

## The application of probability density function in modeling of wind speed on the Polish Baltic Coast

Bartosz Czernecki

Adam Mickiewicz University in Poznań, Department of Climatology, Dzięgielowa Street 27, 61-680 Poznań, Poland,  
e-mail: nwp@amu.edu.pl

**Abstract:** The aim of the research was to identify the potential for the use of probability density functions (*PDF*) in modeling of near-surface wind speed. The approaches of Empirical Orthogonal Functions (*EOF*) and Canonical Correlation Analysis (*CCA*) are used in combination with 2-parametric Weibull distribution. The downscaling model was built using a diagnosed relationship between sea level pressure (*SLP*) patterns over Europe and the Northern Atlantic and estimated monthly values of Weibull parameters at 9 stations along the Polish Baltic Coast.

The obtained scale (*A*) and shape (*k*) parameters make it possible to describe temporal variations of wind fields and their theoretical probability values. This may have further application in the modeling of extreme wind speeds for seasonal forecasting, climate prediction or in historical reconstructions. The model evaluation was done separately for the calibration (1971-2000) and validation periods (2001-2010). The scale parameter was reconstructed reasonably, while there were some problematic issues with the shape parameter, especially in the validation period. The quality of the developed models is generally higher for the winter season, due to larger *SLP* gradients, whereas the results for the spring and summer seasons were less satisfactory. Despite this, the 99th percentile of theoretical wind speeds are in most cases satisfactory, due to the lesser importance of the shape parameter for typical distributions in the analyzed region.

**Keywords:** wind speed, Weibull distribution, canonical correlation analysis, statistical downscaling, Polish Baltic Coast

**Submitted** 15 July 2013, **revised** 29 December 2014, **accepted** 29 December 2014

### 1. Introduction

Wind is a meteorological phenomenon that is mainly governed by the general circulation pattern and local conditions that modify the flow of air masses on smaller scales. This modification is particularly strong in areas with highly variable topographical configuration, with a variety of vegetation cover and variability of albedo (Chromow 1969). Besides mountainous areas, local aspects are particularly important in regions of land-sea interaction, due to heat capacity differences and sharply changing values of the surface friction coefficient (Simpson 1994).

The increasing interest for anemological studies in recent years, especially for mean and extreme wind speeds (Lorenc 1992, 1996, 2012; Arażny et al. 2007), is not surprising as wind properties have a strong influence on other variables characterizing the local climate. It is often emphasized that the increasing tendency of extreme meteorological phenomena (Pryor et al. 2006; IPCC 2007) may comprise a higher frequency of severe wind speed events causing environmental hazards (e.g. hydrological, geomorphological, biosphere) and economic hazards (e.g.

affecting infrastructure). Nevertheless, there is a strong need to quantify the possibilities of using different downscaling techniques, since the analysis and modeling of extreme values is still very challenging, especially for non-continuous datasets (STARDEX 2005).

### 2. Objectives, data and methods used

In this paper quantitative methods of statistical downscaling are employed, by means of applying probability density functions (*PDF*). This approach has the capacity to fulfill most requirements postulated by interested groups of end users of wind data. It must be noted that using parameters of distribution instead of raw data may be applied for creating a synthetic dataset for an evaluated period of time. One of the main advantages of analyzing the variability of phenomena in a statistical way is the use of a much smaller dataset to represent whole data distribution, which leads to a significant reduction of redundant data (Wójcik et al. 2014). Moreover, it is possible to extrapolate wind speed values directly from the density function (Wilks 2008). This approach may be useful for

historical and future climate change reconstructions, and also for seasonal forecast purposes (e.g. as Model Output Statistics).

Modeling of the behavior of the tails of *PDFs* provides the opportunity to focus on extreme events (Maraun et al. 2010), the frequency and/or strength of which are often underestimated while using non-parametric datasets due to the significant reduction of the variance in such a process (Pryor et al. 2005; Buishand, Brandsma 2001). Similar effects have also been noted while using dynamical downscaling methods (Turner et al. 2011). Despite this, there is no clear conclusion in contemporary literature on downscaling strategies as to which approach gives better results for particular regions and meteorological variables. Usually, downscaling results are comparable and undoubtedly statistical-empirical downscaling is not as computationally demanding as dynamical modeling (STARDEX 2005; Huth et al. 2014). Several recent studies couple statistical and dynamical methods (e.g. STARDEX 2005; Bernardin et al. 2009). This approach emphasizes the need to develop and verify tools and predictors for empirical-statistical downscaling for different meteorological phenomena.

The main aim of this paper is to check whether the most common statistical-empirical downscaling techniques are capable of reconstructing the probability density function of wind speed on the Polish Baltic Coast, since the results are important for the modeling of wind speed for various applications. This task requires the identification of relations between the large-scale force field described by the reduced sea level pressure (*SLP*) and regional wind speed variability.

Monthly mean *SLP* patterns were derived from the NCEP/NCAR reanalysis project (Kalnay et al. 1996) for the period of 1971-2010, and cover the area of the North

Atlantic and Europe (35°N-70°N, 50°W-40°E, Fig. 1). The values of predictands were estimated from 3-hourly data from the operational network of the Institute of Meteorology and Water Management (Fig. 1). Site-specific conditions of wind speed measurements at the stations are presented in Tab. 1.

The test region is one of the areas of most potential for wind energy purposes (Lorenc 1992, 1996; Czernecki 2013) and spatial-temporal distribution of wind speed is of wide interest to decision makers, engineers and environmentalists due to its connection to storm surges, coastal erosion and other events. The Polish Baltic Coast is covered by a regular meteorological network with a long history of observations. Despite this, not all available datasets for the region were included in the research due to some technical problems with measurements which led to the inhomogeneous of data series in the period of 1971-2010 (Czernecki 2013). Data used in this study were not homogenized or reduced to a fixed anemometer height. Some problematic issues are connected with the change of locations during these 40 years. For example, at Elbląg at least 3 different types of anemometer were used, the height of which changed from 19 to 20 meters in 1979, and in late 1970's the measurement conditions worsened due to changes in surrounding buildings.

Estimation of parameters of Weibull distribution.

Among probability density functions used for wind speed data, the best fit are typically obtained for Rayleigh and Weibull distributions. However, the use of Weibull distribution is common in climatological and wind energy research (e.g. Lorenc 1996; Pryor et al. 2005; Bernardin et al. 2009). According to Seguro and Lambert (2000) one of the best methods of estimating Weibull parameters is the maximum likelihood approach. This method was applied

Table 1. Meteorological stations maintained by the Institute of Meteorology and Water Management used in this study. Surface roughness class (0-6) according to Lorenc (2012). Mean wind speed and values of Weibull parameters calculated for the period of 1971-2010

Station name	Coordinates	Elevation [m]	Anemometer height [m]	Surface roughness class	Mean wind speed	Weibull's <i>PDF</i> parameters	
						scale ( <i>A</i> )	shape ( <i>k</i> )
Kołobrzeg	15.58°E, 54.18°N	3	18	3	3.10	3.65	1.89
Koszalin	16.15°E, 54.20°N	33	13	2	3.49	4.20	1.85
Ustka	16.87°E, 54.58°N	6	20	3	4.33	4.93	1.77
Łeba	17.53°E, 54.75°N	2	14	0	4.92	5.78	1.72
Łębork	17.75°E, 54.55°N	38	12	3	2.85	3.79	1.69
Hel	18.82°E, 54.60°N	1	23	3	4.33	4.95	2.04
Elbląg	19.43°E, 54.17°N	40	19	2	3.35	4.01	2.01
Świnoujście	14.23°E, 53.92°N	6	11	1	3.72	4.32	1.74
Szczecin	14.62°E, 53.40°N	1	24	2	3.73	4.40	1.98

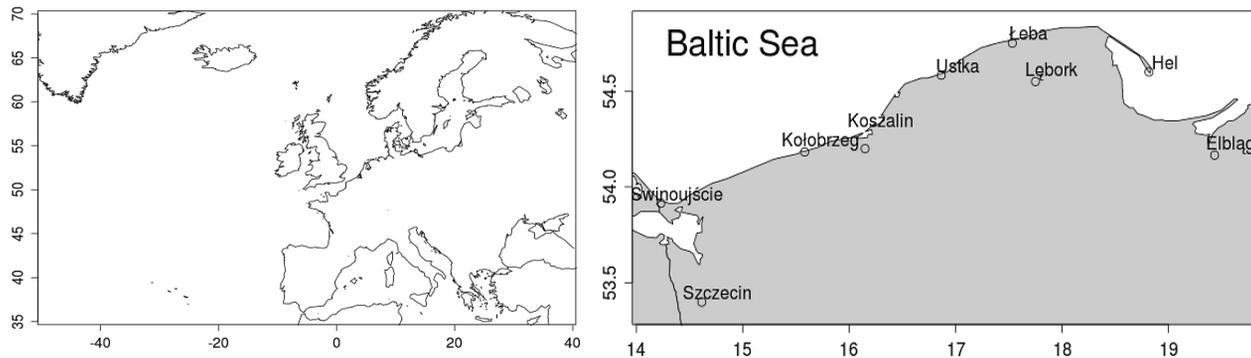


Fig. 1. Range of predictor field domain (left) and locations of stations used in the study (right)

by the “fitdistrplus” package (Delignette-Mueller et al. 2010), which is dedicated to the R programming environment (R Core Team 2013).

The following calculations for extreme values of monthly wind speed are based on transformations of formulas for probability density function and cumulative distribution function as shown in eq. 1 and eq. 2 (Troen, Petersen 1989; Mortensen et al. 2004).

$$f(v) = \left(\frac{k}{A}\right) \left(\frac{v}{A}\right)^{k-1} \exp\left[-\left(\frac{v}{A}\right)^k\right] \text{ for: } v>0, A>0, k>0 \quad (1)$$

$$F(v) = 1 - \exp\left[-\left(\frac{v}{A}\right)^k\right] \quad (2)$$

where:  $v$  – wind speed; positive values only allowed [ $\text{m s}^{-1}$ ];  $A$  – scale factor (determines the scale of the curve) [ $\text{m s}^{-1}$ ];  $k$  – shape factor, which determines shape of the curve [dimensionless].

According to Pramod (2011) the use of Weibull’s distribution has some specific properties connected with the different values of  $A$  and  $k$  factors. The  $A$  factor, to some degree, is related to the mean of distribution, although it always differs from mean wind speed values due to the removal of wind calms from a dataset before its estimation (eq. 1, see also Tab. 1). For  $k$  factor, the *PDF* is an exponential function while  $k = 1$ , it represents a Rayleigh’s distribution for  $k = 2$  and for  $k$  slightly above 3 it becomes similar to Gaussian distribution (Pramod 2011).

In fact, for large datasets, as in this study, even relatively small changes of one Weibull parameter causes noticeable changes in a whole distribution, something that is presented with example datasets created from monthly mean parameters at selected stations in years 1971-2000 (Fig. 2).

The Canonical Correlation Analysis (*CCA*) was chosen as a method for the creation of the empirical-statistical model. This method relies on finding spatially coherent patterns for both regional and local fields with the largest possible temporal correlation (Preisendorfer 1988; Benes-

tad et al. 2008; Wilks 2008). Among different versions of *CCA* one of the most often used is described by Preisendorfer (1988) and is also called “Barnett-Preisendorfer *CCA*” (Benestad et al. 2008). It is based on a decreasing number of dimensions from the recognized patterns by means of the Empirical Orthogonal Function (*EOF*) (Lorenz 1956; Preisendorfer 1988) and on finding spatial structures with the maximum correlations.

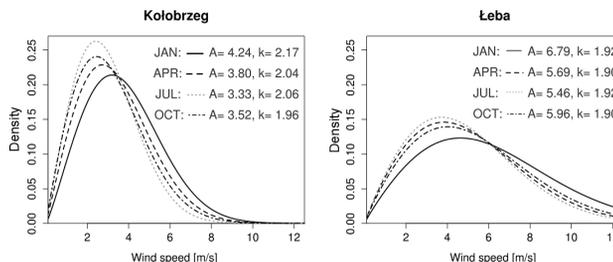


Fig. 2. The theoretical probability density functions (*PDF*) of wind speed at selected stations of the Polish Baltic Coast for the given scale and shape parameters of Weibull distributions (1971-2000)

### 3. Results

#### 3.1. Empirical-statistical downscaling of anemologic conditions described by the parameters of Weibull distribution

Analysis of the temporal and spatial structure of Weibull parameters is presented with the example of characteristic months of the year. This choice reflects the relatively wide range of their calculated annual, seasonal and monthly values. Moreover, the variation of the parameters of Weibull distribution is higher for monthly than for seasonal and annual means. Similarly, the correlation between the predictor and predictands seems to be greater in shorter time periods, while some of the mean *SLP* patterns for seasons and years do not include sharp changes in pressure gradient, which physically influences temporal variation of wind speed.

According to *EOF* analysis done for the Weibull scale ( $A$ ) parameter in selected months, there are 6 (January) to

Table 2. Explained variance [%] by EOFs of Weibull's distribution scale ( $A$ ) parameter of wind speed in the Polish Baltic Coast, 1971-2000

month	EOF1	EOF2	EOF3	EOF4	EOF5	EOF6	EOF7	EOF8	EOF9	Total
JAN	66.55	12.98	8.89	5.07	2.95	1.61				98.05
APR	42.31	23.08	14.17	8.54	3.91	3.47	2.40	1.15		99.03
JUL	46.88	24.87	11.51	5.72	3.36	2.79	2.16	1.53	1.18	99.99
OCT	62.36	13.81	10.94	5.05	3.16	2.04	1.32			98.67

Table 3. Explained variance [%] by EOFs of Weibull's distribution shape ( $k$ ) parameter of wind speed in the Polish Baltic Coast, 1971-2000

month	EOF1	EOF2	EOF3	EOF4	EOF5	EOF6	EOF7	EOF8	EOF9	Total
JAN	51.53	15.53	8.87	7.68	5.46	4.52	3.19	2.04	1.19	98.81
APR	42.06	15.03	12.49	10.64	6.65	4.92	3.62	2.72	1.87	98.13
JUL	34.91	23.61	13.49	9.78	5.61	4.66	4.01	2.73	1.20	98.80
OCT	50.37	16.21	11.27	8.49	5.60	3.48	2.14	1.39	1.06	98.94

9 (July) non-degenerated eigenvectors, which explain over 98% of total variance. The first four eigenvectors explain from 88.1% (April) to 93.4% (January) of total variance, so the contribution of the next few eigenvectors is relatively small (Tab. 2). For the Weibull shape parameter ( $k$ ) total explained variances are comparable and explain from 98.1% to 98.9%, but in each case the number of eigenvectors is equal 9, which is the maximum possible value. Additionally, there is a lesser domination of the first EOFs, something that may suggest that the signal for shape parameter ( $k$ ) is not as obvious as was recognized for scale parameter ( $A$ ).

Canonical correlation (CCA) model of Weibull scale ( $A$ ) and shape ( $k$ ) parameters.

To recognize temporal and spatial interaction between the regional and local field, Canonical Correlation Analysis (CCA) was used. The created CCA models were based on the same number of eigenfunctions as derived from

EOF analysis. The strongest correlation between predictor and Weibull scale parameter was recognized in January and explains about 39% variance of SLP and over 62% of the local field. In other seasons the tele-connections between predictor and predictand are not as obvious and the explained variances do not exceed 30% for the regional field and 20% in the case of the local field for the highest correlated pairs of maps (Tab. 4).

In the case of Weibull shape parameter ( $k$ ) reflecting the variability of wind speed, it must be noted that the explained variance by the regional and local field is significantly smaller (Tab. 5) than was recognized for scale parameter ( $A$ ). However, the correlation coefficient values for the first pair of maps, even if slightly smaller than for scale parameter ( $A$ ), are still relatively high and oscillate around  $r = 0.9$ . In the author's opinion, the problem with modeling the shape parameter is connected with its low variance for monthly time periods. Bearing in mind that empirical-statistical downscaling tend to underestimate

Table 4. Results of Canonical Correlation Analysis in chosen months, 1971-2000 ( $r$  – correlation coefficient, varSLP – explained variance [%] of predictor field, var $_A$  – explained variance [%] of Weibull's distribution scale ( $A$ ) parameter of wind speed)

month:		CCA1	CCA2	CCA3	CCA4	CCA5	CCA6	CCA7	CCA8	CCA9	Total
JAN	r	0.97	0.87	0.72	0.48	0.37	0.14				-
	varSLP	39.06	15.01	9.18	5.84	4.50	10.2				83.79
	var $_A$	62.58	3.99	13.9	6.94	6.45	4.43				98.29
APR	r	0.94	0.85	0.76	0.59	0.48	0.33	0.28	0.24		-
	varSLP	11.46	39.54	2.55	14.06	6.59	13.58	4.88	4.66		97.32
	var $_A$	13.55	12.79	12.94	16.69	11.04	7.8	6.97	17.29		99.07
JUL	r	0.97	0.86	0.75	0.68	0.56	0.39	0.20	0.12	0.10	-
	varSLP	14.38	18.96	7.13	6.79	7.54	10.68	9.49	12.83	7.80	95.60
	var $_A$	34.76	18.15	11.35	2.93	8.53	7.95	4.58	6.74	5.00	99.99
OCT	r	0.93	0.88	0.67	0.50	0.38	0.33	0.14			-
	varSLP	24.35	11.12	10.26	6.54	7.59	23.42	5.56			88.84
	var $_A$	16.14	19.79	12.69	22.96	10.41	5.89	10.79			98.67

Table 5. Results of Canonical Correlation Analysis in chosen months, 1971-2000 ( $r$  – correlation coefficient, varSLP – explained variance [%] of predictor field, var $_k$  – explained variance [%] of Weibull’s distribution shape ( $k$ ) parameter of wind speed)

month:		CCA1	CCA2	CCA3	CCA4	CCA5	CCA6	CCA7	CCA8	CCA9	Total
JAN	r	0.88	0.80	0.70	0.57	0.54	0.36	0.19			-
	varSLP	20.42	23.90	19.92	10.57	8.15	5.27	9.20			97.43
	var $_k$	5.51	17.75	26.15	12.82	8.25	7.66	7.65			85.79
APR	r	0.93	0.88	0.79	0.52	0.45	0.39	0.19	0.10	0.01	-
	varSLP	12.65	16.80	9.31	9.04	7.23	7.26	9.58	16.99	7.40	96.26
	var $_k$	13.57	11.75	11.63	6.40	11.97	12.76	12.39	7.05	12.49	100
JUL	r	0.93	0.81	0.73	0.73	0.58	0.43	0.27	0.17	0.03	-
	varSLP	19.05	6.22	17.24	10.21	13.90	4.92	7.73	9.54	6.58	95.39
	var $_k$	15.38	9.85	5.22	8.90	10.98	11.78	16.71	9.48	11.70	100
OCT	r	0.90	0.83	0.69	0.65	0.36	0.26	0.13	0.01		-
	varSLP	19.21	22.45	14.09	2.94	4.41	16.89	8.52	7.95		96.46
	var $_k$	24.94	12.06	10.65	23.25	6.14	2.42	14.10	5.45		99.01

variance of a modeled value, it causes serious problems with data reconstruction, especially in months where mean the SLP pattern has a low gradient. The diagnosed pattern of  $k$  parameter anomalies during a year are often strongly related to the local conditions of wind measurements on stations. Such examples are Świnoujście and Szczecin, where much stronger anomalies were recognized in comparison to other stations. The reason for this may be the shadow effect of Rugia island (Lorenc 1996), especially in situations with a domination of winds from western sectors. Also, anomalies of  $k$  parameter in Lębork often tend to have the opposite values to anomalies on neighbouring stations, probably because of its location on the edge of an urstromtal where tunneling effects are observed.

### 3.2. Downscaling model evaluation

In order to distinguish the realism of the model simulation, a reconstruction for the entire period of the analysis was carried out for both Weibull parameters and for the 99<sup>th</sup> percentile of mean wind speed. Reconstructed datasets were compared with observations using correlation coefficients, normalized standard deviations, and root mean square errors (RMSEs) – this was done separately for the calibration periods (1971-2000) and validation periods (2001-2010).

It was noted that CCA models computed for whole seasons tend to have smaller correlation coefficients than those reconstructed on the basis of monthly CCA values in the calibration period 1971-2000 (Tab. 6). Due to higher correlations of datasets obtained from monthly CCA models, further calculations were carried out for seasonal datasets, as above, based on independent monthly CCA reconstruction.

Summary results are presented on Taylor diagrams (Taylor 2001) separately for calibration periods (1971-2000) and verification periods (2001-2010) – see Fig. 3, 4. The blue dotted lines indicate normalized standard deviations (i.e. standard deviations of modeled data divided by standard deviation of observation data), the green lines indicate RMSE values and the black dotted lines show correlation coefficient values.

#### 3.2.1. Weibull’s scale parameter ( $A$ ) – calibration period (1971-2000)

The highest values of agreement between modeled and observational data for Weibull scale parameter ( $A$ ) were obtained for winter seasons with area average correlation coefficient 0.67. Values of this coefficient vary from 0.46 in Hel and 0.48 in Elbląg, up to 0.82 in Szczecin, 0.80 in

Table 6. Correlation coefficient of observed and reconstructed data sets on selected stations for  $A$  (left) and  $k$  (right) parameters of Weibull distribution, 1971-2000. Correlation calculated for seasonal CCA models signified by (1), calculations based on monthly CCA models (mean from monthly correlations) signified by (2)

$A$	Kołobrzeg		Łeba		Elbląg	
	(1)	(2)	(1)	(2)	(1)	(2)
Winter	0.71	0.72	0.76	0.80	0.39	0.48
Spring	0.54	0.67	0.62	0.64	0.28	0.44
Summer	0.50	0.62	0.70	0.82	0.44	0.58
Autumn	0.67	0.69	0.71	0.82	0.26	0.55

$k$	Kołobrzeg		Łeba		Elbląg	
	(1)	(2)	(1)	(2)	(1)	(2)
Winter	0.40	0.61	0.31	0.53	0.57	0.62
Spring	0.24	0.51	0.46	0.51	0.41	0.49
Summer	0.31	0.50	0.43	0.64	0.30	0.39
Autumn	0.45	0.70	0.26	0.54	0.50	0.58

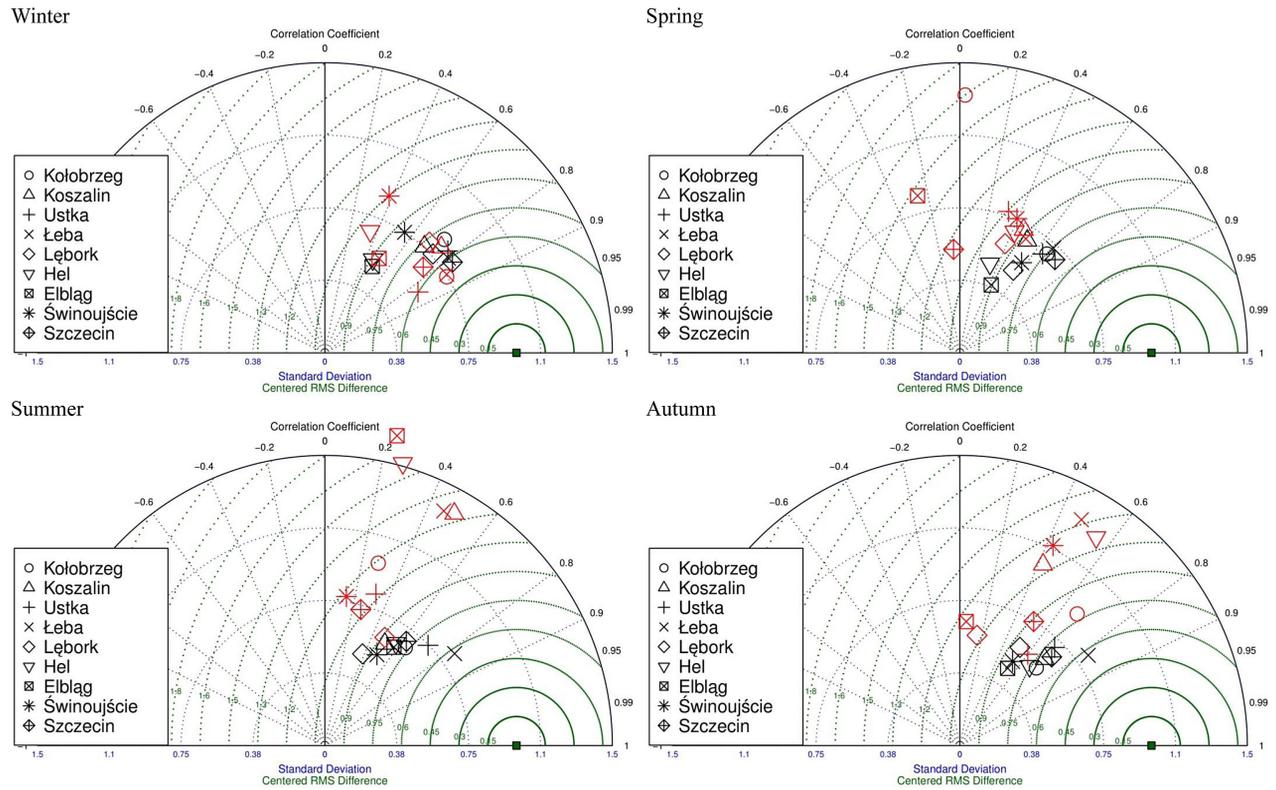


Fig. 3. Taylor diagrams for Weibull scale ( $A$ ) parameter in calibration (black) and verification period (red)

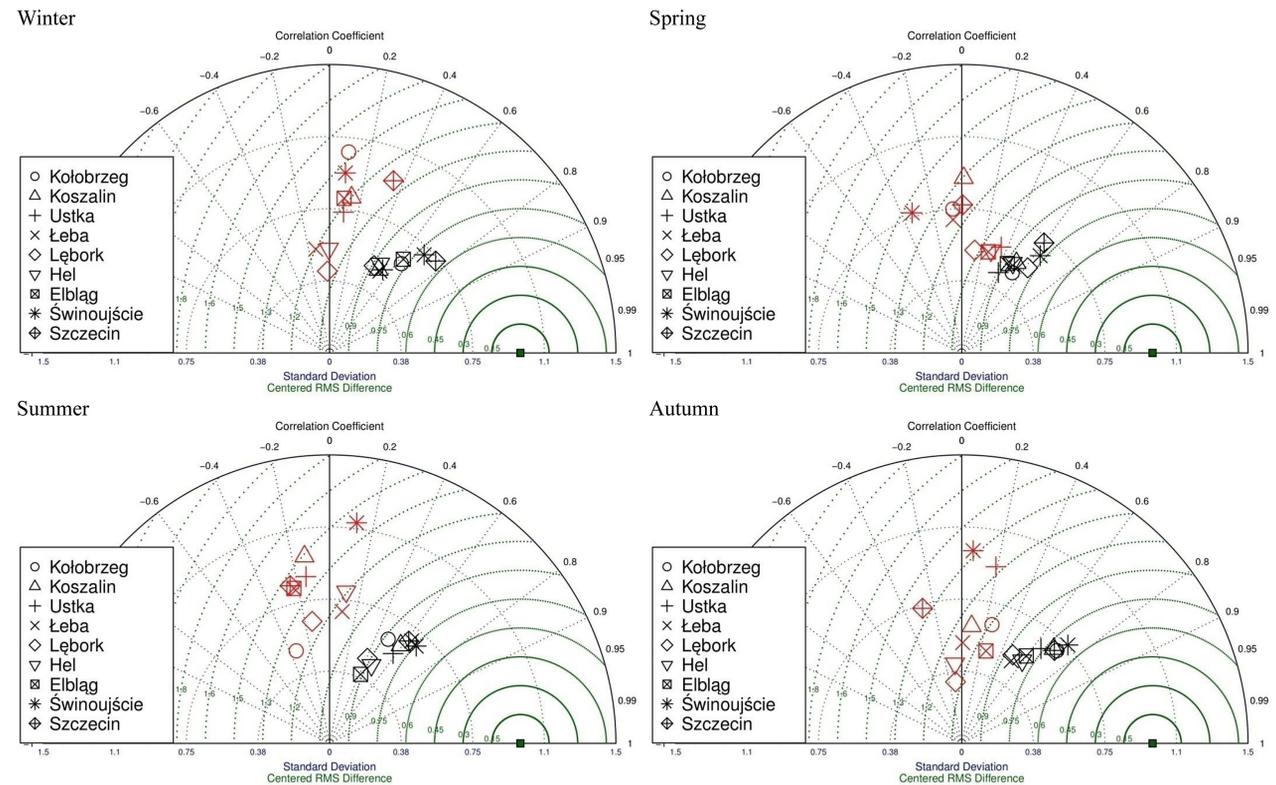


Fig. 4. Taylor diagrams for Weibull shape ( $k$ ) parameter in calibration (black) and verification period (red)

Łeba and 0.77 in Ustka. As presented on the Taylor diagram (Fig. 3), the quality of the reconstructed dataset is worst in Świnoujście, where lower values of correlation coexist with higher than average values of standard deviation and  $RMSE$ .

A significantly lower quality of model was obtained for spring seasons. The most outlying correlations were calculated for Hel ( $r = 0.32$ ) and Elbląg ( $r = 0.42$ ), while the best results were obtained for Szczecin ( $r = 0.72$ ) and Łeba ( $r = 0.67$ ). Spring is also a season with the lowest

value of area average correlation ( $r = 0.56$ ) during a year. In the cases of Hel and Elbląg lower correlations coexist with high values of *RMSE*. The best results for normalized standard deviation were obtained for Elbląg, Hel, Lębork and Świnoujście.

The average correlation between observations and modeled data in summer seasons is only slightly better ( $r = 0.53$ ) than the one presented above for spring. The best reconstructions in terms of correlation coefficients were obtained for Łeba ( $r = 0.82$ ) and Ustka ( $r = 0.72$ ), with the lowest in Lębork ( $r = 0.32$ ). Area average *RMSE* in summer is 0.48 and varies from 0.31 in Szczecin up to 0.63 in Świnoujście.

Significantly better results of Weibull shape parameters were calculated for autumn seasons, with average correlation coefficient 0.65. Again, the highest values were obtained in the middle part of the Polish Baltic Coast (Łeba 0.82) and the lowest in Świnoujście (0.53). The range of *RMSE* values is comparable to that described above for other seasons, i.e. between 0.6 and 0.9.

### 3.2.2. Weibull's shape parameter ( $k$ ) – calibration period (1971-2000)

The level of agreement for Weibull shape ( $k$ ) parameter in calibration period (1971-2000) varies less in particular seasons and stations than is recognized for scale parameter ( $A$ ). One of the highest average seasonal correlation coefficient values for all stations was obtained for winter ( $r = 0.58$ ) (Fig. 4). The above average correlations were recognized in Kołobrzeg, Świnoujście and Szczecin. Also, the *RMSE* values are one of the lowest, something that confirms the good quality of the model for the winter in these locations. For the rest of the stations the model quality is slightly worse, with correlation coefficients from 0.46 to 0.54, but values of model agreement with observations (also for *RMSE* and standardized standard deviation) are still of a satisfactory level.

The worst correlation coefficient results for  $k$ -parameter in the calibration period were obtained in spring months (III-V) with  $r = 0.53$ . There is a noticeable differentiation between particular stations, as the correlation coefficients vary from 0.42 (Ustka) to 0.63 (Świnoujście). Typical values of *RMSE* are around 0.80 and normalized standard deviations vary from 0.50 to 0.75. It is worth noting that the low value of correlation coefficient in the case of Ustka is compensated by a lower normalized standard deviation value.

The quality of the downscaling model in the summer is comparable to that presented for spring. Area average correlation coefficient for summer is 0.54 and varies from

0.41 (Lębork, Elbląg) to 0.67 (Świnoujście). These values are only slightly better than during the spring, while the range of *RMSE* is almost the same (0.80). Similarly, as was observed in the case of spring, stations with lower correlations have at the same time the lowest values of normalized standard deviations.

The highest level of agreement between observations and model data in the calibration period is calculated for autumn season, with area average correlation coefficient  $r = 0.64$ . Most analyzed stations also have the highest values of correlation in this season, with a maximum of up to 0.74 (Świnoujście). Range of *RMSE* and standard deviations are comparable to that described in the case of winter, something that confirms the good quality of the *CCA* model for the autumn season.

### 3.2.3. Weibull's scale parameter ( $A$ ) – verification period (2001-2010)

In verification period (2001-2010) agreement between the observed and modeled Weibull scale parameter ( $A$ ) significantly decreases. For summer, autumn and partially in spring seasons the increase of *RMSEs* are in some cases even twice as great as those calculated for the calibration period (Fig. 3). Only *RMSEs* and correlation coefficients for the winter (area average  $r = 0.66$ ) are comparable to those obtained in the calibration period ( $r = 0.67$ ). Additionally, correlation coefficients in Kołobrzeg, Ustka and Łeba are higher than in the reference period ( $r \sim 0.85$ ).

In other seasons the agreement with observations is not so high, something that may be particularly noticeable in the example of area average correlation coefficient in the spring season, with  $r = 0.24$  (the worst result in all seasons). As well as this, there are slightly negative correlation values for Elbląg and Szczecin. Better model quality was obtained for summer (average  $r = 0.33$ ), but none of the stations have a correlation greater than 0.50 (Koszalin and Łeba  $r = 0.49$ ). For autumn, the average correlation is 0.43, but there is a huge span in model quality for some of the analyzed stations, such as Elbląg ( $r = 0.05$ ) and Kołobrzeg ( $r = 0.67$ ). Among all the stations, the best models in the annual cycle were obtained for Łeba ( $r = 0.57$ ), Koszalin ( $r = 0.53$ ) and Ustka ( $r = 0.52$ ). The worst results in the seasonal cycle were confirmed in the case of Elbląg ( $r = 0.13$ ).

### 3.2.4. Weibull's shape parameter ( $k$ ) – verification period (2001-2010)

The variation of monthly wind speed is represented weakly in the downscaling model beyond calibration pe-

riod (Fig. 4). Among all seasons area average correlation coefficients vary from -0.09 in summer up to 0.06 in spring and autumn, and 0.08 in winter. For summer 6 of 9 stations have negative values of this coefficient and for other seasons there are always 3 such incidences. The range of correlation coefficients varies from -0.34 (Świnoujście in spring, Kołobrzeg in summer) up to 0.35 (Szczecin in winter, Ustka in spring). Also, there are negative average correlations for annual values of Weibull's shape parameter, especially in the case of Szczecin, Łębork ( $r = -0.04$ ), Łeba and Świnoujście ( $r = -0.02$ ). Even the highest correlations (Ustka  $r = 0.13$ , Elbląg  $r = 0.10$ , Hel  $r = 0.08$ ) are not satisfactory for reliable reconstructions of wind speed variations.

Similarly, as was described for the scale parameter, values of *RMSE* are significantly higher than in the calibration period and their range is usually between 1.0 and 1.4.

### 3.3. Variation of Weibull parameters and extreme wind speeds (1971-2010)

Due to problems with the downscaling model, especially in dealing with Weibull's shape parameter ( $k$ ) beyond the calibration period, some attempts were made to check the effects of these misrepresentations on extreme wind speeds. There is no unambiguous method of checking the size of the influence of overestimation or underestimation of  $k$ -parameter on the heads and tails of *PDFs*. However, this problem should not significantly influence

the middle range of cumulative distribution function (i.e. mean wind speed) because the scale parameter, which is connected to median, is reconstructed for most analyzed stations to, in the author's opinion, a satisfactory level.

Multiannual variation of both Weibull parameters and (calculated on their basis) the 99<sup>th</sup> percentile of mean wind speed are presented for characteristic months in the example of Łeba (Fig. 5), which represents one of the most windy areas in Poland (Lorenc 2012). High correlations for A-parameter, especially in winter, are confirmed in agreement between observations and model reconstructions. Lower correlations beyond the calibration period are seen in some noticeable discrepancies, especially in seasons with no clear high-gradient pressure pattern, as presented in the example of July (Fig. 5). Nonetheless, the range of changes, and in most cases their annual trends, are properly represented by the model.

The reconstructed datasets for  $k$ -parameter indicate a problem with both the range and direction of multiannual change. Especially after the calibration, there are significant problems in representing the variability of this parameter. Additionally, as presented in the chart for July, the obtained anomalies are constantly too high, although the error seems not to be as relevant as suggested by correlation coefficient values. Partial confirmation of this thesis may be well seen in July, in years 2002-2006, when the calculated 99<sup>th</sup> percentiles of wind speed are almost equal to those estimated on the basis of the observation dataset. This may also suggest that the influence of  $k$ -parameter

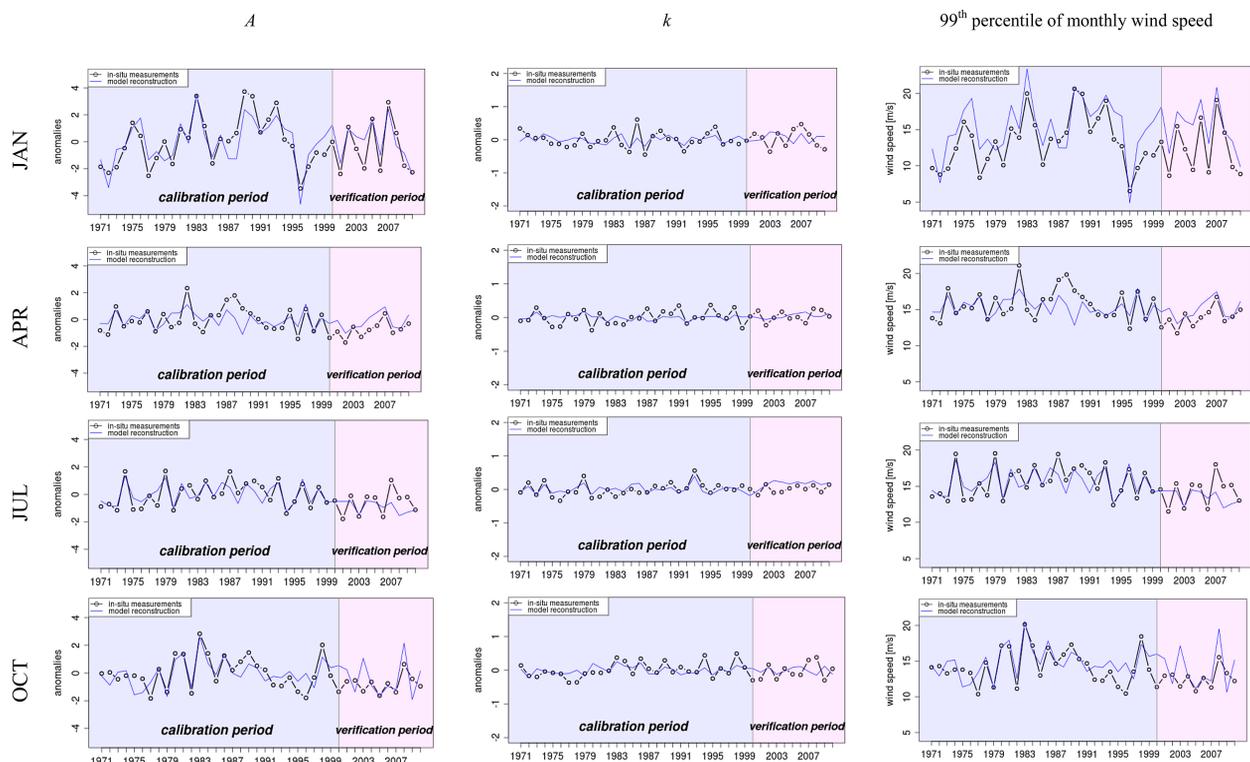


Fig. 5. Variation of Weibull parameters and 99<sup>th</sup> percentile of wind speed according to historical and modeled dataset in Łeba (1971-2010)

is much less when compared to the  $A$ -parameter, something that would explain the relatively high similarity of the multiannual course of Weibull  $A$ -parameter with high wind speeds in all of the analyzed seasons (Fig. 5).

As presented in the example of Łeba, characteristics of model quality for Weibull parameters are usually also observed at other stations. However, there are some site-specific conditions near station locations which can modify large scale air flow and sometimes these are not fully represented by the downscaling model. Calculated values of the 99<sup>th</sup> percentile of mean wind speed for all stations show that in general the model slightly tends to overestimate high wind speeds. Model values are in 58% of all cases higher than observations, and average overestimation is about  $0.10 \text{ m s}^{-1}$  in the calibration period and  $0.26 \text{ m s}^{-1}$  in the verification period. 7 out of 9 stations in the calibration period have overestimated values in January and October, while the inverse relation is characteristic for July (Fig. 6a). At 4 stations average absolute errors for monthly mean values of the 99<sup>th</sup> percentile are less than  $0.1 \text{ m s}^{-1}$ . The diagnosed low model quality in the case of Elbląg impacts on the biggest mean error, equal to  $-0.88 \text{ m s}^{-1}$  (model underestimation).

Similar regularities are found in the verification period, with a dominance of high wind speed overestimations in January, April and October and underestimations in July (Fig. 6b). However, worse model quality in comparison to years 1971-2000 impacts on a greater spectrum of errors. Mean overestimation of model wind speed in January is  $0.96 \text{ m s}^{-1}$ , in April  $0.44 \text{ m s}^{-1}$  and in October  $0.40 \text{ m s}^{-1}$ . In June mean model underestimation is equal to  $-0.76 \text{ m s}^{-1}$ . Among all stations the mean errors calculated from values of characteristic months are greater than  $1 \text{ m s}^{-1}$  for 5 of 9 stations (Ustka  $-2.80 \text{ m s}^{-1}$ , Świnoujście  $1.47 \text{ m s}^{-1}$ , Koszalin  $1.27 \text{ m s}^{-1}$ , Hel and Łeba  $1.07 \text{ m s}^{-1}$ ), while the lowest error was obtained for Szczecin ( $-0.18 \text{ m s}^{-1}$ ). Significantly worse model quality is noticed in case of Ustka, where underestimation of modeled data may be higher than  $3.5 \text{ m s}^{-1}$  (April).

#### 4. Discussion and conclusions

The presented results of coupling statistical downscaling approach with probability density function show that there are certain limitations to local wind speed predictability. Even if the *CCA* models include most regional and local fields' variances (Tab. 2, 3) there is a significant decrease of model quality for Weibull parameters beyond the calibration period, something which influences problems with local wind speed reconstruction. The Large number of recognized *EOFs*, especially for the shape parameter ( $k$ ), means that some of them may be omitted in the modeling of time periods not included in the calibration periods.

On the other hand, it must be noted that use of Weibull parameters instead of raw data helped avoid problems related to significant reduction of variance (Pryor et al. 2005; Buishand, Brandsma 2001) and allowed the calculation of the range of wind speed occurrence, including extreme events. This kind of information, especially for site-specific locations, is often diminished while using coarse-grid climate models and often does not fulfill end user needs, something which may underline the potential of using the PDF approach in climate predictions and reconstructions (Maraun et al. 2010).

Despite problems with model quality for Weibull shape parameter ( $k$ ), it was found that the more important issue is the quality of the scale parameter ( $A$ ), which is responsible to a greater degree for the calculated percentile values. Moreover, the variability of scale parameter ( $A$ ) is almost linearly correlated with the variability of extreme wind speeds (Fig. 5). The reason for this fact may be the relatively small rate of change of the shape parameter ( $k$ ), which in most cases is close to  $k = 2$  (similar to the Rayleigh distribution). However, the importance of the  $k$ -parameter is noticeable while calculating values for heads and tails of the Weibull cumulative distribution function. As presented in Fig. 5, 6a and 6b, differences between modeled and observed values slightly increase in years when problems with shape parameters occurred.

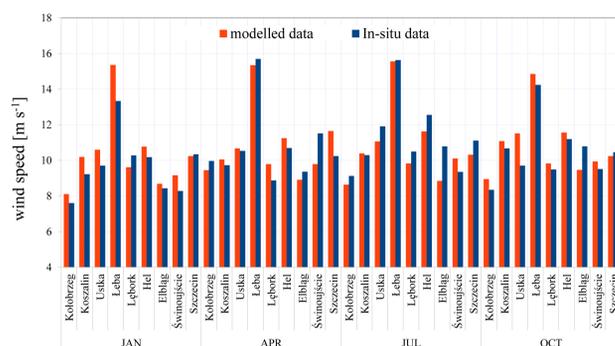


Fig. 6a. The 99<sup>th</sup> percentiles of wind speeds calculated from observed and modeled dataset in calibration period (1971-2000)

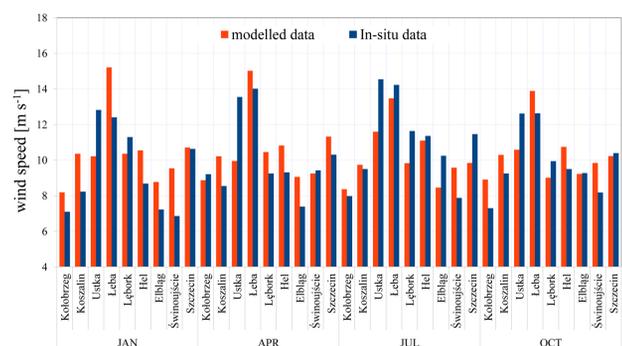


Fig. 6b. The 99<sup>th</sup> percentiles of wind speeds calculated from observed and modeled dataset in verification period (2001-2010)

The best results of *CCA* models were obtained for winter, while the worst are typical for spring and summer seasons. This is probably connected with a clearer signal of predictor field, with the largest *SLP* gradient during a year (Miętus 1999) and greater wind velocities in coastal areas. The inverse situation occurs in summer, when high-gradient *SLP* patterns do not occur so often. Geographically, the highest correlation values were found in the middle part of the Polish Baltic Coast, usually between Kołobrzeg and Łeba.

Another extremely important factor is the matter of data quality and its representativeness for particular stations. The local tunneling effect observed in Lębork, the shadow effect of Rugia Island in Świnoujście and in Szczecin, or changes of anemometer height and measurement instruments during the calibration period mean that some analyzed stations have noticeably worse reconstructed datasets and may also influence the increasing number of obtained *EOFs*. This may suggest that in further research, e.g. on a national scale, additional analysis of metadata must be done to exclude some station datasets in order to improve a model's quality.

The assumption of a physical relationship between regional large-scale *SLP* patterns and the temporal variation of local wind field described by Weibull distribution may be also more promising for shorter time scales. The use of monthly means as a predictor field cannot provide all necessary information that influences short-term local wind speed variability, which is further used to estimate *PDF* parameters. This thesis may be indicated by the better model quality obtained for monthly than for seasonal *CCA* models (Tab. 6). Besides creating models for shorter time periods, the use of vector wind components should also be considered in further research, which may be more promising and may lead to the avoiding of reconstruction problems for one of a distribution parameter.

**Acknowledgments.** This research was partially supported by the statutory programme in the field of climatology (DS-K) run by the Institute of Meteorology and Water Management – National Research Institute, Warsaw, Poland.

#### Bibliography

- Arażny A., Przybylak R., Vizi Z., Kejna M., Maszewski R., Uscka-Kowalkowska J., 2007, Mean and extreme wind speed in Central Europe during the period 1951-2005 (On the basis of data from NCEP/NCAR Reanalysis Project), *Geographia Polonica*, 80 (2), 69-78
- Benestad R.E., Hanssen-Bauer I., Chen D., 2008, *Empirical-Statistical Downscaling*, World Scientific Publishing Co.
- Bernardin F., Bossy M., Chauvin M., Drobinski C., Rousseau P., Salameh A., 2009, Stochastic downscaling method: application to wind refinement, *Stochastic Environmental Research and Risk Assessment*, 23, 851-859, DOI: 10.1007/s00477-008-0276-9
- Buishand T.A., Brandsma T., 2001, Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest neighbour resampling, *Water Resources Research*, 37 (11), 2761-2776, DOI: 10.1029/2001WR000291
- Chromow S.P., 1969, *Meteorologia i Klimatologia*, PWN, Warszawa
- Czernecki B., 2013, Creating wind field time-series over the Southern Baltic area using dynamical downscaling approach, *Meteorologische Zeitschrift*, 22 (5), 587-593, DOI: 10.1127/0941-2948/2013/0454
- Delignette-Muller M.L., Pouillot R., Denis J.-B., Dutang C., 2010, Fitdistrplus: help to fit of a parametric distribution to non-censored or censored data
- Huth R., Miksovsky J., Stepanek P., Belda M., Farda A., Chladova Z., Pisoft P., 2014, Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature, *Theoretical and Applied Climatology*, DOI: 10.1007/s00704-014-1190-3
- IPCC, 2007, *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Geneva, 104 pp.
- Kalnay E., Kanamitsu M., Kistler R., et al., 1996, The NCEP/NCAR 40-year Reanalysis Project, *Bulletin of the American Meteorology Society*, 77, 437-471, DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2
- Lorenc H., 1992, *Zasoby wiatru w Polsce*, Materiały Badawcze IMGW, seria: Meteorologia, 18, IMGW, Warszawa
- Lorenc H., 1996, *Struktura i zasoby energetyczne wiatru w Polsce*, Materiały Badawcze IMGW, seria: Meteorologia, 25, IMGW, Warszawa
- Lorenc H., 2012, *Maksymalne prędkości wiatru w Polsce*, IMGW-PIB, Warszawa, 100 pp.
- Lorenz E.N., 1956, *Empirical Orthogonal Function and Statistical Weather Prediction*, Statistical forecasting project, Scientific Report No. 1, MIT Cambridge
- Maraun D.F., Wetterhall A.M., Ireson A.M., Chandler R.E., Kendon E.J., Widmann M., Brienen S., Rust H.W., Sauter T., Themessl M., Venema V.K.C., Chun K.P., Goodess C.M., Jones R.G., Onof C., Vrac M., Thiele-Eich I., 2010, Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, 48 (3), 1-34, DOI: 10.1029/2009RG000314

- Miętus M., 1999, Rola regionalnej cyrkulacji atmosferycznej w kształtowaniu warunków klimatycznych i oceanograficznych w polskiej strefie brzegowej Morza Bałtyckiego, Materiały Badawcze IMGW, seria: Meteorologia, 29, IMGW, Warszawa, 157 pp.
- Mortensen N.G., Landberg L., Troen I., Petersen E.L., Rathmann O., Nielsen M., 2004, WASP Utility Programs. Risø-I-2261(EN), Risø National Laboratory Roskilde, 52 pp.
- Pramod J., 2011, Wind Energy Engineering, The McGraw-Hill Companies, Inc., 330 pp.
- Pryor S.C., Schoof J.T., Barthelmie R.J., 2006, Winds of change? Projections of near-surface winds under climate change scenarios, *Geophysical Research Letters*, 33 (11), L11702, DOI: 10.1029/2006GL026000
- Pryor S.C., Schoof J.T., Barthelmie R.J., 2005, Empirical downscaling of wind speed probability distributions, *Journal of Geophysical Research: Atmospheres*, 110 (D19), D19110, DOI: 10.1029/2005JD005899
- R Core Team, 2013, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, online access on <http://www.R-project.org/>
- Preisendorfer R.W., 1988, Principal component analysis in meteorology and oceanography. Elsevier, Amsterdam, 436 pp.
- Seguro J.V., Lambert T.W., 2000, Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *Journal of Wind Engineering and Industrial Aerodynamics*, 85 (1), 75-84, DOI: 10.1016/S0167-6105(99)00122-1
- Simpson J.E., 1994, Sea breeze and local winds, Cambridge University Press, 264 pp.
- STARDEX, 2005, Recommendation on the more robust statistical and dynamical downscaling methods for the construction of scenarios of extremes, Deliverable D 16 – Summary Report
- Storch von H., Zwiers F.W., 2001, Statistical Analysis in Climate Research, Cambridge University Press, 499 pp.
- Taylor K.E., 2001, Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, 106 (D7), 7183-7192, DOI: 10.1029/2000JD900719
- Troen I., Petersen E.L., 1989, European Wind Atlas, Risø National Laboratory, Roskilde. 656 pp.
- Turner R., Zheng X., Gordon N., Uddstrom M., Pearson G., de Vos R., Moore R., 2011, Creating Synthetic Wind Speed Time Series for 15 New Zealand Wind Farms, *Journal of Applied Meteorology and Climatology*, 50 (12), 2394-2409, DOI: 10.1175/2011JAMC2668.1
- Wilks D.S., 2008, Statistical Methods in the Atmospheric Sciences, 2nd ed., International Geophysics Series, 59, Academic Press, 627 pp.
- Wójcik R., Pilarski M., Miętus M., 2014, Statistical downscaling of probability density function of daily precipitation on the Polish coast, *Meteorology Hydrology and Water Management*, 2 (1), 27-36