Atmospheric and Climate Sciences, 2016, 6, 205-223
Published Online April 2016 in SciRes. http://www.scirp.org/journal/acs
http://dx.doi.org/10.4236/acs.2016.62018

An Extreme Value Analysis of Wind Speed over the European and Siberian Parts of Arctic Region

Alexander Kislov, Tatyana Matveeva

Department of Meteorology and Climatology, Lomonosov Moscow State University, Moscow, Russia
Email: avkislov@mail.ru

Received 11 January 2016; accepted 5 April 2016; published 8 April 2016

Abstract

Multiyear observed time series of wind speed for selected points of the Arctic region (data of station network from the Kola Peninsula to the Chukotka Peninsula) are used to highlight the important peculiarities of wind speed extreme statistics. How largest extremes could be simulated by climate model (the INM-CM4 model data from the Historical experiment of the CMIP5) is also discussed. Extreme value analysis yielded that a volume of observed samples of wind speeds are strictly divided into two sets of variables. Statistical properties of one population are sharply different from another. Because the common statistical conditions are the sign of identity of extreme events we therefore hypothesize that two groups of extreme wind events adhere to different circulation processes. A very important message is that the procedure of selection can be realized easily based on analysis of the cumulative distribution function. The authors estimate the properties of the modelled extremes and conclude that they consist of only the samples, adhering to one group. This evidence provides a clue that atmospheric model with a coarse spatial resolution does not simulate special mechanism responsible for appearance of largest wind speed extremes. Therefore, the tasks where extreme wind is needed cannot be explicitly solved using the output of climate model. The finding that global models are unable to capture the wind extremes is already well known, but information that they are members of group with the specific statistical conditions provides new knowledge. Generally, the implemented analytical approach allows us to detect that the extreme wind speed events adhere to different statistical models. Events located above the threshold value are much more pronounced than representatives of another group (located below the threshold value) predicted by the extrapolation of law distributions in their tail. The same situation is found in different areas of science where the data referring to the same nomenclature are adhering to different statistical models. This result motivates our interest on our ability to detect, analyze, and understand such different extremes.

How to cite this paper: Kislov, A. and Matveeva, T. (2016) An Extreme Value Analysis of Wind Speed over the European and Siberian Parts of Arctic Region. *Atmospheric and Climate Sciences*, **6**, 205-223. http://dx.doi.org/10.4236/acs.2016.62018

Keywords

Extreme Wind Speed Analysis, Arctic Circulation, Modelled Extreme Wind Speed

1. Introduction

This paper focuses on changes in extreme wind events in the Arctic region. For the purpose of ensuring the safety of infrastructure (particularly at exposed sites such as bridges, high buildings, wind turbines and radio masts), it is usually a requirement to estimate the extreme loads they might be subjected to during their service time. In the maritime sectors, the extremes of low-level winds can generate huge oceanic waves and storm surges that consequently may lead to the damage of marine structures (ships, drilling platforms) and coastal erosion. This is especially the case during the cold period in the Arctic, where regular events of intense wind velocities are typically observed. The quantitative analyses of spatial variation of extreme wind patterns are important for effective wild fire management and sustainable long-term urban development on fire-prone landscapes. It is therefore worthwhile to properly assess the distribution of wind extremes and their origin.

Research in the statistical analysis of extreme values has flourished over the past several decades: new probability models, inference and data analysis techniques have been introduced; and new application areas have been explored [1]-[3].

Extreme value analysis of wind speeds (U) is generally performed through implementation of the following idea. Beginning with a parent distribution whose cumulative distribution function (cdf) is $F(U)$, the distribution is sampled n times, and the maximum value of the n samples is obtained. This maximum value has a cumulative distribution function of its own of simply $F^n(U)$. This relationship leads to the extreme value theory noted by Fisher and Tippett [4] that for sufficiently long sequences of independent and identically distributed random variables, the maxima of these sequences can be fit to one of three limiting distributions. This result has been quantitatively refined by Gnedenko [5]. One representative of these three limiting distributions is the Weibull distribution, which has traditionally been used for the statistical modelling of wind extremes (see [6]).

In another approach, the Pareto distribution is applied to the peaks of independent storms that exceed a sufficiently high threshold (see [6] and [7]).

Many studies of the estimation of extreme wind speeds are commonly expressed in terms of the quantile value $U(p)$, or U_T, the maximum wind speed (which is exceeded, on average, once every T years), and the return period (the corresponding return period is given by $T = 1/(1 - p)$). In this situation, the data are generally fitted to a theoretical distribution curve in order to calculate the quantiles. To insure the independency of the data, a certain minimum separation time is maintained among the data selected for the analysis.

Statistical method of the extreme value analysis of wind speeds is important because it allows us to detect their statistical model. Note, that the same statistical distribution suggests a common originating mechanism. We plan to use such idea to interpret of the extreme wind records.

A striking aspect of anticipated global climate change in response to increasing greenhouse gases is that the largest warming is predicted to occur in the Arctic. This observed warming has affected glaciers, sea ice, ecosystems, permafrost and the coastal geomorphology. It is likely that such warming affects meteorological regimes (e.g., extreme conditions). Because the climate models are the tools used to simulate climate change, it is very important to understand to what extent wind speed extremes can be reproduced by a general circulation atmospheric models within those climate models. The use of station data will make it possible to evaluate the consistency, in terms of reproduction of statistical behavior, between model simulation products and near-surface observations.

The next section reviews the study area and dataset. The following sections describe the evidence for Weibull distribution in station observation data and in model data as well. The last section concludes the paper.

2. Study Area, Dataset and the Question of Statistical Independence

The study was performed over the Arctic region from the Kola Peninsula until Chukotka Peninsula including both coastal area (predominantly) and inland territory. Strong wind speed events are often noted in the region in the cold time of year during the passage of meteorological synoptic storms. Wind speeds of more than 30 $m \cdot s^{-1}$

are observed during this time over the marine surface, inducing high waves of more than 4 m.

A dataset of observed hourly 10-minute mean wind speed data from stations was obtained, with the record period varying from station to station. For the present study, we used the period 1966-2013, which was covered by data of all stations (**Table 1**). Their location is shown in **Figure 1**. Note that we do not consider here the matter of anemometer exposure. As with all analyses of wind data, the results of an extreme value analysis will be flawed if the data on which they are based are taken from an anemometer with a non-standard exposure (e.g., sheltered in one or more directions, or at a height above the ground different from the 10 m standard). Metadata for stations can be obtained from Meteorological Service of Russia.

It was interesting to observe exceptional outliers for several values (60 - 70 $m \cdot s^{-1}$) in the dataset. As part of our analysis, we questioned whether to neglect this information as errors or spurious outliers. We investigated the reanalysis dataset to find such values. For this aim, a dataset of 3-hourly 10-m wind of the 20th *Century Reanalysis* dataset (1.9 × 1.875 deg. Lat. × Long.) for the period 1979-2004 was obtained [8] [9]. These data correspond to diagnostic values at the equivalent anemometer level. They differ little from other products [10] that are typically used for the Arctic region. This product was related directly to observations and in all cases we saw that it does not contain observed exceptional outliers. Therefore it seems logical that such velocities should were diagnosed as errors and ignored. However, as will be shown below, in some cases their appearance is possible.

A dataset of wind simulation of the INM-CM4 climate model for the period 1966-2005 was also obtained [11]. These data (1.5 × 2 deg. Lat. × Long.) correspond to the Historical experiment of the Coupled Model Intercomparison Project, Phase 5 (CMIP5) [12]. The use of station data will make it possible to evaluate the consistency between wind simulation products and near-surface observations. Observational data and the INM CM4 data cover not the same years, but it could not influence the statistical results due to specification of used numerical experiment, which focuses on generation of common features of modern climate conditions. Apart from, the appearances of extremes are not regular (**Figure 2**). Note that pictures in **Figure 2(a)** and **Figure 2(b)** are different in spite of the distance between stations is no more than 250 km. In this figures we can identify the long-term clusters of extreme events. Due to short observation period, it is not clear whether this is a sign of climate changes or a trace of long-lasting variations. The period of 1966-2013 characterizes the period of climate change [13]; however, diagnosed events did not appear in statistics of wind speed events.

It is a condition of extreme value analysis that the extremes selected for examination have to be independent. Annual (or seasonal) maximum wind speeds chosen from each year are statistically independent. However, when several data points are taken from each season, there may well be several clustered maximum speeds from a single storm. Such events are unlikely to be statistically independent. Various strategies are invoked to remove dependent events before proceeding with a statistical analysis. A simple method is to require a minimum time separation or "deadtime" between selected events. Working with Arctic wind climate, we use the autocorrelation coefficient $r(\tau)$ to establish a deadtime between consequent wind fluctuations. Its value is a measure of the correlation of neighbouring wind events. It was shown to be less than 0.05 for τ equal to 48 or 72 hours. Therefore, we use a deadtime of 72 hours. The same values (48 - 60 h) were used earlier [14]-[16].

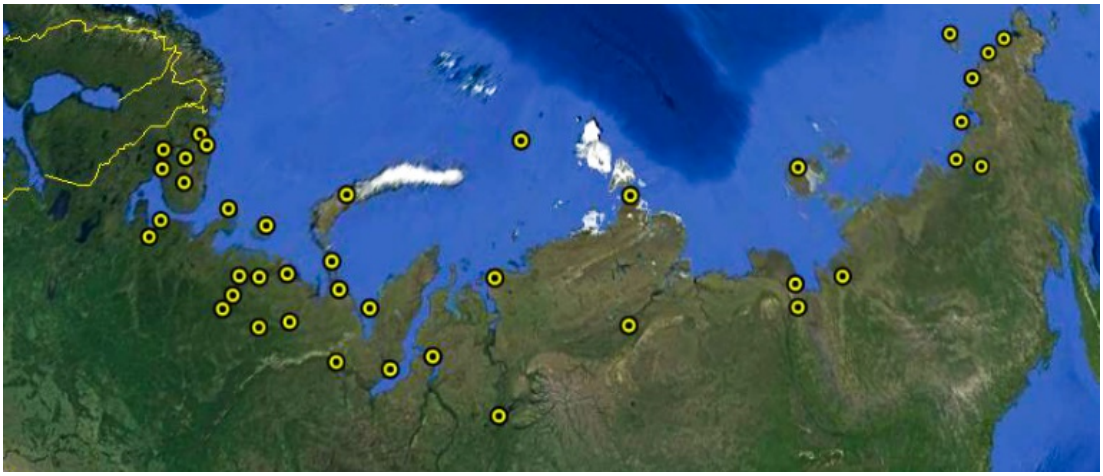
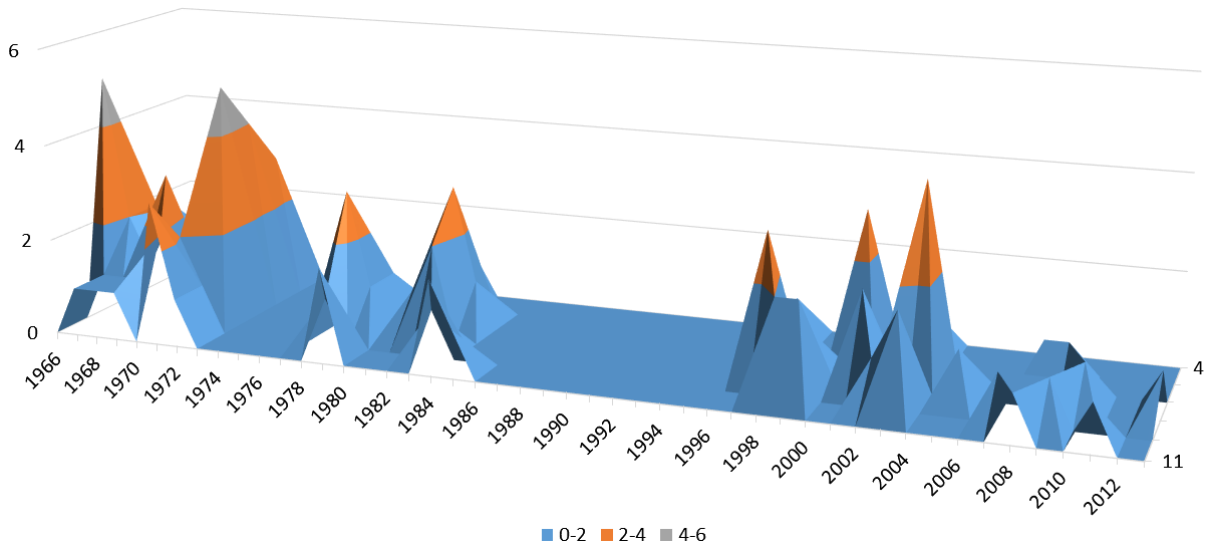


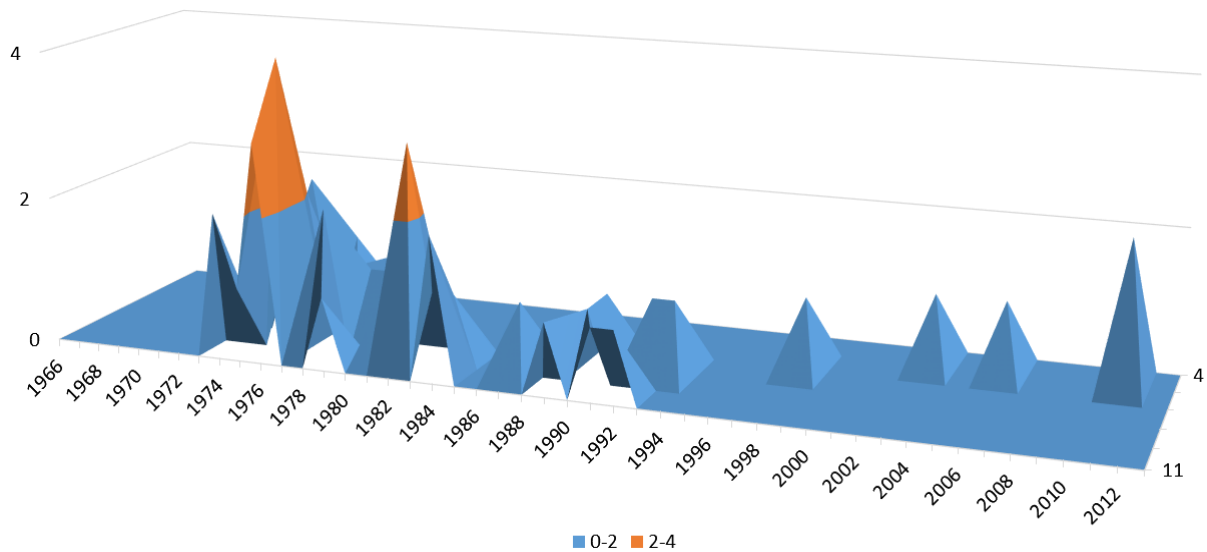
Figure 1. Location of observation stations.

Table 1. Wind measurement stations across the Arctic coastal zone of Russia (http://meteo.ru/).

Station	Lat., N	Lon., E	Height above sea level
Teriberka	69.2	35.1	33
Murmansk	69	33.1	55
Lovozero	68.1	34.8	161
Krasnoshchelye	67.4	37.0	155
Kandalaksha	67.1	32.4	26
Umba	66.7	34.3	39
Arkhangelsk	64.6	40.5	3
Zimnegorsky Mayak	65.5	39.7	85
Cape Kanin	68.7	43.3	48
Kolguyev Island Northern	69.5	49.1	23
Kotkino	67	51.2	18
Naryan-Mar	67.7	53.0	4
Ust-Usa	65.9	56.9	77
Ust-Tsilma	65.4	52.3	78
Okunev Nos	66.3	52.6	20
Hoseda Hard	67.1	59.4	84
Anderma	69.8	61.7	53
MalyKarmakula	72.4	52.7	18
Marresale	69.7	66.8	24
Novy Port	67.7	72.9	11
Antipauta	69.1	76.9	2
Dikson	73.5	80.2	42
Fedorov Observatory	77.7	104.3	13
Bolvansky Nos	70.5	59.1	13
Khatanga	72.0	102.3	30
Vize Island	79.5	77	11
Tiksi	71.6	128.6	6
Wrangel Island	71.0	181.5	2
Cape Konstantinovsky	68.6	55.5	7
The Yubileynaya	70.8	136	25
Vankarem	67.8	183.5	3
Ambarchik	69.6	162.3	23
Cape Schmidt	68.9	180.7	3
Ayon Island	69.9	168.0	13
Ostrovnoye	68.1	164.2	94
Cape Billings	69.9	175.8	2
Salekhard	66.5	66.5	15
Igarka	67.5	86.4	25
Kotelny Island	76	140.5	10
Kyusyur	70.7	127.5	36



(a)



(b)

Figure 2. Appearance of wind speed extremes (>21 m/s), based on wind observation at the Teriberka (a) and the Zimnegorsky Mayak (b) for cold period (from November (number of month is 11) to April (number is 4)) of each year.

3. The Weibull Distribution in Station Observation Data

Because it is widely accepted that the Weibull distribution is a good model for wind speed distributions, empirical extremes are modelled by the cdf:

$$\frac{n}{N} \approx F(U) = 1 - \mathrm{e}^{-AU^k} \tag{1}$$

This expression (stretched exponential distribution) can be replaced by

$$\ln\left[-\ln\frac{N-n}{N}\right] = k\ln U + \ln A. \tag{2}$$

Such representation allows a straight representation of the empirical function on the coordinate axis of the Weibull distribution. The model parameters (A and k) can be estimated using the maximum likelihood approach.

To estimate the success of approximation the coefficient of determination, denoted R^2, is traditionally calculated. It provides a measure of how well observed outcomes are replicated by the model of linear regression, it is the square of the sample correlation coefficient. Such approach allows us to determine (almost visually) whether a simple estimate can approximate expression (1).

In **Figure 3** we plot several empirical cdfs on the bases of station measurements. Configuration of empirical points in the form of columns is determined by the fact that the data are quantized due to specified accuracy of measurement. Pictures are the "Weibull Plots", which are a specific nonlinear transformation of the data, and a straight line is recovered if the sample is Weibull. We can see that at all sites, we found that the empirical cdfs consistently deviate from the theoretical line starting with certain large threshold values (U_{th}). This means that the empirical tail diverges from the Weibull model, indicating that a different model might describe the data well. As a rule, there are only few such values ($U > U_{th}$); however, their presence has profound significance because they are the greatest extremes.

To approximate these empirical cdfs we use the same technique but applying Gumbel distribution. Again, as was expected, we found that the empirical cdfs consistently deviate from the theoretical model.

It is possible in such a situation to choose another function (including three parameters) for approximating the behaviour of observed data. However, we interpret these results in another way. It seems that data indicate that there is a violation of the above-mentioned condition of identically distributed random variables. The shape of the curve suggests that the volume of samples is composed of two sets of variables, each described by its own Weibull function. **Figure 4** shows approximations of the empirical cdfs of wind speed extremes for different regions (same as in **Figure 3**) by two different Weibull distributions. Very large R-squared values (>0.95) denote the high success of this approximation. Coefficients of regression equations allow us to estimate parameters k and A in each cases.

For example, in **Figure 5** we compare empirical cdf of wind speed extremes on the bases of the Teriberka station measurements (only data shown in **Figure 4(a)**) and corresponding Weibull distribution using calculated parameters: $A = 1.64 \times 10^{-5}$ and $k = 3.97$. To decide if samples come from a population with Weibull distribution the special statistical tests could be utilized. First of all note that R-squared values, denoting the success of the maximum likelihood approach, reflects in some aspects the Cramer-Mises-Smirnov (C-M-S) test, because it uses the integral of the squared difference between the empirical and the estimated distribution functions. If $R^2 \to 1$ this means that the integral converge to zero. However, the C-M-S test cannot be used, because the information about the empirical function is incomplete since we have only values corresponding to $U \le U_{th}$. Similarly, the Anderson-Darling criterion cannot be applied because it places more weight on observations in the tails of the distribution which are out of reach. In this case more suitable the Kolmogorov-Smirnov (K-S) test because it uses the supremum of the absolute difference between the empirical and the estimated distribution functions. Here we take into consideration (using forms of regression lines—see **Figure 4**) that the supremum is not located in tail zone.

Because test is applied in contexts where a family of distribution is being tested, in which case the parameters of that family need to be estimated and account must be taken of this in adjusting either the test-statistic or its critical values. Revised critical values for Weibull distribution are given by [17]. Using the K-S test we assess that the null hypothesis, which asserts that the sample come from a population with Weibull distribution, cannot be rejected. Analogous procedure was used to decide if samples depicted in **Figure 4(b)** come from a population with Weibull distribution using another calculated parameters: $A = 0.0120$ and $k = 1.77$. It was concluded again that Weibull distribution fits well to the data (**Figure 5(b)**). The same result was established for other examples of **Figure 4** and in all studied stations (**Table 1**).

Note additionally that application of a sufficiently high threshold and, consequently, detection of especially high wind speeds allows us to describe for their approximation the peaks-over-threshold modelling approach, using the Pareto distribution. It has a cumulative distribution function

$$\Phi(U) = 1 - \left(\frac{U_{th}}{U}\right)^{\gamma}. \tag{3}$$

It is worth mentioning here that the threshold value is not assigned a priori (as is usually done [7]) but is explicitly estimated previously.

Generally, the implemented analytical approach allows us to detect that the majority of extreme wind speed

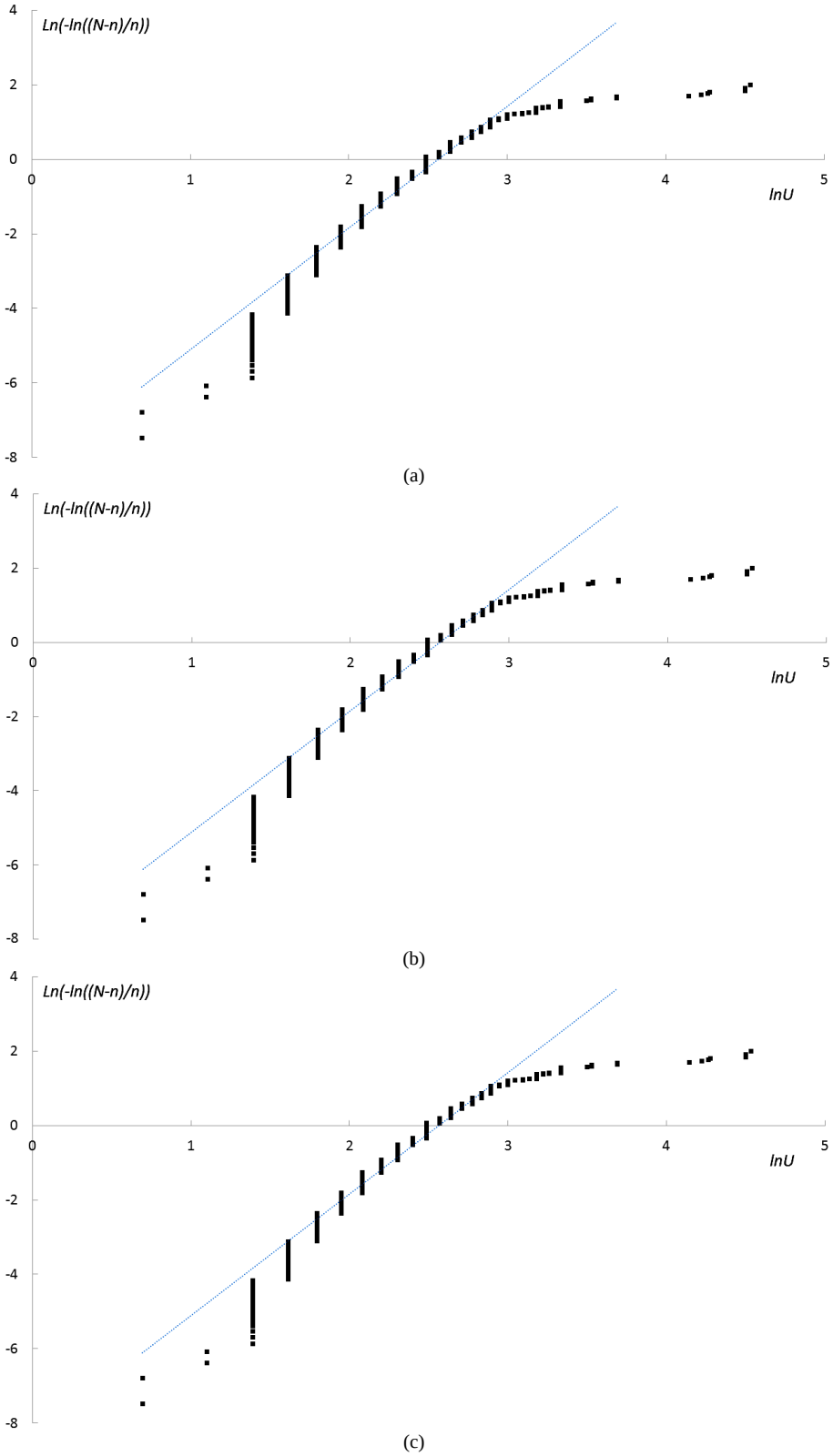
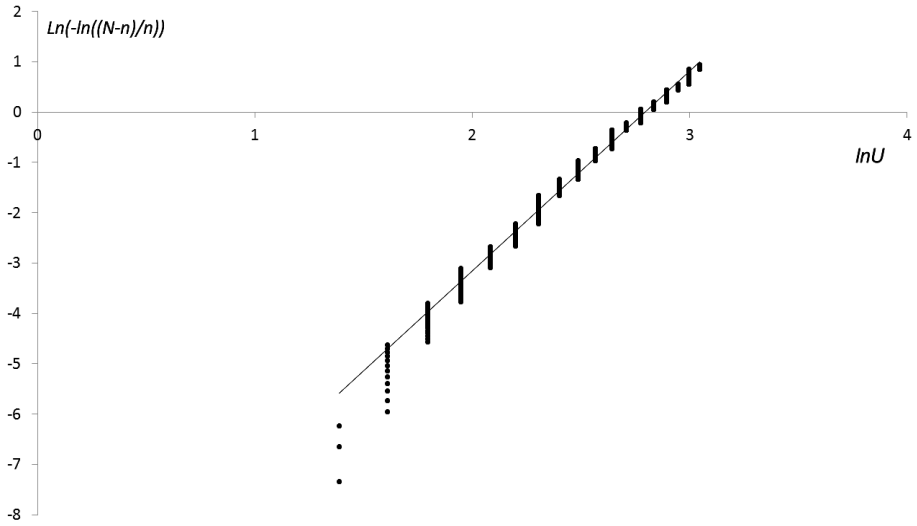
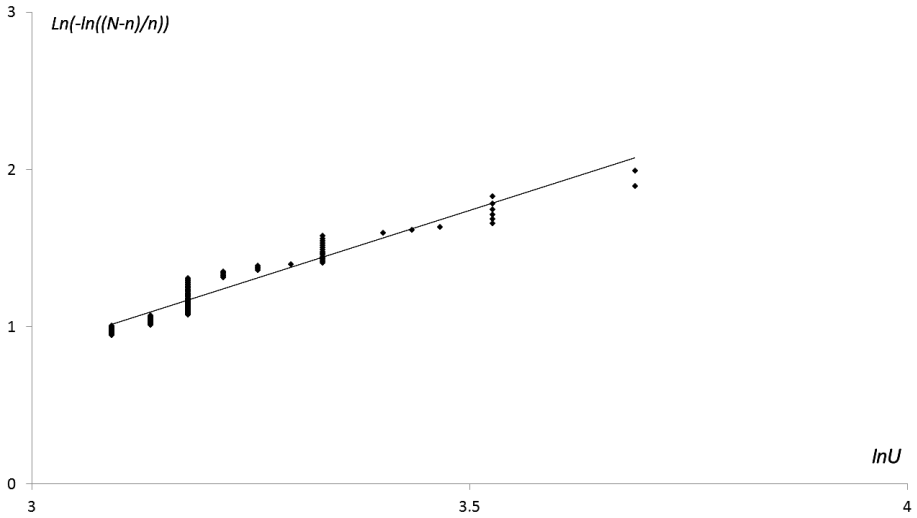


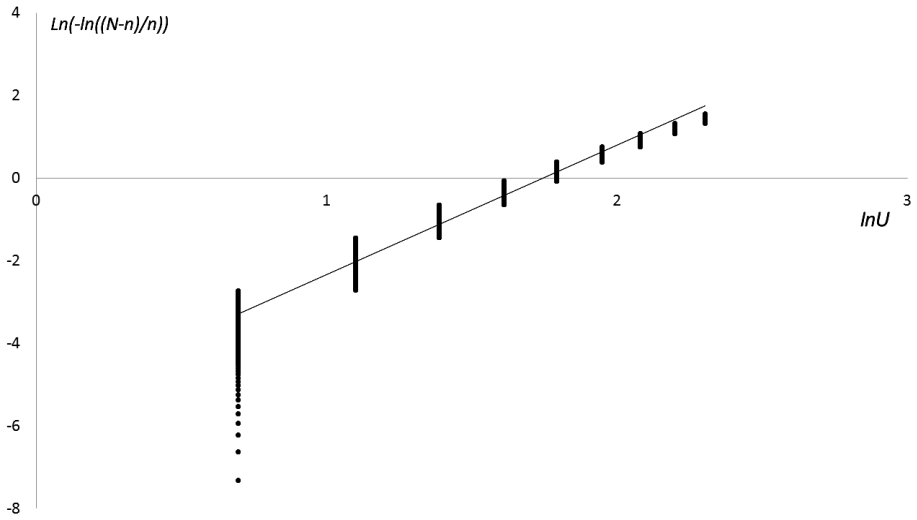
Figure 3. Empirical cumulative distribution functions of wind speed maxima (station observations, cold period) for 72 hours time step records straightening on the coordinate axis of the Weibull distribution, and linear regression line corresponding to the Weibull function. (a) Teriberka, (b) Krasnochelie, (c) Zimnegorsky Mayak.


Ln(-ln((N-n)/n))
lnU
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
0
1
2
3
4

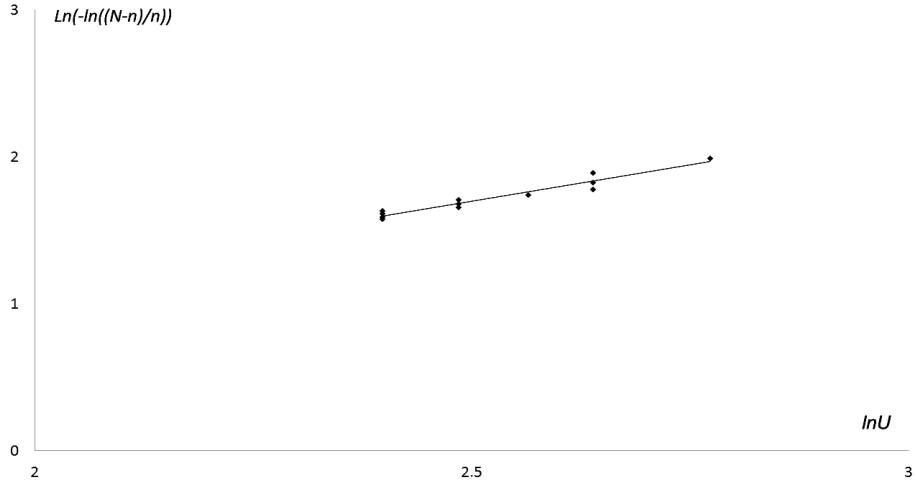
(a)


Ln(-ln((N-n)/n))
lnU
3
2
1
0
3
3.5
4

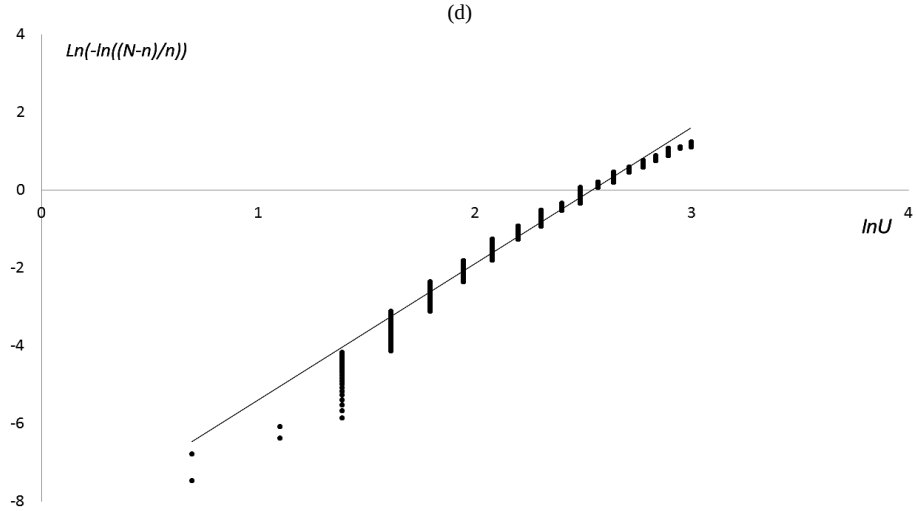
(b)


Ln(-ln((N-n)/n))
lnU
4
2
0
-2
-4
-6
-8
0
1
2
3

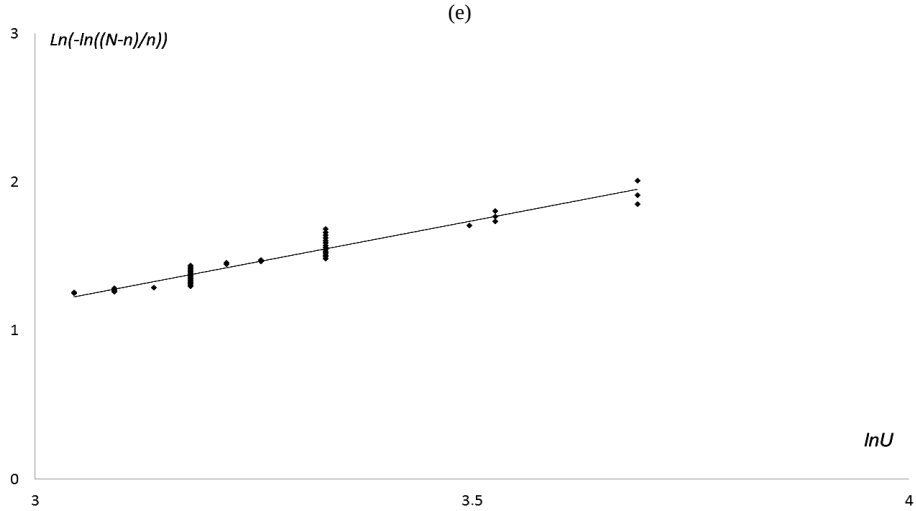
(c)



(d)



(e)



(f)

Figure 4. Empirical cumulative distribution functions of wind speed maxima (station observations, cold period) for 72 hours time step records straightening on the coordinate axis of the Weibull distribution, and linear regression line corresponding to the Weibull function. (a), (b) Teriberka, (c), (d) Krasnochelie, (e), (f) Zimnegorsky Mayak; (a), (c), (e) denotes the Weibull distribution for range $U \le U_{th}$ (so-called "swans"—see below), (b), (d), (f) denotes the Weibull distribution for range $U > U_{th}$ (so-called "dragons"—see below). In all cases $R^2 > 0.96$.

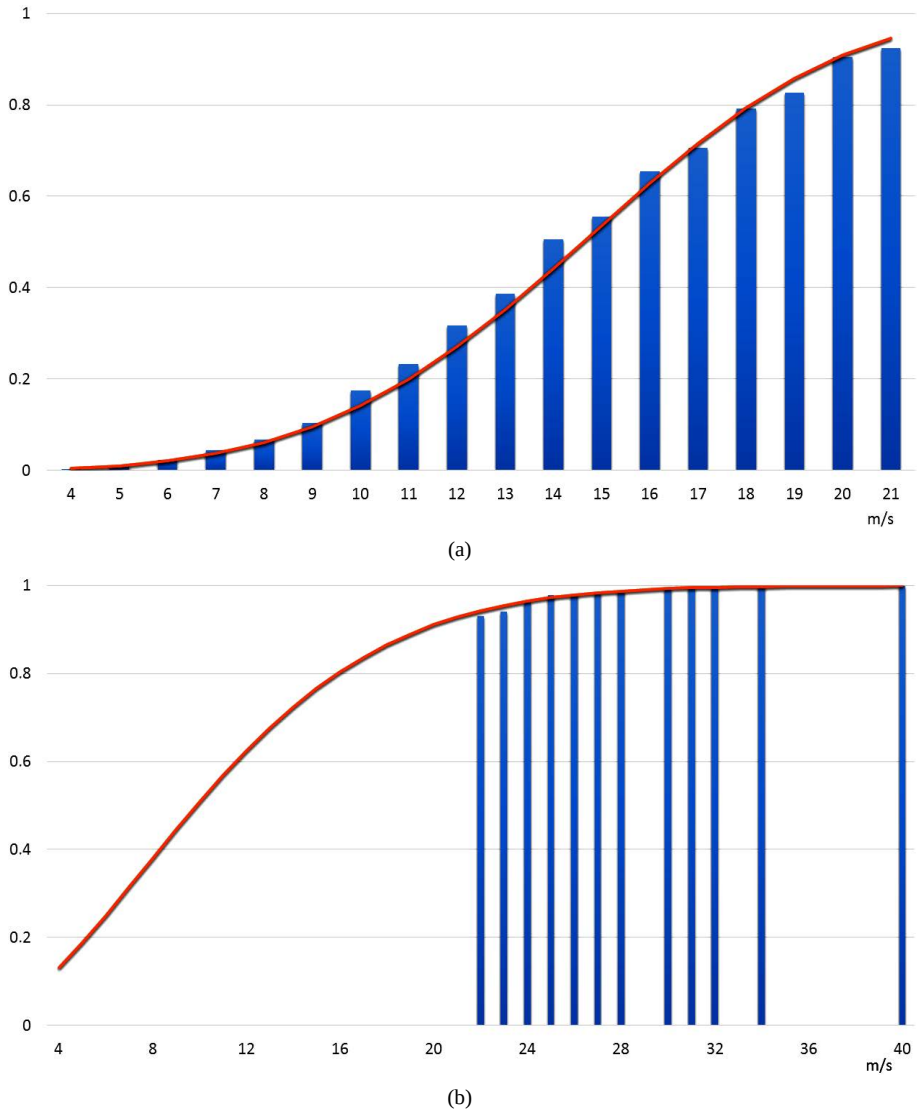


Figure 5. Fragments of empirical cdfs of wind speed extremes on the bases of the Teriberka station measurements and corresponding Weibull distribution using previously calculated parameters: (a) for data shown in **Figure 4(a)**; (b) for data shown in **Figure 4(b)**.

events (below the threshold value) adhere to the Weibull distribution. The same statistical distribution of population could be considered a result of the same organization principle, and this suggests a common generating mechanism for each representative of this population. This idea allows us to understand that a large extreme is not distinguished from its small siblings apart from its large power. The occurrence of large extremes looks like the appearance of a few black swans in a flock of white swans. This terminology was introduced by N.N. Taleb [18], as a metaphor to describe an event that comes as a surprise.

However, there are extreme wind speed events located above the threshold value that are much more pronounced than predicted by the extrapolation of "black swans" law distributions in their tail. They adhere again to the Weibull distribution. Such events were termed "kings" (taking into account the special position of the fortune of kings, which appear to exist beyond the Zipf law distribution of the wealth of their subjects [19]) or "dragons" (to stress that we address a completely different, beyond the normal, type of animal). D. Sornette [20] introduced the concept of dragon-kings to refer to such extreme events. The same statistical distribution suggests a common generating mechanism different from that responsible for extrema at $U \le U_{th}$.

It is not clear to what extent such excessively metaphoric terminology is required for our case because it was

originally introduced to describe unique extraordinary events. However, it allows us to mark events that adhere to different Weibull distributions. Therefore, we will use below these terms: "swans" or "black swans" (hereafter Ss or BSs) and "dragons" (hereafter the Ds).

A very important message is that we can easily diagnose events adhering to the BSs or the Ds (in many cases the diagnostic of Ds is not simple and requires different methods adapted to the specific problem [20]). In **Figure 3**, indeed, the Ds can be detected evidently based on obvious breaks in the tail of wind distributions. In statistics there is a test of the null hypothesis that two samples come from the same population against an alternative hypothesis, especially that a particular population tends to have larger values than the other. This is the Mann-Whitney *U* test (the Kruskal-Wallis test extends the Mann-Whitney U test when there are more than two groups). However, for comparing two (or more) sets of observations, they have to be a priori selected. In the case of wind speed extremes (or other meteorological extrema) such method is not possible, because all data refer to the same nomenclature.

Note that mentioned above exceptional outliers ("super extreme" wind speed events) are denoted along the line corresponding to the Ds distribution on the Weibull Plot. It means that their presence is not prohibited.

The Weibull distribution parameters calculated for all stations are shown in **Table 2**. Parameters are given separately for two groups of wind speed extremes come from the Ss and Ds populations. Note that events belonging to the Ds are not diagnosed in several cases.

4. The Question of Quantile Estimation

As was demonstrated, the Weibull distribution fits well with the data in all cases. Therefore, the estimated parameters (k and A) allow us to calculate the quantile (inverse cumulative distribution) function for the Weibull distribution as follows:

$$U(p) = \left(\frac{1}{A}\ln\frac{1}{1-p}\right)^{1/k}. \tag{4}$$

Quantile wind speed values are calculated differently for the Ss and the Ds (**Table 3**). In our analysis, we have divided the data relating to cold and warm seasons. We have taken into account that in the Arctic region, July and August are the only true summer months, while the winter season covers not only December, January and February but also includes the whole interval from November to April. In these cases, taking into account the data volume (1966-2013) and remembering that time step records are 72 hours, the value of $U(0.99)$ for the summer characterizes the maximum wind speed, which is exceeded, on average, once every five warm seasons. Similarly, the value of $U(0.99)$ for the cold period of the year characterizes the maximum wind speed, which is exceeded, on average, two times every three cold seasons.

The dissimilarity between the Ds and the BSs can reach up to 30%, demonstrating both a difference in statistical properties and, probably, the differences of origin. The most pronounced feature of the geographical distribution of the quantile wind speed values is that the maxima (both the BSs and the Ds) are in the coastal area. As an example, for winter, $U(0.99) = 24\ \mathrm{ms}^{-1}$ (the BS) and $U(0.99) = 29\ \mathrm{ms}^{-1}$ (the Ds) are at the Teriberka station, (corresponding to 19 and 27 ms^{-1} at another coastal station, the Zimnegorsky Mayak), while for the Krasnochelie (the inland station of the Cola Peninsula), $U(0.99)$ is 9 and 10 ms^{-1} for the BS and the Ds, respectively. During the summer, the geographical peculiarities are the same; however, absolute values are almost two times lower.

The "winter acceleration" of wind over the coastal area is not simply a consequence of a smooth sea surface, compared to land. An important role is played by storms that are typically much more active over the sea, especially during the cold season under the conditions of the non-freezing surface of the Barents Sea. During the warm season, the coastal/inland difference is not so pronounced and the quantile values are smaller.

The wind speed extremes observed at the surface should be a function of the meso-scale circulation [21]-[23]. It is a well-known fact that increased wind (e.g., wind gusts) originates from air parcels flowing at higher levels in the boundary layer that are deflected downward to the surface. Apart from meso-scale convective complexes we can note that some low level extremes involve the role of complex terrain as well as combinations of these processes. The most striking example (within the area of investigation) is demonstrated by the Maly Karmakula station where bora winds affect the eastern shore of the Barents Sea and wind speed can be ~40 ms^{-1} (see **Table 3**). Boras develop when cold air from the Kara Sea (typically covered by ice) is blocked by the Novaya Zemly Mountains, which rise to ~1000 m, but it can cross the mountains.

Table 2. The Weibull distribution parameters (*k* and *A*) calculated separately for two groups of wind speed extremes come from the Ss and Ds populations over the selected period 1966-2013 years (for wind speed in m/s).

Station	Population	Cold season		Warm season	
		k	*A*	*k*	*A*
Teriberka	Ss	3.97	1.6E−05	4.39	3.1E−05
	Ds	1.77	0.0120	2.12	0.0081
Murmansk	Ss	3.95	0.0001	4.94	3.6E−05
	Ds	1.34	0.1039	2.56	0.0062
Lovozero	Ss	3.19	0.0013	4.45	0.0003
	Ds	1.69	0.0429	2.30	0.0202
Krasnoshchelye	Ss	3.14	0.0043	3.04	0.0012
	Ds	0.99	0.4608	1.45	0.0664
Kandalaksha	Ss	3.50	0.0017	4.20	0.0006
	Ds	1.22	0.2322	1.45	0.1545
Umba	Ss	3.63	0.0006	4.56	0.0002
	Ds	1.70	0.0508	0.9285	0.4050
Arkhangelsk	Ss	3.60	0.0016	4.00	0.0010
	Ds	1.48	0.1159	1.49	0.1372
Zimnegorsky Mayak	Ss	3.50	0.000145	3.80	0.0002
	Ds	1.13	0.1125	1.40	0.0759
Cape Kanin	Ss	4.80	0.2E−05	4.40	1.9E−05
	Ds	2.40	0.0017	1.30	0.0835
Kolguyev Island Northern	Ss	4.50	0.7E−05	6.10	0.1E−05
	Ds	1.50	0.0309	2.90	0.0013
Kotkino	Ss	2.90	0.0032	3.56	0.0013
	Ds	0.40	1.6109	1.66	0.0869
Naryan-Mar	Ss	3.12	0.0017	4.45	0.0002
	Ds	1.54	0.0620	1.88	0.0389
Ust-Usa	Ss	3.70	0.0006	5.20	3.0E−05
	Ds	1.25	0.1515	1.72	0.0620
Ust-Tsilma	Ss	4.20	0.0002	5.10	4.9E−05
	Ds	0.90	0.3854	1.85	0.0561
Okunev Nos	Ss	3.40	0.0014	4.40	0.0002
	Ds	0.52	1.1722	0.98	0.3816
Hoseda Hard	Ss	3.00	0.0011	4.50	8.8E−05
	Ds	0.98	0.2695	2.20	0.0159
MalyKarmakula	Ss	3.40	5.4E−05	4.08	3.6E−05
	Ds	1.90	0.0042	1.86	0.0087
Anderma	Ss	3.60	7.7E−05	4.26	6.7E−05
	Ds	1.85	0.0127	2.20	0.0108
Marresale	Ss	3.60	0.0001	4.80	1.8E−05
	Ds	1.67	0.0264	1.86	0.0237

Continued

Novy Port	Ss	3.65	0.0002	5.00	1.2E−05
	Ds	1.60	0.0303	1.80	0.0254
Antipauta	Ss	3.19	0.0006	4.60	4.4E−05
	Ds	2.14	0.0093	1.88	0.0266
Dikson	Ss	3.23	0.0002	5.14	0.5E−05
	Ds	2.00	0.0088	2.70	0.0023
Bolvansky Nos	Ss	3.70	6.5E−05	4.15	5.9E−05
	Ds	1.65	0.0254	2.1	0.0115
Khatanga	Ss	3.70	0.0005	5.20	2.1E−05
	Ds	1.76	0.0319	1.50	0.0756
Vize Island	Ss	3.3	0.0003	4.4	3.6E−5
	Ds	2.1	0.0063	2.0	0.0154
Tiksi	Ss	1.9	0.0088	4.7	2.0E−5
	Ds	-	-	2.6	0.0029
Wrangel Island	Ss	2.7	0.0008	3.1	0.0009
	Ds	-	-	2.4	0.0042
Cape Konstantinovsky	Ss	3.9	5.3E−5	5.3	0.3E−5
	Ds	2.1	0.0057	2.5	0.0032
The Yubileynaya	Ss	2.7	0.0081	4.2	0.0002
	Ds	1.3	0.1358	1.4	0.1060
Vankarem	Ss	3.3	0.0002	5.2	0.6E−5
	Ds	1.0	0.2166	2.6	0.0032
Ambarchik	Ss	3.0	0.0005	4.8	1.8E−5
	Ds	1.6	0.0259	3.0	0.0013
Cape Schmidt	Ss	3.2	0.0003	4.2	5.5E−5
	Ds	1.3	0.0632	2.6	0.0033
Ayon Island	Ss	3.12	0.0011	5.0	1.8E−5
	Ds	1.8	0.0188	1.8	0.0294
Ostrovnoye	Ss	2.4	0.0115	4.1	0.0005
	Ds	-	-	1.5	0.1125
Cape Billings	Ss	3.3	0.0005	4.9	1.3E−5
	Ds	1.5	0.0377	2.2	0.0070
Salekhard	Ss	2.9	0.0034	3.8	0.0004
	Ds	1.6	0.0518	1.6	0.0561
Igarka	Ss	3.2	0.0015	4.6	0.0001
	Ds	1.5	0.0852	1.3	0.1848
Kotelny Island	Ss	3.3	0.0004	4.1	4.9E−5
	Ds	1.43	0.00620	2.0	0.0104
Kyusyur	Ss	2.1	0.0107	4.9	2.4E−5
	Ds	-	-	1.9	0.0254

Table 3. Quantile wind speed values $U(0.99)$ in ms^{-1} (1966-2013) for wind data from measurement stations calculated separately for two groups of wind speed extremes come from the Ss and Ds populations.

Station	BSs	Ds	BSs/D	BSs	Ds	BSs/D
	Cold season			Warm season		
Teriberka	24	29	0.83	15	20	0.75
Murmansk	15	17	0.88	11	13	0.85
Lovozero	13	16	0.81	9	11	0.89
Krasnoshchelye	9	10	0.90	8	11	0.82
Kandalaksha	10	12	0.83	9	10	0.90
Umba	12	14	0.86	10	11	0.91
Arkhangelsk	9	12	0.75	8	11	0.73
Zimnegorsky Mayak	19	27	0.70	14	19	0.74
Cape Kanin	21	27	0.78	17	22	0.77
Kolguyev Island Northern	19	28	0.68	12	17	0.71
Kotkino	12	14	0.86	8	11	0.73
Naryan-Mar	13	16	0.81	10	13	0.77
Ust-Usa	11	15	0.73	10	12	0.83
Ust-Tsilma	11	16	0.69	11	16	0.69
Okunev Nos	11	14	0.79	10	13	0.77
Hoseda Hard	16	18	0.89	11	13	0.85
MalyKarmakula	28	40	0.70	18	29	0.62
Anderma	21	24	0.88	13	18	0.72
Marresale	19	22	0.86	13	17	0.77
Novy Port	17	23	0.74	13	18	0.72
Antipauta	16	18	0.89	12	16	0.75
Dikson	21	23	0.91	14	17	0.82
FedorovObservatory	18	23	0.78	12	17	0.71
Bolvansky Nos	20	23	0.87	15	17	0.88
Khatanga	12	17	0.71	11	16	0.69
Vize Island	19	23	0.82	15	17	0.88
Tiksi	?					
Wrangel Island	25	-		16	18	0.89
Cape Konstantinovsky	19	24	0.79	14	19	0.74
Yubileynaya	10	16	0.63	11	16	0.69
Vankarem	20	22	0.91	13	17	0.77
Ambarchik	22	26	0.85	14	15	0.93
Cape Schmidt	20	27	0.74	14	17	0.82
Ayon Island	15	20	0.75	12	17	0.71
Ostrovnoye	12	-		10	11	0.91
Cape Billings	16	25	0.64	14	19	0.74
Salekhard	12	17	0.71	11	16	0.69
Igarka	12	14	0.86	10	11	0.91
Kotelny Island	18	20	0.90	16	21	0.76
Kyusyur	18	-		12	16	0.75

5. The Weibull Distributions in Data of Climate Model Simulation

The next step of the analysis is to investigate to what extent the above-mentioned peculiarities of wind extremes are simulated by climate models. We analysed a dataset of wind simulation of the INM-CM4 climate model. The establishment of the correspondence between wind simulation products and near-surface observations could help us to assess the quality of modelling products and their capability to reproduce the wind extremes. Apart from that, it is important to advance our understanding of the origin of the BSs and the Ds.

In **Figure 6(a)**, **Figure 6(b)**, **Figure 6(c)** we plot several cdfs on the bases of the INM-CM4 simulations. For this aim we chose the INM-CM4 grid points located near the stations. We can conclude that the Weibull distribution is a good approximation of the modelled wind speed extremes. We find very small deviation in cdfs from the theoretical line starting with certain large threshold values. Let us remind that noticeable deviation was a typical feature of empirical cdfs (**Figure 3** and **Figure 4**). Using our terminology, we can conclude that the INM-CM4 model wind speed extremes are the Ss and that there are no the Ds.

This conclusion is supported not only by the specific location of points along a theoretical line but also by the fact that modelled wind speed extremes themselves are close to observation data adhering the BSs besides the Zimnegorsky Mayak data and the Teriberka data, where observed $U(0.99)$ are almost half times greater than modelled values (**Table 4**). Probably, this is due to inadequate distribution of land and sea in the INM-CM4. Their geographical peculiarities are the same (the maxima are at the coastal area).

The discovered phenomenon of the absence of the representatives of the Ds in modelling data is very important. Let us investigate this effect more precisely. Because extreme wind at the surface originates from air parcels that are deflected downward from the top of the boundary layer to the surface, we will focus on extremes in 850 hPa wind, which is likely to be more reliable than surface wind in an atmospheric model, as surface wind is more affected by unresolved topography and land/sea mask, and the model boundary layer scheme. Extremes in 850 hPa wind may be related to the potential for extreme surface wind speed [24].

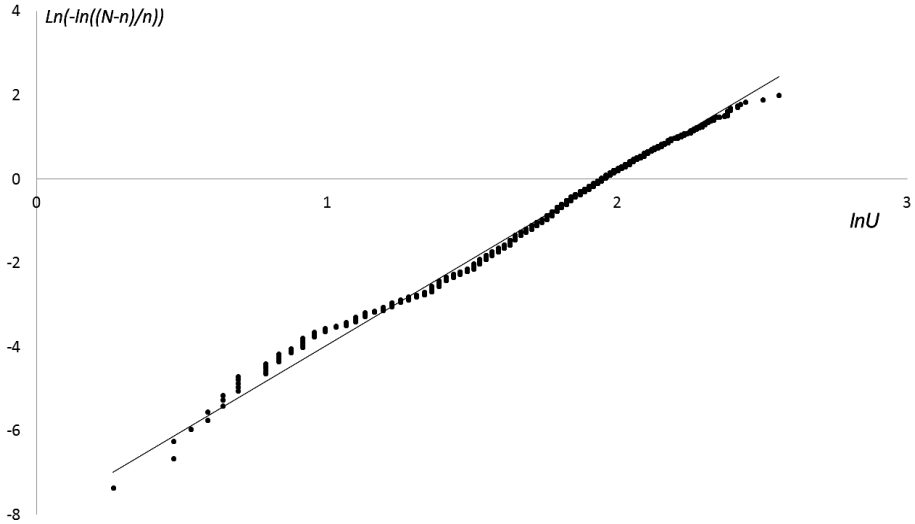
The results obtained (**Figure 6(b)**, **Figure 6(d)**, **Figure 6(f)**) indicate that the Weibull distribution is a good approximation of modelled wind speed extremes at the level H850. We find again the absence of deviation in cdfs from the theoretical line and can conclude that above the atmospheric boundary layer the set of modelled wind speed extremes do not consist of the Ds, they are the Ss. Note that in spite of difference of grid points location (inland area or coastal zones) the statistical features of modelled wind extremes are the same. For example, the quantile wind speed values at the level 850 hPa $U(0.99) = 25 - 26$ m/s for winter season and $U(0.99) = 19$ m/s for summer season (see **Table 5**). It means, that the geographical peculiarities of the near-surface wind speed extremes (see **Table 4**) originate due to air parcel subsidence occurring differently in the modelled atmospheric boundary layer over the coastal and inland area. In vicinity of coastal zone the modelled wind speed extremes at the level H850 are close to near-surface observation data, but within the inland territory the model overestimates the results of measurements (**Table 5**). However, this comparison is not meaningful because compared values are related to different populations (BSs and Ds) and hence their origin is different.

6. Conclusions

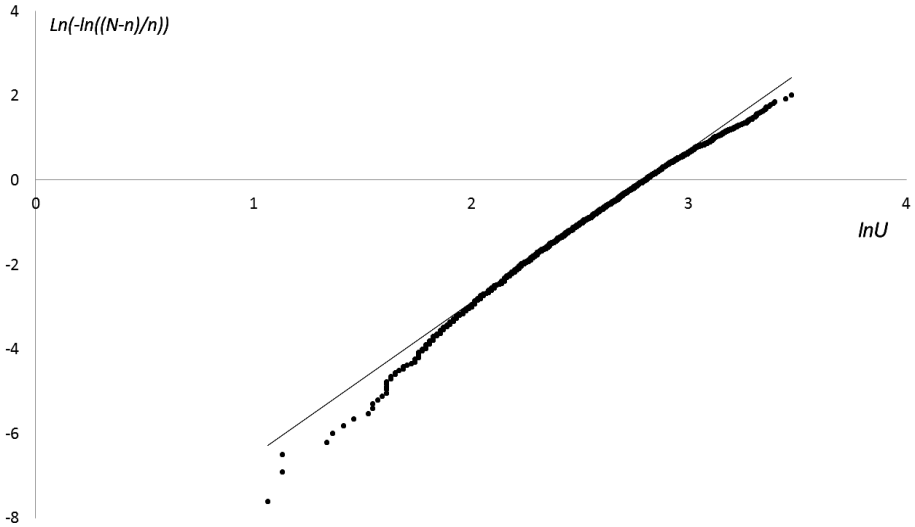
Extreme value analysis has been implemented to estimate the statistical properties of extreme wind speed over the European and Siberian parts of Arctic region from the Kola Peninsula to the Chukotka Peninsula. The application was made on 10-m wind speed data taken from the INM-CM4 climate model dataset and observation stations.

It was shown that for all stations a volume of observed samples of extreme wind speed are composed of two sets of variables. All samples of each population have the same statistical properties but one population is sharply different from another. So different origin of strong wind events adhering to two groups can be concluded. Using metaphoric terminology, we marked these events as the Ss (power extremes are the BSs) and the Ds. However, the modelled (INM-CM4 data) extreme wind speeds consist of only the Ss. Dissimilarity of the Ds and the BSs can reach up to 30%, hence, atmospheric model underestimates extreme wind speeds. The finding that global climate models are unable to capture the wind extremes is already well known, but information that the modelled (INM-CM4 data) extreme wind speeds do not consist of the Ds provides new knowledge.

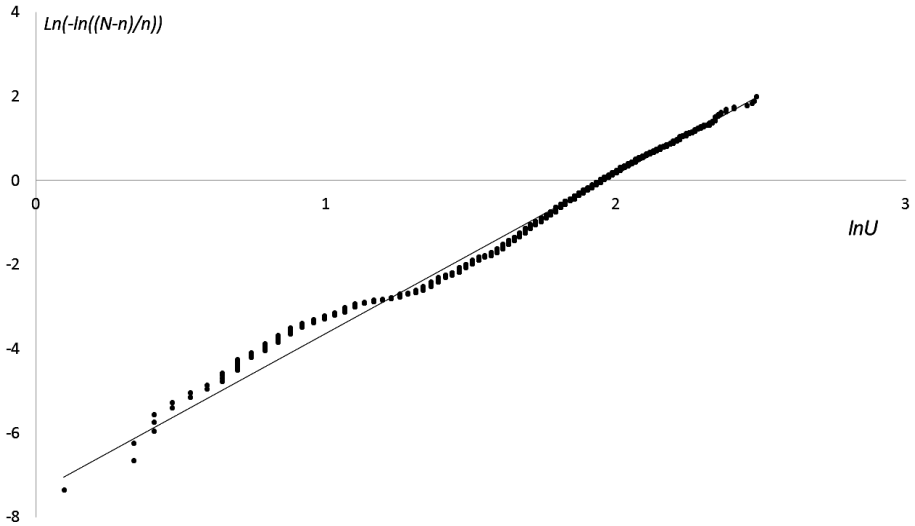
This evidence indicates that the special mechanisms of the Ds are not reproduced by climate models which are utilized as a tool used to simulate climate change. Hence, the problem of identification of pronounced extreme wind speeds based on modelling data remains unresolved.


Ln(-ln((N-n)/n))
lnU
4
2
0
-2
-4
-6
-8
0
1
2
3

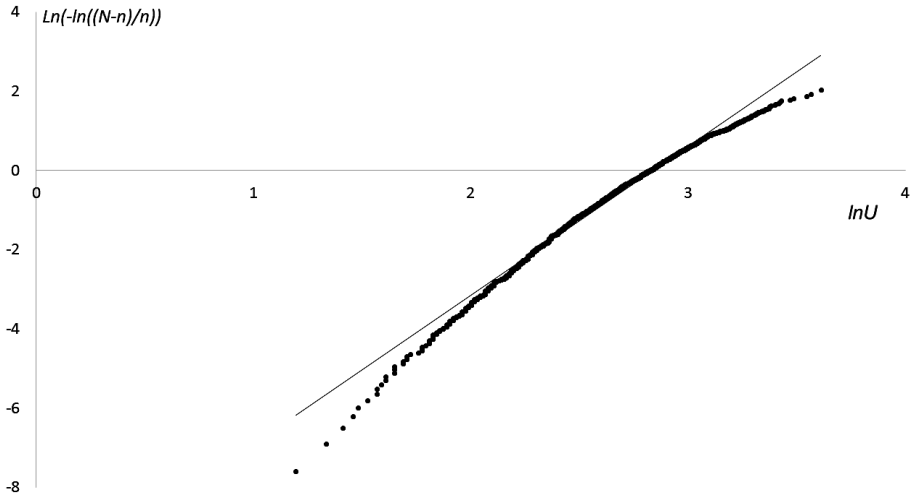
(a)


Ln(-ln((N-n)/n))
lnU
4
2
0
-2
-4
-6
-8
0
1
2
3
4

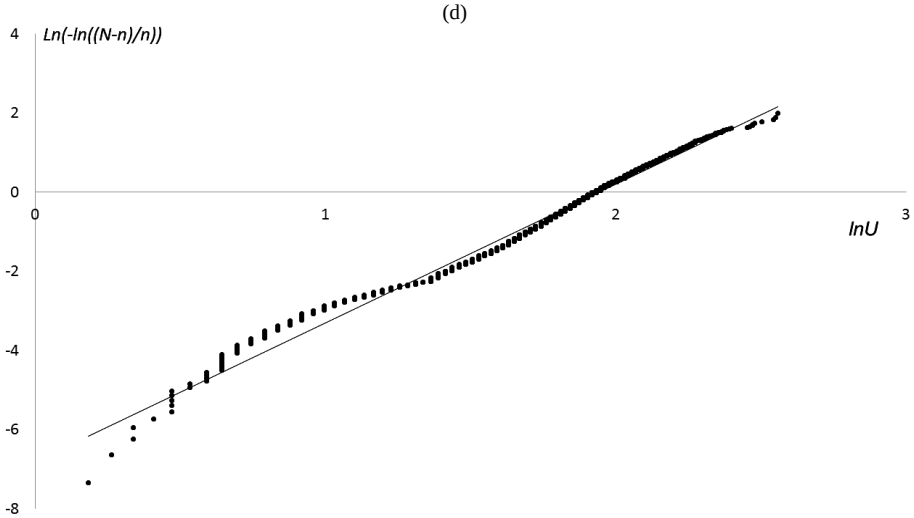
(b)


Ln(-ln((N-n)/n))
lnU
4
2
0
-2
-4
-6
-8
0
1
2
3

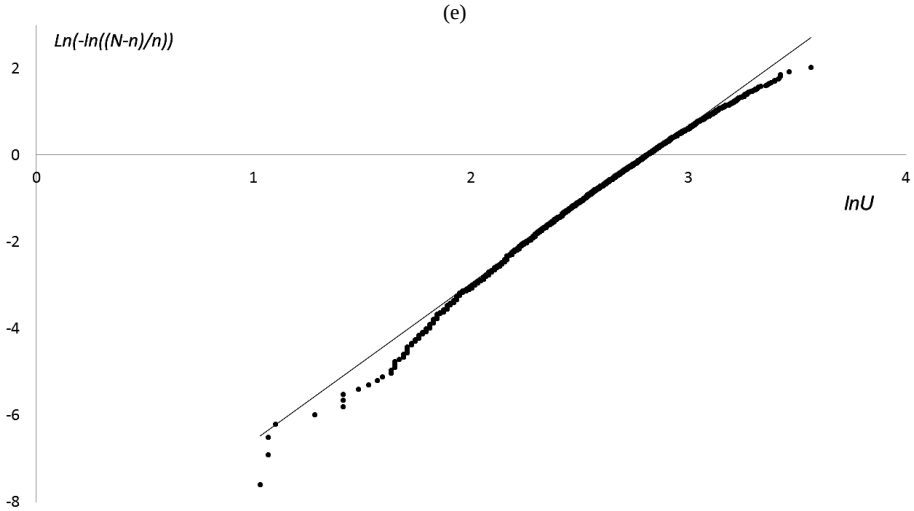
(c)



(d)



(e)



(f)

Figure 6. Cumulative distribution functions of wind speed maxima for cold period simulated by the INM-CM4 in the framework of the Historical experiment (CMIP 5) in grid points corresponded to the Zimnegorsky Mayak (a), (b), Kandalaksha (c), (d) and Krasnoshchelye (e), (f), near the surface (a), (c), (e) and at 850 hPa (b), (d), (f) for 72 hours time step records straightening on the coordinate axis of the Weibull low, and linear regression line corresponding to the Weibull function. In all cases $R^2 = 0.99$.

Table 4. Quantile wind speed values ($U(0.99)$, ms^{-1}) near the surface for grid points corresponding to wind measurement stations across the Kola Peninsula and coastal zones of the Barents Sea and the White Sea (see **Figure 1**) based on data of the INM-CM4 and measurement stations.

Grid points, corresponding to stations					
Teriberka	Lovozero	Krasnochelie	Kandalaksha	Umba	Zimnegorsky Mayak
Winter season, the INM-CM4/the BSs (observation-see **Table 3**)					
19/24	12/13	11/9	11/10	11/12	11/19
Summer season, the INM-CM4/the BSs (observation-see **Table 3**)					
15/15	9/9	9/8	9/9	9/10	9/14

Table 5. Quantile wind speed values ($U(0.99)$, ms^{-1}) at the level 850 hPa for grid points corresponding to wind measurement stations across the Kola Peninsula and coastal zones of the Barents Sea and the White Sea (see **Figure 1**) based on data of the INM-CM4, and quantile wind speed values ($U(0.99)$, ms^{-1}) near the surface based on data of measurement stations adhering to BSs and Ds (see **Table 3**).

Grid points, corresponding to stations					
Teriberka	Lovozero	Krasnochelie	Kandalaksha	Umba	Zimnegorsky Mayak
Winter season, the INM-CM4/BSs/Ds					
25/24/29	25 /13/16	25/9/10	26/10/12	25/12/14	25/19/27
Summer season, the INM-CM4/BSs/Ds					
19/15/20	19/9/11	19/8/11	19/9/10	19/10/11	19/14/19

It is well-known that large wind speed extremes observed at the surface are governed by the mesoscale atmospheric phenomena (embedded into strong synoptic storms) including both convective processes and effects of gravity waves, connecting to specific circulations (like the bora). Because such processes are not fully simulated by coarse spatial resolution atmospheric model, we could conclude that the largest wind speed extremes are not recreated by the climate models. It is important because the tasks demanding the information about wind speed extreme (for example, the task of projection of storminess intensity depending on the surface wind field) cannot be explicitly solved using the output of current climate model.

The mesoscale atmospheric models cover several aspects of such processes, the use of which has huge potential.

Acknowledgements

Funding for this research was provided by grants from the Russian Science Foundation (Grant No. 14-37-00038).

References

[1] Beirlant, J., Goegebeur, Y., Segers, J., JozefTeugels, J., De Waal, D., and Ferro, C. (2004) Statistics of Extremes: Theory and Applications. Willey Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester.

[2] Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics. Springer-Verlag. http://dx.doi.org/10.1007/978-1-4471-3675-0

[3] de Haan, L. and Ferreira, A. (2006) Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering. Springer. http://dx.doi.org/10.1007/0-387-34471-3

[4] Fisher, R.A. and Tippett, L.H.C. (1928) Limiting Forms of the Frequency Distribution of the Largest or Smallest Members of a Sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190. http://dx.doi.org/10.1017/S0305004100015681

[5] Gnedenko, B. (1943) Sur la distribution limite du terme maximum d'unesériealéatoire. *Annals of Mathematics*, **44**, 423-453. (In French) http://dx.doi.org/10.2307/1968974

[6] Palutikof, J.P., Brabson, B.B., Lister, D.H. and Adcock, S.T. (1999) A Review of Methods to Calculate Extreme Wind Speeds. *Meteorological Applications*, **6**, 119-132. http://dx.doi.org/10.1017/S1350482799001103

[7] Brabson, B.B. and Palutikof, J.P. (2000) Tests of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds. *Journal of Applied Meteorology and Climatology*, **39**, 1627-1640. http://dx.doi.org/10.1175/1520-0450(2000)039<1627:TOTGPD>2.0.CO;2

[8] Compo, G.P., Whitaker, J.S. and Sardeshmukh, P.D. (2006) Feasibility of a 100-Year Reanalysis Using Only Surface Pressure Data. *Bulletin of the American Meteorological Society*, **87**, 175-190. http://dx.doi.org/10.1175/BAMS-87-2-175

[9] Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M. C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D. and Worley S.J. (2011) The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1-28. http://dx.doi.org/10.1002/qj.776

[10] Lindsay, R., Wensnahan, M, Schweiger, A. and Zhang, J. (2014) Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic. *Journal of Climate*, **27**, 2588-2606. http://dx.doi.org/10.1175/JCLI-D-13-00014.1

[11] Volodin, E.M., Diansky, N.A. and Gusev, A.V. (2013) Simulation and Prediction of Climate Changes in the 19th to 21th Centuries with the Institute of Numerical Mathematics, Russian Academy of Science, Model of the Earth Climate System. *Izvestia, Atmospheric and Oceanic Physics*, **49**, 347-366. http://dx.doi.org/10.1134/S0001433813040105

[12] Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, **93**, 485-498. http://dx.doi.org/10.1175/BAMS-D-11-00094.1

[13] Seitola, T. and Järvinen, H. (2014) Decadal Climate Variability and Potential Predictability in the Nordic Region: A Review. *Boreal Environment Research*, **19**, 387-407.

[14] Coles, S.G. and Walshaw, D. (1994) Directional Modelling of Extreme Wind Speeds. *Journal of the Royal Statistical Society. Series C* (*Applied Statistics*), **43**, 139-157. http://dx.doi.org/10.2307/2986118

[15] Cook, N.J. (1985) The Designer's Guide to Wind Loading of Building Structures. Part 1: Background, Damage Survey, Wind Data and Structural Classification. Building Research Establishment. Garston and Butterworths, London.

[16] Gusella, V. (1991) Estimation of Extreme Winds from Short-Term Records. *Journal of Structural Engineering*, **117**, 375-390. http://dx.doi.org/10.1061/(ASCE)0733-9445(1991)117:2(375)

[17] Lemeshko, B.Yu. and Lemeshko, S.B. (2009) Distribution Models for Nonparametric Tests for Fit in Verifying Complicated Hypotheses and Maximum-Likelihood Estimators. Part 1. *Measurement Techniques*, **52**, 555-565. http://dx.doi.org/10.1007/s11018-009-9330-3

[18] Taleb, N.N. (2010) The Black Swan: The Impact of the Highly Improbable. 2nd Edition, Penguin, New York.

[19] Laherrère, J. and Sornette, D. (1999) Stretched Exponential Distributions in Nature and Economy: Fat Tails with Characteristic Scales. *The European Physical Journal B*, **2**, 525-539. http://dx.doi.org/10.1007/s100510050276

[20] Sornette, D. (2009) Dragon-Kings, Black Swans and the Prediction of Crises. *International Journal of Terraspace Science and Engineering*, **2**, 1-18. http://dx.doi.org/10.2139/ssrn.1596032

[21] Stensrud, D.J. (1996) Importance of Low-Level Jets to Climate: A Review. *J. Climate*, **9**, 1698-1711. http://dx.doi.org/10.1175/1520-0442(1996)009<1698:IOLLJT>2.0.CO;2

[22] Brasseur, O. (2001) Development and Application of a Physical Approach to Estimating Wind Gusts. *Monthly Weather Review*, **129**, 5-25. http://dx.doi.org/10.1175/1520-0493(2001)129<0005:DAAOAP>2.0.CO;2

[23] Schulz, J.-P. and Heise, E. (2003) A New Scheme for Diagnosting Near-Surface Convective Gusts. *COSMO Newsletter*, **3**, 221-225.

[24] Yin, J.H. and Branstator, G.W. (2007) Extremes in Global Climate Models. *Proceedings 'Aha Huliko'a Hawaiian Winter Workshop*, University of Hawaii at Manoa, 23-26 January 2007, 67-72.