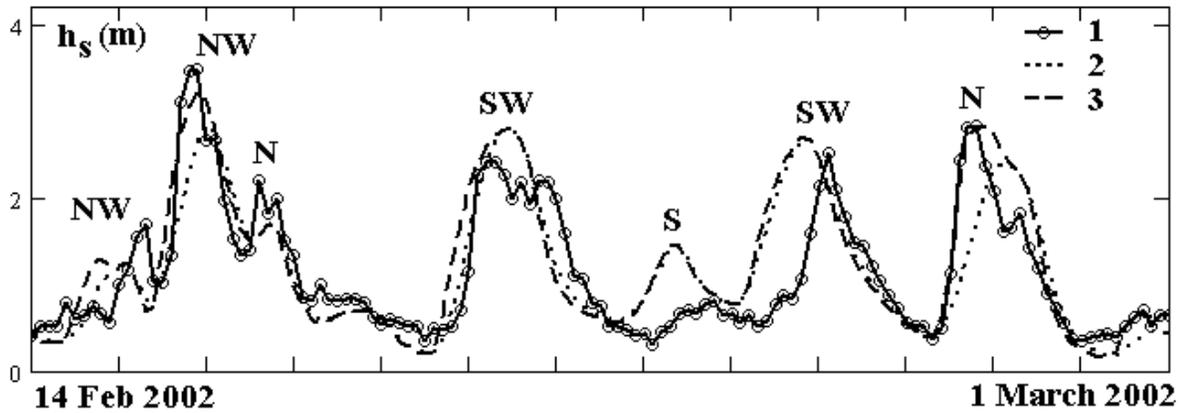(a,c) –without assimilation, (b,d)- with assimilation.

Fig. 4. Parts of significant wave height $h_s(t)$ time series. North Caspian, point 44.10N, 48.49E (depth 24 m).
1 – measurements. 2 - SWAN with step 1 hour. 3 – SWAN with step 15 minutes.
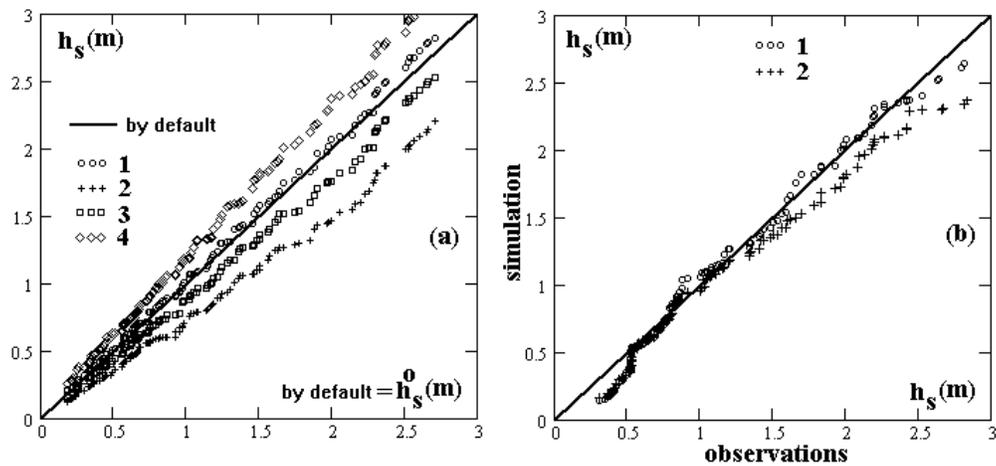

Fig. 5. Q-Q plots of SWAN calculations for N. Caspian (February, 2002).
(a) – values from table 2, (1-4 – the same as in table 2), (b) – different time steps; 1- 15 minutes, 2 – 1 hour.

(a) 13 Feb 1952, 09.00 GMT

(b) 02 March 1952, 12.00 GMT

(c) 21 Nov 1957, 03.00 GMT
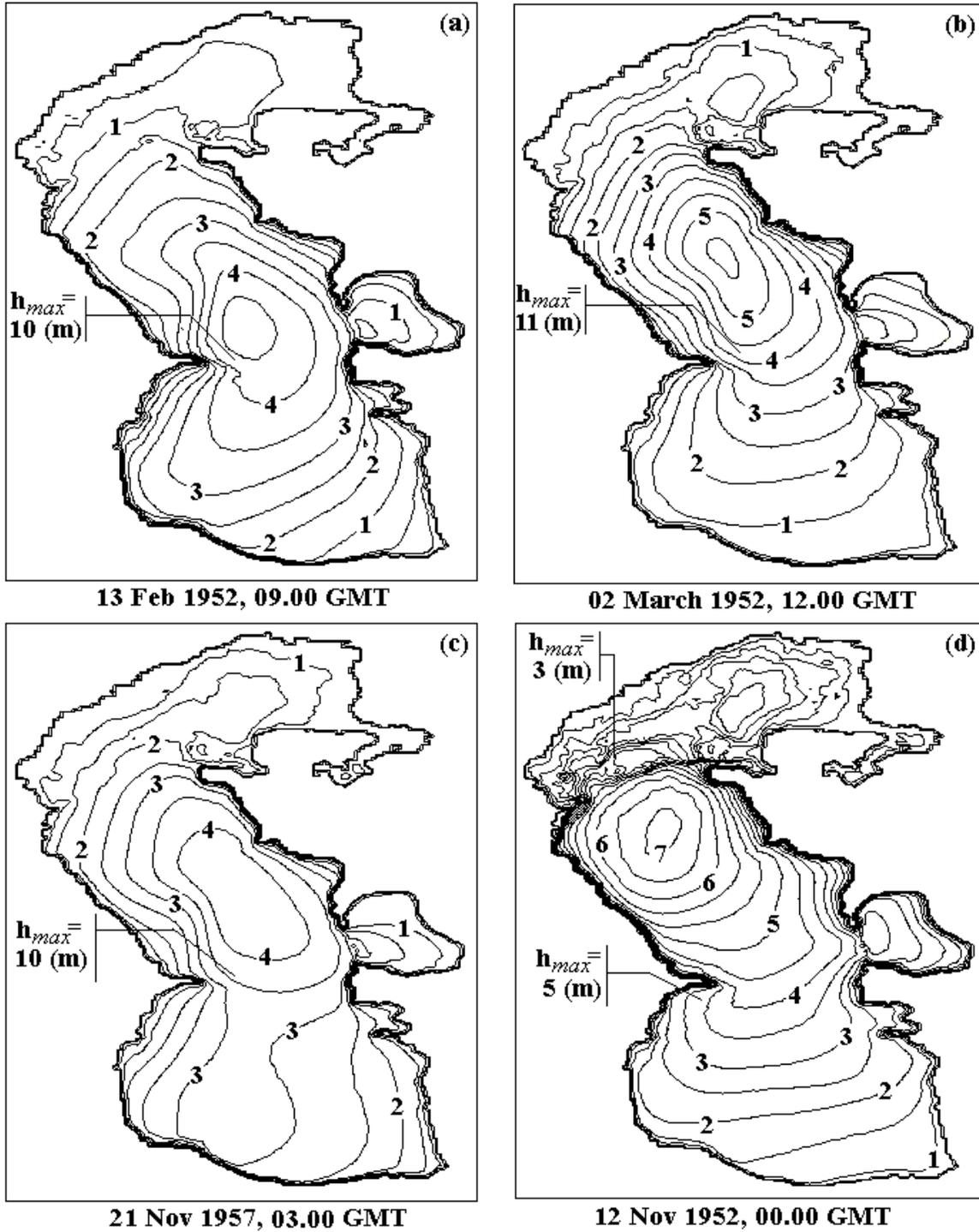
(d) 12 Nov 1952, 00.00 GMT

Fig. 6. Significant wave heights during peak of sever storms. The highest *observed* wave heights also are pointed.

Really to any ensemble of calculated $S(\omega,\theta,\vec{r},t)$ we may set mathematical expectation $m_s(\omega,\theta,\vec{r},t)$ and, if needed, covariance function $K_s(\omega_1,\theta_1,\omega_2,\theta_2,\vec{r}_1,\vec{r}_2,t_1,t_2)$, of $\{\omega_i,\theta_i,\omega_k,\theta_k\}$, but even in stationary (for t) case this statistic is characterized by more than 50 millions of number. High measure of these values and the complexity in interpretation induces to adopt some simplifications in calculations of wave climate statistics. This simplification is based on the parameterization of the directional spectra and further evaluation of the ensemble of these parameters.

The parameterization allow to write wind wave, $S_{ww,}$ and swell spectra S_{sw}, as non-random functions of a set of random arguments Ξ:

$$S(\omega,\theta) = S(\omega)Q(\omega,\theta) \overset{def}{=} S(\omega,\theta \mid \Xi(\vec{r},t)). \quad (5)$$

In the present study as parameters in Ξ are selected wave height, period, spectral shape, frequency of spectral peak ω_{max}, and main wave direction θ_{max}. A single peak wave spectra $S_p(\omega,\theta \mid \omega_{max},\theta_{max})$ is completely determined by these parameters and more general spectra $S(\omega,\theta)$ are obtained as

$$S(\omega,\theta) = m_{00} \sum_{k=1}^{n_{fields}} p_k S_k(\omega,\theta \mid \omega_{max}^{(k)},\theta_{max}^{(k)}), \quad (6)$$

where m_{00}, the zero moment of the spectrum, is equal to the total variance of wave field, and p_k are weight factors for each system so that $\sum_{k=1}^{N} p_k = 1$. Other parameters of spectrum (5) can be expressed as nonlinear functions of spectral moments m_{kj}. The most simple way to determine the parameters $\omega_{max}^{(k)},\theta_{max}^{(k)}$ is the minimization of the deviation index, Liu (1983).

$$DI = \sum_i \sum_j \frac{\left|S(\omega_i,\theta_j) - S_{ij}^*\right|}{S_{ij}^* m_{00}} \underset{\left\langle p_k,\omega_{max}^{(k)},\theta_{max}^{(k)},n_{fields}\right\rangle}{\longrightarrow} min. \quad (7)$$

Where S_{ij}^* are the values of the model output (for frequency ω_i and direction θ_j). For the solution of (7) the adaptive multi-step Monte-Carlo optimization approach is used.

The model (5-7) allows easily to distinct one, two and multipeaked (by variables ω and θ) spectra. Hence, the genetic classification may be presented, see Fig. 8 (for the SW part of the North sea). All the details of classification are published in (Lopatoukhin et al, 2002). The following five classes of wave spectra are selected:

- Wind waves (k=1);
- Swell (k=2);
- Wind waves and "young" swell with close frequencies (k=3);
- Wind wave and "old" swell separated both by frequency and direction (k=4);
- Wind waves and swell without separation (complicate sea) (k=5).

If we associate each class with the stable state of the sea with number k, the synoptic variability of sea waves may be presented as the Markov chain $k = k(t)$ with the transient probability matrix $p_{ij}^{(t,t+1)} = P\{k^{(t+1)} = i \mid k^{(t)} = j\}$, $i,j = \overline{1,m}$

and limit probability vector $\pi_j = P\{k^{(t)} = j\}$, $j = \overline{1,m}$. In the Fig. 8 the transitions between classes are also shown as a "star" diagram, where the arrows correspond to different transient probabilities. E.g., the probability of transition during 3 hours from Wind waves (Class 1) to Wind waves and "young" swell (Class 3) is 5%, and 29% - return. The probability of the cases with the same class after 6 hours, is pointed on the arcs; e.g. for the wind waves this value is 89%. The spatial distribution of limit probability (as unconditional occurrence of each class of the spectra) in the North Sea is shown in the Fig. 7.
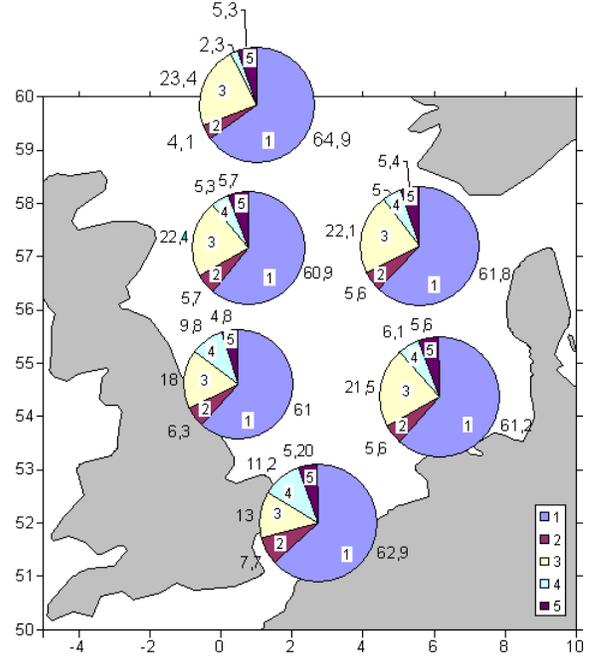

Fig. 7. Spatial distribution of the occurrence of 5 classes of directional spectra in the North Sea.
1÷5 are the classes of spectra

It is clearly seen, that the wind waves are prevailing in over the sea. The occurrence of complex sea with "fresh" swell is decreased from North to South. The model (5-7) allows to estimate the probability characteristics of directional spectra $S(\omega,\theta)$ – mean value, r.m.s, probability, tolerant and confidence intervals of the spectra by means of the correspondent characteristics of the parameters Ξ. E.g., in the Fig.9 the results of the estimation of mean spectra with 70% probability intervals for each class, for SW part of the North Sea, are presented. All the directions are showed in Wave Watch III notation (zero is the East, and rotation counter clockwise).

Thus, the parameterization procedure (5) is the powerful tool for the decreasing of the data dimensionality. It allows reducing the analysis of the set of directional spectra to the spatio-temporal random fields of its parameters $\Xi(\vec{r},t)$. In this paper we only consider the most basic parameter, the significant wave height $h(\vec{r},t)$, defined as four times the standard deviation of the surface height.
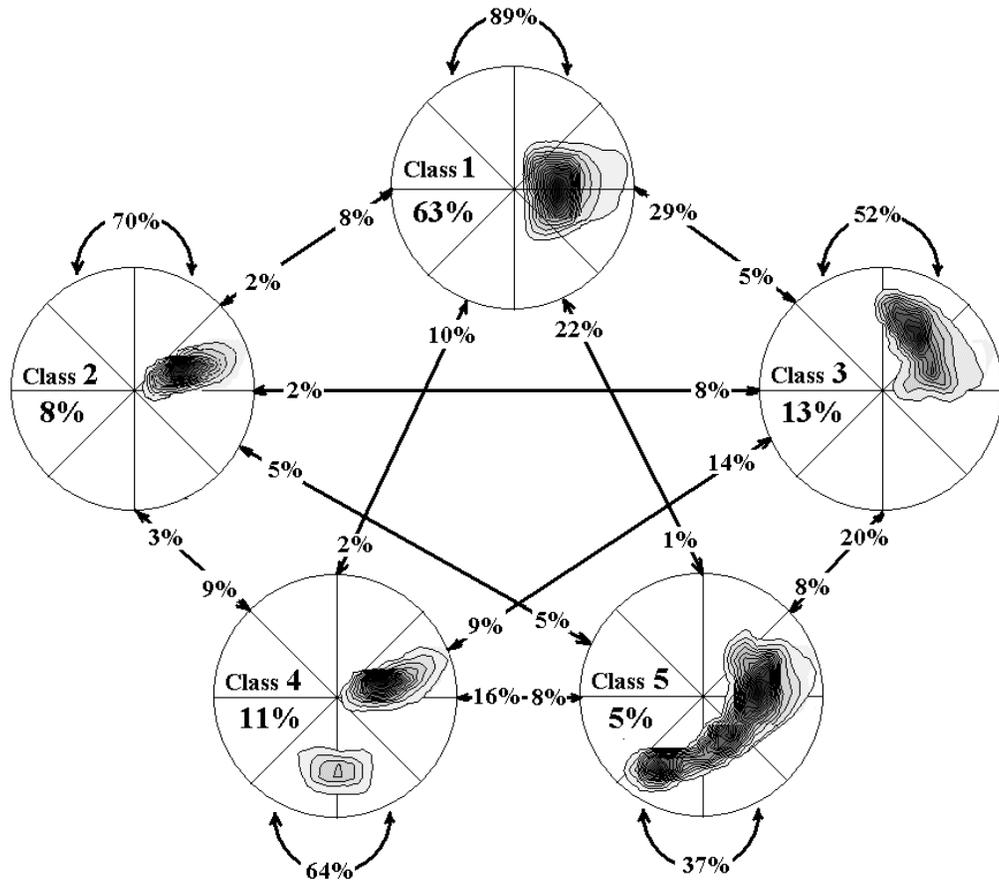
Fig. 8. Transient "star" diagram for directional spectra variability. SW part of North Sea.
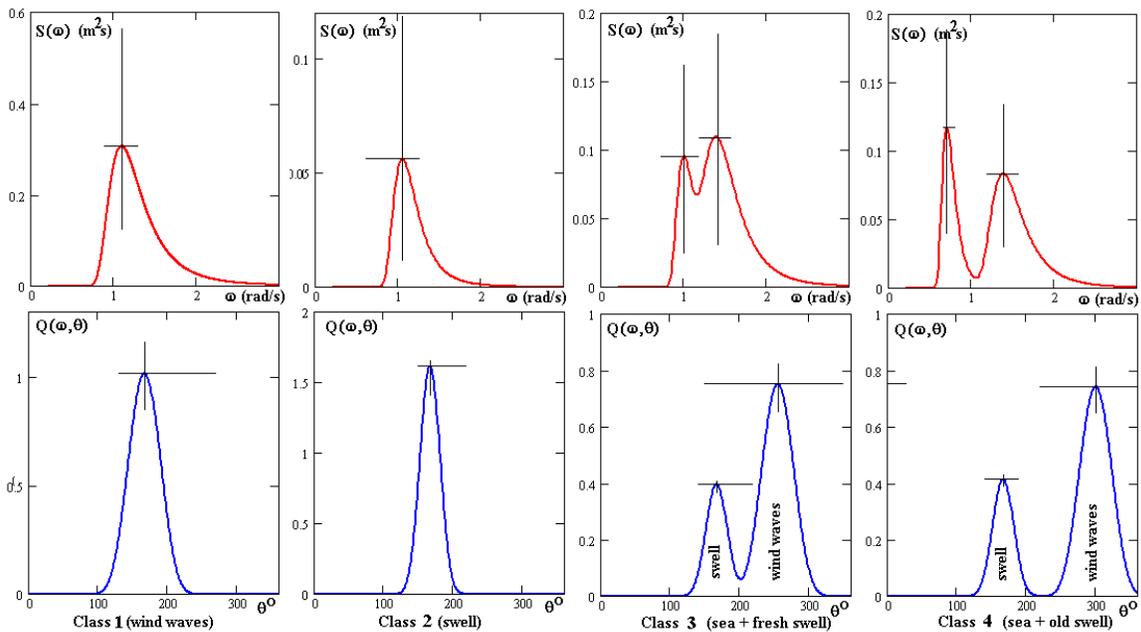

Fig. 9. Mean values and 70% probability interval for frequency spectra and angular distributions of directional spectra in Eq. (5). Classes 1-4, SW part of the North Sea.

5. MULTISCALE STOCHASTIC MODELING OF WIND WAVE CLIMATE WITH APPLICATION TO EXTREMES

Metocean fields, like ocean waves, have a complex spatial and temporal variability (synoptic, annual, year-to-year). Traditionally, the approach for statistical formalization of such phenomena has been based on a *multiscale* hypothesis as proposed by Andrey Monin (1986). The hypothesis suggests modelling the total variability by means of a set of stochastic models for each temporal scale separately, and with the interdependence taken into account parametrically.

5.1. Synoptic variability

The synoptic variability corresponds to temporal scales from a few hours to some days. For atmospheric processes, the associated spatial scales are 2000–3000 km, as opposed to 500–700 km for the oceanic ones. The nature of the synoptic variability may be explained as a stochastic alternation between *storms* and *calms* (Lopatoukhin et al., 2002). It is possible to generalize definition of storms to the spatio-temporal domain,

$$\Omega(t) = \left\{ \mathbf{r} : \quad h(\mathbf{r},t) \geq z \right\}, \tag{8}$$

Where z is the level of the storm and additional parameters are defined in Table 3. Some additional explanations are presented on the Fig. 10(a).

Note that $\{h^+, \mathbf{r}^+\}$ characterize the extreme and $\{\bar{h}, \mathbf{r}_0\}$ the general behaviour of the storm in space. The mean, r.m.s. and 95% quantiles for h^+, $\bar{h}$, and L show that the parameters of the storms are strongly dependent on z and have a clear annual variability.

Table 3. Parameters of a storm.

Description	Notation	Definition
Area	$S_\Omega(t)$	$\int_{\Omega(t)} d\mathbf{r}$
Equivalent diameter	$L(t)$	$2\sqrt{S_\Omega(t)/\pi}$
Averaging wave height	$\bar{h}(t)$	$\int_{\Omega(t)} h(\mathbf{r},t)d\mathbf{r} \Big/ S_\Omega(t)$
Geometric centre ("centre of gravity")	$\mathbf{r}_0(t)$	$\int_\Omega h(\mathbf{r},t)\mathbf{r}d\mathbf{r} \Big/ \int_\Omega h(\mathbf{r},t)d\mathbf{r}$
Maximum wave height	$h^+(t)$	$\max_{\mathbf{r}\in\Omega(t)}\left[h(\mathbf{r},t)\right]$
Location of the maximal wave height	$\mathbf{r}^+(t)$	$\left\{ \mathbf{r} : h(\mathbf{r},t) = h^+(t) \right\}$

Fig. 10 shown these parameters, marked on the map. The storm velocity $\mathbf{W} = \partial \mathbf{r}_0 / \partial t$, and although the mean velocity of the storms is only 3.4-11.9 (km/h), the variations may reach 50 (km/h) with rather high variability.
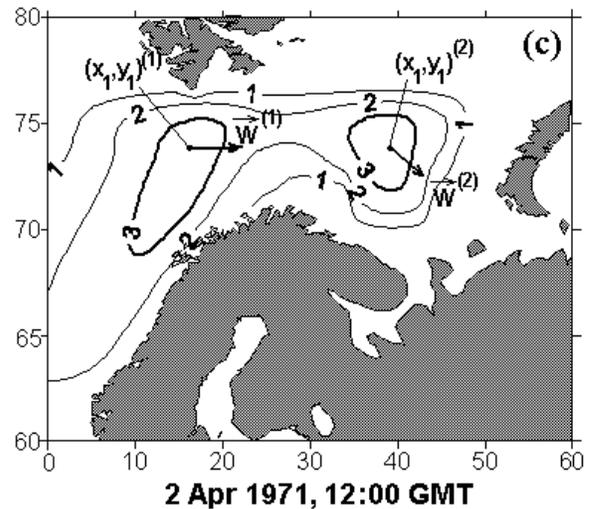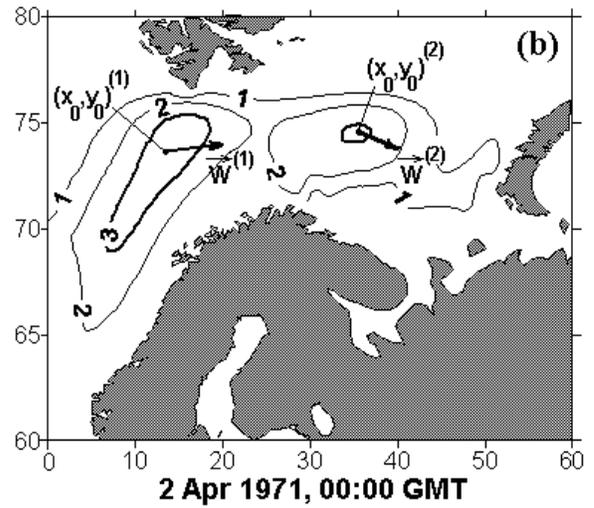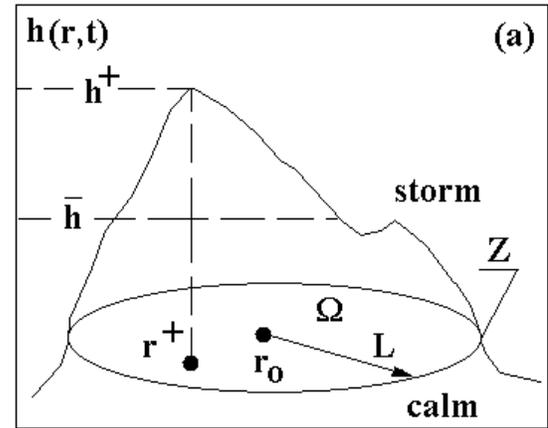

Fig. 10. Illustrations of the storm impulses parameterisation

In general, only one storm occurs at a time, and the position of highest wave is not far from the storm's geometrical centre, $\mathbf{r}_0$. Moreover, $\{h^+, L\}$ may be applied to define the

storm's spatio-temporal behaviour. It allows simplify the solution of stochastic differential equations, driving the variability of sea wave fields

$$\frac{\partial h}{\partial t} + \mathbf{W} \cdot \nabla h = G(\mathbf{r}, t) .\tag{9}$$

Where $G(\mathbf{r}, t)$ a source function. The solution may be presented as the impulse stochastic process, using Galerkin techniques

$$h(\mathbf{r}, t) = \sum_k a_k(t) \Phi_k(\mathbf{r}, t \mid \Xi_k) ,\tag{10}$$

where $\{\Phi_k\}$ are spatio-temporal basis functions, depending on a set of parameters Ξ_k, and $\{a_k\}$ the random corresponding coefficients. The Lagrange approach considered below writes the field as moving *spatio-temporal impulse structures*, and estimates the characteristics $a_k(t), \Xi_k$ directly from the initial data set, without considering the source function.

Let us consider a storm $\{\mathbf{r}_0(t), h^+(t), S_\Omega(t)\}$, $t \in [t_0, t_0 + \Im]$. Our aim is to parameterize the storm impulses in terms of the overall maximum wave height H^+ and the associated storm area S^+. This parameterization, given in terms of a set of parameters $\{H^+, S^+, \Im\}$ and the Markov process for $\mathbf{r}_0(t)$, generalizes the BOLIVAR approach (Lopatoukhin et al., 2002b) from time series to spatio-temporal fields. The data have shown that $(H^+, \Im)$ and (H^+, L^+) are highly dependent: (correlation is 0.7–0.9).

Since H^+ is an extreme value, conditional on $\Im$ distribution is approximated by 1st limit (Gumbel) distribution: this has been validated in (Rozhkov et al., 1999).

$$F\left(H^+ \mid \Im\right) = \begin{cases} \exp\left[-\exp\left(-\left(H^+ - A(\Im)\right)/B(\Im)\right)\right], & H^+ \geq z, \\ 0, & H^+ < z, \end{cases}\tag{11}$$

Relation (11) is valid for deep water, though infinitely large waves may arise. For shallow water third parameter, connected with limit wave height must be introduced. Then, instead of (11), we have

$$F(H^+ \mid \Im) = \begin{cases} \exp\left[-\exp\left[-\dfrac{H^+ - A(\Im)}{B(\Im)}\right] + C\right], & z \leq H^+ \leq h_{\lim} \\ 1, & H^+ > h_{\lim} \\ 0, & H^+ < z \end{cases}\tag{12}$$

Parameter C is connected with limit wave height h_{lim} in specific place:

$$C = \exp\left[-\frac{h_{\lim} - A_1}{B_1}\right].\tag{13}$$

Limit wave height is different for two limiting situations. The first one is propagation of the waves from deep water to shallow. The second – wave development in a shallow water basin. (Classical example is closed Azov Sea with extreme depth about 13 (m), or N. Caspian in the cases of northern winds). For the first case, on the grounds of waves of limited amplitude (Sarphaya, Issakson, (1981)):

$$\frac{h_{\lim}}{g\tau^2} = C_1 \tanh\left[C_2 \frac{H}{g\tau^2}\right]\tag{14}$$

here τ – wave period. Coefficients in (14) are $C_1 = 0.02711$ и $C_2 = 28.77$.
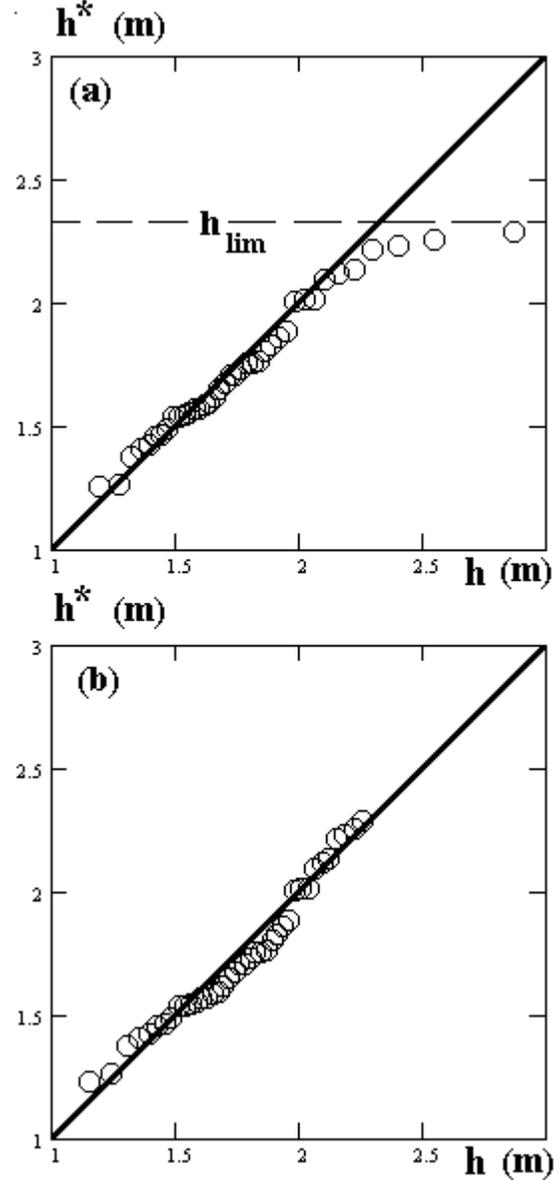

Fig. 11. Quantile biplots of maximal waves in a storm H^+ for shallow water of the North Caspian Sea (depths ~12 m). (a) – Distribution (11), (b) – distribution (12)

In the second case due to some Russian Handbooks and Design Codes:

$$\bar{h} = 0.062 V^{0.4} H^{0.8} ,$$
$$\text{or}\tag{15}$$
$$g\bar{h}/V^2 = 0.06(gH/V^2)^{0.5} ,$$

here $\bar{h}$ - mean wave height, H- water depth, V- wind speed.

Fig.11 definitely shows the difference of approximations (11) and (12) for shallow water.

It now remains to specify a time function for $h^+(t)$ and a spatial field function $\Phi(\mathbf{r})$ as in Eq. (10). It is shown in (Lopatoukhin et al., 2000) that a piecewise linear function in time is sufficient, whereas for $\Phi(\mathbf{r})$ it is possible to use an elliptic cone (1st order), or a elliptic paraboloid (2nd order) (Boukhanovsky et al., 2003b). This parameterisation, given in terms of a set of parameters $\{H^+, S^+, \mathfrak{I}\}$ and the Markov process for $\mathbf{r}_0(t)$, generalizes the BOLIVAR approach from time series to spatio-temporal fields.

5.2. Annual and year-to-year variability

The storm threshold in Eq. (12) really varied in time and space due to annual and year-to-year variability. In this case the scalar analogue of Eq. (3) in Eulerian form is used:

$$\zeta(\mathbf{r},t) = m(\mathbf{r},t) + \sum_k a_k(t)\varphi_{kt}(\mathbf{r},t) + \varepsilon(\mathbf{r},t), \qquad (16)$$

where $m(\bullet)$ is mean value of wave heights field, $\varphi_k(\bullet)$ are the empirical orthogonal functions, and $\varepsilon(\mathbf{r},t)$ is residual white noise (both in space and time). The noise variance is typically not more than 20% of σ_ζ^2. The stochastic processes $a_k(t)$ may be presented in the form of independent scalar autoregressive models:

$$a_k(t) = \sum_{j=1}^{p} \Phi_j^{(k)} a_k(t-j) + \delta^{(k)}(t). \qquad (17)$$

Here the $\Phi_j^{(k)}$ coefficients are calculated from the covariance function $K_{a_k}(\tau)$, and $\delta^{(k)}(t)$ is a constant variance Gaussian white noise, only dependent on k. The relations (15) and (16) constitute a stochastic model of the inhomogeneous (by $\mathbf{r}$) and periodically correlated (by t) random field $\zeta(\mathbf{r},t)$ of significant wave height.

6. EXTREMES IN THE POINT AND IN THE SPACE

6.1. Extremes in a point

There are a lot of approaches to calculations of extreme wave heights in a point. The main are IDM (Initial Distribution Method), AMS (Annual Maxima Series), POT (Peak Over Threshold) and BOLIVAR. Their advantages and disadvantages are investigated elsewhere, Lopatoukhin at al (2000). Short resume with the example for one region of the Mediterranean is presented in the fig 12 and explained below. Comparison of various approaches for estimation of wave heights is shown in the table 4.
IDM method estimates the extreme wave height h_{max} of certain return period as quantile h_p of wave height distribution F(h) with probability p (see fig. 12a). For log-normal long-term wave height distribution, the quantile with probability p can be computed as follows:

$$h_p = h_{0.5} \exp\left(\frac{U_p}{s}\right). \qquad (18)$$

U_p is quantile of the standard normal distribution. Here quantile h_p should be understood as wave height, which is likely to be observed once (at the standard synoptic observation times) in T years. In applied studies the period T is called "return period", and the corresponding probability is defined as

$$p = \frac{\Delta t}{24 \cdot 365 \cdot T}. \qquad (19)$$

Where Δt is interval (in hours) between subsequent observations (say, 6 hours). Then we get *p = 0.000684/T*. For *Δt = 3(hr)*, we get *p=0.000342/T*.
AMS approach defines h_{max} as the last term of the ranked independent series of wave heights h (see fig 12b). Thus it is a random value with Gumbel distribution

$$F(x) = \exp(-\exp(-a(x-b))), \qquad (20)$$

where a, b – parameters.
In the POT approach the k strongest storms with the heights greater than selected threshold. In the fig 12c, threshold is 4.5m. Thus, the POT method estimates depend on the choice of threshold and approximations for corresponding distributions. Unlike other methods, in the POT approach the uncertainty is connected both with the wave height h_p^* and return period. For example, the 25-year wave height estimate in fig. 12d is found to be in the range of 7.2 – 8.4 m, and return period is in the range of 20-45 years.
BOLIVAR approach considered *n* samples, consisting of heights h_{ij}^+ of the largest waves in the *k* the strongest storms in year number *i,*(i=1,..,n; j=1,…,k) BOLIVAR approach exclude the limitations of the POT method and take into account the asymptotic characteristics of AMS.

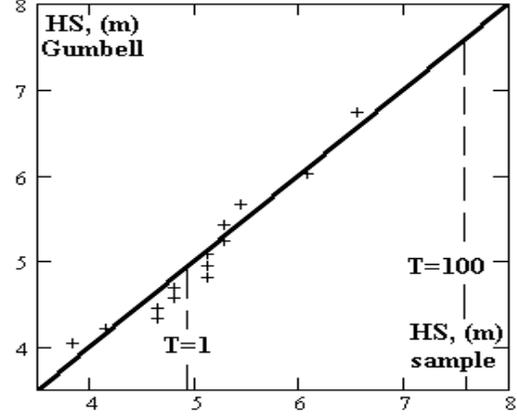Table 4. Values of extreme (significant) wave heights, calculated by various approaches. NW Mediterranean.

| Approach	T, years				
	1	10	25	50	100
IDM	5.2	7.0	7.8	8.5	9.1
AMS	4.8	6.0	6.6	7.1	7.6
POT	5.5	5.9	6.3	6.7	7.0
BOLIVAR 1st maxima	4.8	6.0	—	7.1	7.6
BOLIVAR 2nd maxima	3.8	4.8	—	5.5	5.8
BOLIVAR 3rd maxima	3.4	4.2	—	4.7	4.9

The AMS method has the most solid theoretical foundation. The BOLIVAR method represents its further development that includes into consideration the second, third and, other maximums in a year. Each of the considered methods has its advantages and disadvantages and has to be used accordingly.
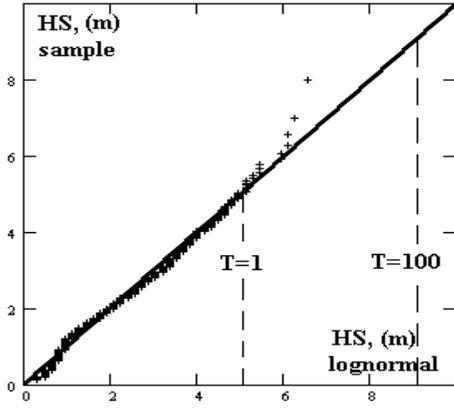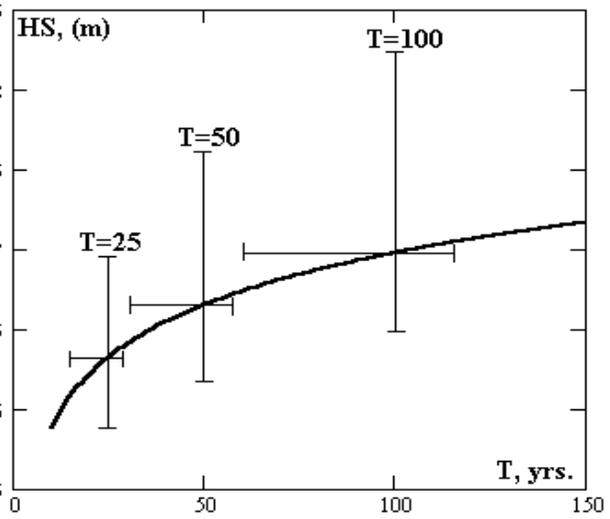
(a)

(b)
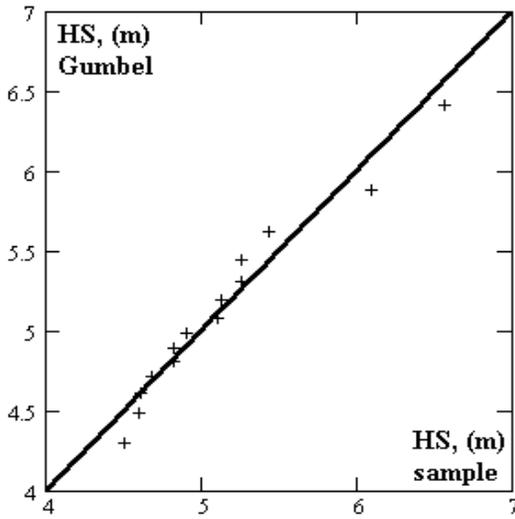
(c)

(d)

Fig.12. Distributions of extreme (significant) wave height HS (m).
IDM approach (a), AMS (b), POT (c,d). NW Mediterranean, 1969–1984.

6.2. Extremes in a field

Storm evolution in any basin may be presented as an impulse random field (10). At any time the impulse $\{\Phi_k\}$ can be presented as an elliptic cone. The size of the storm $\{r_0(t), h^+(t), S_\Omega(t)\}$ is equal to the fraction of total area of the region, where wave heights larger than z, see Table 3. Parameterize of storm in a space impulses in terms generalizes the BOLIVAR approach from time series to spatio-temporal fields.

The behaviour of the extreme wave in a single storm in a fixed point is known, Boukhanovsky et al (1998). For spatial

region this problem more complex, because unique enumeration available only for two-dimensional waves. In the simplest case, with a narrow angular spreading of sea waves, the generalized distribution of maximal wave in a spatial storm region is

$$F_m(h) = \exp\left[2\pi\int_0^L\left(-\exp\left(-\frac{\pi}{4}\left(\frac{h}{\overline{h}(r)}\right)^2\right)\right)\frac{r\,dr}{\lambda(r)}\right] \quad (21)$$

Here 2L is the equivalent diameter of the storm, where $L = 2\sqrt{S_\Omega(t)/\pi}$, $S_\Omega(t) = \int_{\Omega(t)} d\mathbf{r}$. For small-amplitude waves

$\lambda(r) \approx 36\overline{h}(r)$. The storm impulse $\overline{h}(r)$ is approximated by expression

$$\overline{h}(r) = h^+ - (h^+ - z)(r/L)^m, \qquad (22)$$

where m is the shape parameter of storm impulse (m=1 – cone, m=2 – parabolic etc).

In first step simulation procedure consist of estimation the parameters for each synoptic term t. The level z may depend on the season and obtained from models of annual variability, as described in (Boukhanovsky et al., 2003b). The second step is the spatio-temporal impulse parameterisation of the time series $\{h^+(t), L(t)\}$, the parameters of equation (10), see table 3, and the parameters (transition and limit probabilities) of the Markov chain model for $\{\mathbf{r}_0(t)\}$.

The simulation starts with the Monte-Carlo simulation. Based on the realizations, the durations $\{\mathfrak{I}_k\}$ are found and the value of H^+ is simulated from Eq. (11) or (12) The diameter L^+ is obtained of a regression on H^+, and finally, all synthesized parameters are substituted into Eq. (10). Generating a synthetic field for all points $(\mathbf{r}, t)$, where the shape of the storms is approximated by (22). Thus, the final results of the simulation are the set of spatiotemporal fields with the same stochastic properties, as the initial data. It allows applying such approach for estimation of non-observable extreme events.

The main question is: does the spatiotemporal field model (10) reproduce the estimates of T-years extremes in the fixed points estimated by approaches like BOLIVAR or AMS (Lopatoukhin et al., 2000b)? For answering this questions the extremes for return period of 1, 10, and 100 years were calculated for the points 1-3 in Fig. 13 by the AMS approach (with parametrical confidence intervals) and by the Lagrange stochastic simulation (see Table 5). All simulations were carried out for two types of the spatial impulse approximation $\Phi(\mathbf{r})$ in (10), see also (22). The 1ˢᵗ order shape is an elliptic cone, and in this case the model (10) underestimates the T-years extremes. By assuming the more reasonable form in (22) of an elliptical paraboloid for $\Phi(\mathbf{r})$, the values become in better agreement with the AMS method.

Table 5. Sample and simulated values of *T*-year wave heights in Barents Sea.

Point #		1		2		3	
Return period T		1	100	1	100	1	100
Sample estimates	Point data	10.3	16.7	9.6	13.7	9.0	13.2
	90% CI*	9.3 11.8	15.0 19.2	8.6 11.0	12.3 15.8	8.1 10.4	11.9 15.2
Simu-lation	1ˢᵗ order	8.0	13.0	6.9	9.0	6.7	8.5
	2ⁿᵈ order	9.8	15.1	9.2	14.5	8.4	13.4

CI is the confidence interval

Fig. 13 shows the extreme wave heights (0.1% probability) with return period 100 years. This figure (as similar, published in different papers, handbooks an atlases) is a result of calculation at various points and driving isolines. Data of such figure represent the wave heights estimates that are possible in any point, *but not in all points simultaneously*. In the last case the return period of such events will more rare, than 100 years.

This argument became clearer from the Fig. 14. There are shown annual maxima $h_s^{(A)}$ in the point "A" and conditional values $h_s^{(B/A)}$ in the points B₁ and B₂. They are at distance 120 and 240 km from point "A". It is seen, that in spite of significant distance between points, the values of wave heights of rare probability are dependent. This also means that the same return period may appropriate to different combinations of waves. E.g., on the Fig 14a, 100 years wave in the point "A" is 14.4m, and then wave height in the point "B₁"is 13.7m (i.e. with return period 50 years). Another conclusion from the Fig. 14 is that 100-year event may result from a set of events each of it is less than 100 year. E.g., 100 years event in the field will be when simultaneously:

- in the point "A" $h_s^A = 12.1m$ (10 years return period);
- in the point "B₁" $h_s^{B_1} = 13.2m$ (30 years return period);
- in the point "B₂" $h_s^{B_2} = 13.8m$ (60 years return period).

The results of the model verification are satisfying and give confidence to use the model for analysis and numerical studies of spatio-temporal variations of extreme synoptic events.
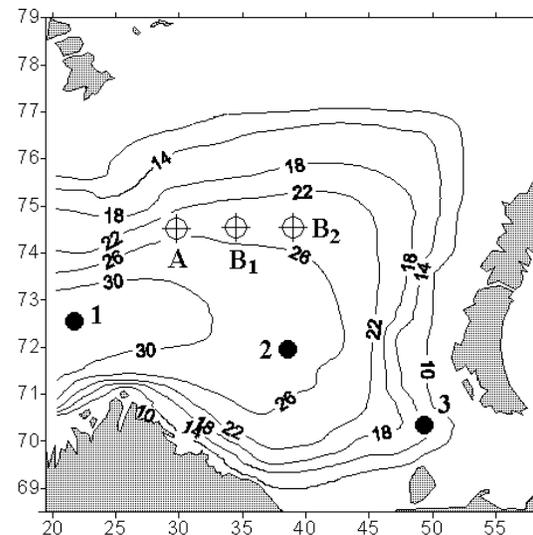

Fig. 13. Spatial estimates of maximal waves (0.1%) once 100 years in the Barents sea.

Complete solution of the problem of estimation of extreme waves joint variability calls for multiscale stochastic models, which simulate ensemble of spatiotemporal fields in the scales of synoptic, annual and year-to-year variability. This is the task of further investigations, but results similar to presented at Fig. 13, 14 and their interpretation allows solving numerous applied problems.
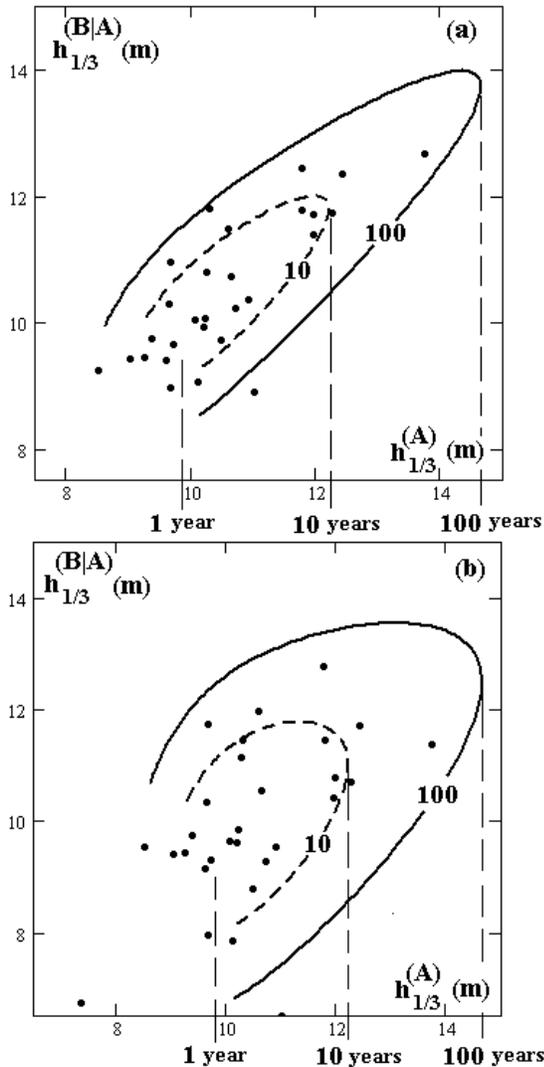
Fig. 14. Points and isolines of return periods of annual maxims. Significant wave heights $h_{1/3}^A$ in the point "A" and conditional wave heights $h_{1/3}^{B|A}$ in the points «B₁» (a) and «B₂» (b), at the same time.

REFERENCES

Akima H. (1978) A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points, ANC Transactions of Mathematical Software, 4, 1978, pp. 160-164

Boukhanovsky A.V., Lopatoukhin L.J., Ryabinin V.E. (1998). Evaluation of the highest waves in a storm./ Marine Meteorology and Related Oceanographic Activities/ World Meteorological Organization. Report N 38, WMO/TD -N 858, 1998, 19p.

Boukhanovsky A.V., Krogstad H.E., Lopatoukhin L.I., Rozhkov V.A. (2003a). "Stochastic simulation of inhomogeneous metocean fields. Part I: Annual variability". *Proc. Of ICCS'03, Lectures Notes in Computer Science. LNCS 2658*, pp. 213-222.

Boukhanovsky A.V., Krogstad H.E., Lopatoukhin L.I., Rozhkov V.A., Athanassoulis G.A., Stephanakos C.N. (2003b). "Stochastic simulation of inhomogeneous metocean fields. Part II: Synoptic variability and rare events". *Proc. Of ICCS'03, Lectures Notes in Computer Science. LNCS 2658*, pp. 223-233.

Cyclostationarity in communications and signal processing. (1993). Ed. W.A. Gardner, IEEE Press, 1993.

Gill M., Malanotte-Rizoly P. (1991). Data assimilation in Meteorology and Oceanography. Adv. Geophys, v.33, p.141-266.

Graham C., Cardone V.J., Ceccacci, E..A., Parsons M.J. Cooper C., Stear J. (2002). Challenges to wave hindcasting in the Caspian Sea. 7th Int. Workshop on Wave Hindcasting and Forecasting. 2002. Banff, Alberta, Canada.

Himmelblau D. (1970). Process analysis by statistical methods. John Willey & Sons, Inc., NY, 1970.

Liu P.C. (1983). A representation for the frequency spectrum of the wind-generated waves. Ocean Engng., vol. 10, No 6, 1983, pp. 429-441.

Lopatoukhin L.J., Boukhanovsky A.V., Rozhkov V.A. (2000a) "Approach and method of calculation of extreme wind waves in a deep and shallow water". *Proc. Int Conference LITTORAL 2000. "Responsible Coastal zone management. The Challenge of the 21st Century". Period Biol,* vol. 102, Supplement 1, Zagreb, p.513-517.

Lopatoukhin L., Rozhkov V.A., Ryabinin V.E., Swail V.R, Boukhanovsky A.V., Degtyarev A.B. (2000b). "Estimation of extreme wind wave heights" // *World Meteorological Organisation (WMO).* WMO/TD-No. 1041, 2000, JCOMM Technical Report.

Lopatoukhin L., Rozhkov V., Boukhanovsky A., Krogstad H., Athanassoulis G., Stephanakos C., Degtyarev A. (2002) "The spectral wave climate in the Barents sea". *Proceedings of Int. Conf OMAE'02*, June 23-28, 2002, Oslo, Norway.

Monin A.S. (1986) "An Introduction to the Theory of Climate". *D. Reidel*, 1986.

Ogorodnikov V.A., Prigarin S.M. (1996) Numerical modeling of random processes and fields: algorithms and applications. *VSP, Utrecht, the Netherlands*, 1996, 240 p.

Rozhkov V.A., Boukhanovsky A.V., Lopatoukhin L.J. (1999). "Method for calculation of extreme metocean events". *Proc. Int. Conf. MEDCOAST'99*, Turkey, 1999

Sarpkaya, T., and Isaacson, M. (1981) "Mechanics of Wave Forces on Offshore Structures", *Van Nostrand Reinhold,* New York.