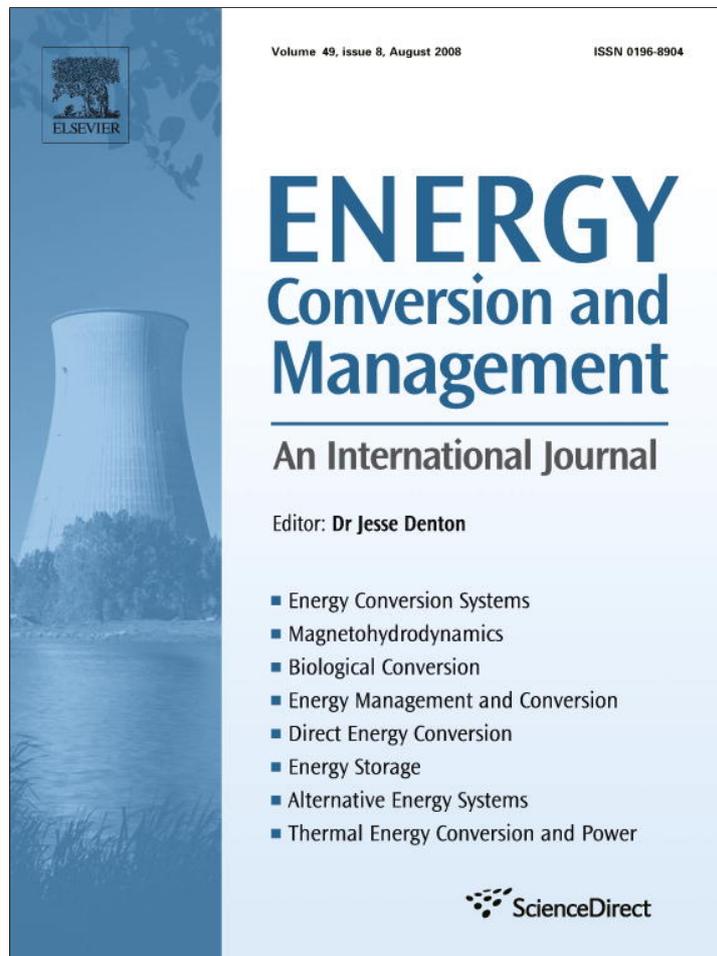


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconmanComprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe[☆]Péter Kiss^{*}, Imre M. János

Department of Physics of Complex Systems, Loránd Eötvös University, Pázmány Péter sétány 1/A, Budapest, Hungary

ARTICLE INFO

Article history:

Received 19 June 2007

Accepted 20 February 2008

Available online 10 April 2008

Keywords:

Wind speed

Wind energy

Weibull distribution

Generalized gamma distribution

ERA-40 re-analysis

ABSTRACT

As a first step of an exhaustive assessment of wind energy potential over Europe, here, we provide a unified description of the wind speed probability distribution both over sea and land. We evaluated surface wind velocity records of the ERA-40 data base covering 44 years with a temporal resolution of 6 hours. We tested the well known distribution functions (Rayleigh, binormal, Weibull, lognormal etc.) and observed that the popular Weibull function performs supremely, however, it fails at many locations over land. We found that the generalized gamma distribution, which has independent shape parameters for both tails, provides an adequate and unified description almost everywhere. The geographical distribution of the fitted parameters reveals the possible climatological origin of different wind speed distributions.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Wind energy is the world's fastest growing renewable source of electricity; the global capacity has more than quadrupled between 2000 and 2006 [1]. The atmospheric flows are strongly volatile, and therefore, the average wind speed at a given location is a very poor predictor of the energy output of a wind turbine. The basic requirement for wind power estimates is an adequate characterization of the empirical probability distribution of wind speeds since wind direction is less important because of the well developed methods of yaw control for modern turbines [2]. The statistical description is highly simplified when a measured histogram can be accurately fitted by an analytical probability density function (PDF) with a few parameters.

The traditional approach of modeling the wind speed PDF is based on the Rayleigh and the more flexible Weibull distributions [2–8]. However, several authors noted that Weibull fits of empirical data have low quality at several locations, mostly over land [10]. Various analytical forms of skewed distributions were proposed as possible alternatives, such as the lognormal [11,12], square root normal [13,14], chi [11], inverse Gaussian [15], generalized gamma [11], generalized extreme value [16] or extended exponential functions [17,18].

[☆] Wind speed Wind energy Weibull distribution Generalized gamma distribution ERA-40 re-analysis.

^{*} Corresponding author. Tel: +36 1 372 2878; fax: +36 1 372 2866.

E-mail addresses: kisspeter@complex.elte.hu (P. Kiss), janosi@lecco.elte.hu (I.M. János).

In this work, we report on a detailed analysis of surface wind speed distribution over Europe. The main goal was to find an effective and optimal description of the PDF both for terrestrial and sea surface locations. We evaluated wind data from the ERA-40 re-analysis [19] of the European Centre for Medium Range Weather Forecasts (ECMWF). Our tests unambiguously demonstrate that the generalized gamma (GG) distribution provides an improved fit for an overall statistical characterization of surface wind speed. We show that the GG fit outperforms the Weibull function, especially at the large speed tails, which is the most relevant region with respect to wind power estimates.

After the description of the data and the methods used, we summarise tests with various distributions by fitting empirical wind speed histograms. Besides the fundamental Rayleigh, the commonly used Weibull and lognormal distributions, we show the results for the generalized gamma distribution and demonstrate its superiority. Next, we briefly discuss the spectral properties of wind speed records and analyse wind speed fluctuations by removing seasonal periodicities from the original time series. A summary and conclusions are given in the final section.

2. The data and methods

We evaluated the ECMWF's ERA-40 re-analysis data [19] consisting of the u (eastward) and v (northward) orthogonal components of the horizontal wind field at 10 m above ground level. The data base covers a time period of 44 whole years between 1 September 1958 and 31 August 2002. Four instantaneous values are recorded each day for the main synoptic hours of 00, 06, 12

and 18 UTC at each location. The spatial resolution is $1^\circ \times 1^\circ$ (lat/long), and a given value for an atmospheric variable is considered to be representative for the whole cell. Our analysis is restricted to a geographical area covering Europe: 2501 grid points between 35°N and 75°N latitude and 20°W and 40°E longitude.

The main quantity of interest in this study is the scalar wind speed $s = \sqrt{u^2 + v^2}$. For each grid point, the standard statistical characteristics (mean, standard deviation, skewness and kurtosis) are determined, and then the Rayleigh, binormal, Weibull, lognormal and generalized gamma (GG) distributions are fitted by the standard method of maximum likelihood estimates [5,20]. No pre-processing (filtering or removal of seasonal periodicities) was implemented prior to the computations above. The goodness of fit is characterized by computing the coefficient of determination R^2 (the fraction of the total squared error that is explained by the model). The different models are evaluated by comparing the unexplained percentage variance $100(1 - R^2)$ for a given data set.

In order to characterize the temporal behavior of wind speed records, frequency domain analysis is performed by the usual Fourier spectral method. The effects of seasonal periodicities are evaluated by removing the long range average values for a given hour of a given calendar day from the original time series, as usual.

The average value \bar{s} and standard deviation σ_s of ERA-40 surface wind speeds computed over the whole period of 44 years (Fig. 1) illustrate the gross features of wind climatology over Europe. The strong coupling between the values of average speed and standard deviation is apparent, the coefficient of variation is around $\sigma_s/\bar{s} \approx 0.5$, except for a few isolated regions (for example around Corsica).

3. Models for wind speed histograms

3.1. Rayleigh distribution

The most transparent model for scalar wind speed distribution is based on the assumptions that the orthogonal u and v components are independent and identically distributed (iid) Gaussian random variables with zero means and equal standard deviations of $s_0/\sqrt{2}$ (we adopt this notation to get simpler mathematical formulae below). Of course, all the higher moments (skewness, kurtosis etc.) are identically zero. In this case, $s = \sqrt{u^2 + v^2}$ obeys Rayleigh probability density distribution [21] of the functional form

$$P_R(s; s_0) = \frac{2}{s_0} \left(\frac{s}{s_0}\right) \exp\left[-\left(\frac{s}{s_0}\right)^2\right], \quad (1)$$

where the only free parameter is s_0 (the so called scale parameter).

A trivial consequence of the basic assumptions behind a Rayleigh distribution is that the mean vector wind should be zero as

well. However, it is well known that the long range vectorial averages are significantly different from zero, in particular over the oceans (see e.g. [22,23]). Actually, these nonzero values define the prevailing wind systems (e.g. trade winds). This is the first reason why the Rayleigh distribution has a limited applicability, especially for sea winds [8,9].

The mean vector wind, however, is often close to zero over land [22,23]. Therefore, a next plausible test on the validity of the basic assumptions is to determine the normalized third and fourth central moments, the skewness (Sk) and kurtosis (K) for the individual wind components u and v :

$$Sk(x) \approx \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n\sigma_x^3}, \quad K(x) \approx \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n\sigma_x^4} - 3, \quad (2)$$

where x_i is either u_i or v_i , and n is the number of observations. Eq. (2) are not exact equalities because they are biased estimators of skewness and kurtosis, but in this case, the sample size is very large ($n = 64240$) and Eq. (2) can be considered to be very good approximations. The results are shown in Fig. 2. The maps clearly illustrate that very few geographical locations exhibit a pure Gaussian probability distribution ($Sk = 0, K = 0$) for the individual wind vector components.

The standard method to test interdependence of the components u and v is based on computing the correlation coefficient r_{uv} defined as

$$r_{uv} = \frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{(n-1)\sigma_u\sigma_v}, \quad (3)$$

where an overline indicates average value, and σ_u and σ_v are the corresponding standard deviations, as before. Fig. 3 shows that the assumption of independence fails in general; strong correlations of magnitude 0.6–0.8 are present at several geographical locations. (Note that $r_{uv} \approx 0$ does not necessarily mean statistical independence.)

Nonzero correlations are usually taken into account by considering joint probability distributions. For example, when u and v are assumed to be Gaussian random variables with mean values \bar{u} and \bar{v} , standard deviations σ_u and σ_v , and correlation coefficient r_{uv} , then the joint PDF can be written as

$$P(U, V) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-r_{uv}^2}} \exp\left(-\frac{U^2 - 2r_{uv}UV + V^2}{2(1-r_{uv}^2)}\right), \quad (4)$$

where $U = (u - \bar{u})/\sigma_u$ and $V = (v - \bar{v})/\sigma_v$ denote standardized variables.

There are two plausible methods to obtain standardized velocity components U and V . First, the mean values \bar{u} and \bar{v} can be computed over the whole length of the time series assuming a well defined prevailing wind. The second way is to consider velocity

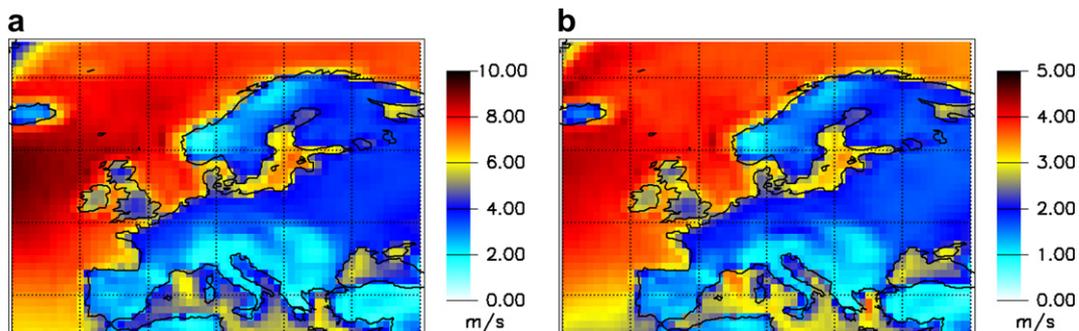


Fig. 1. (a) Average value, and (b) standard deviation of ERA-40 surface wind speeds in the period 1958–2002, in units of m/s. Note that the color scales are different by a factor of 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

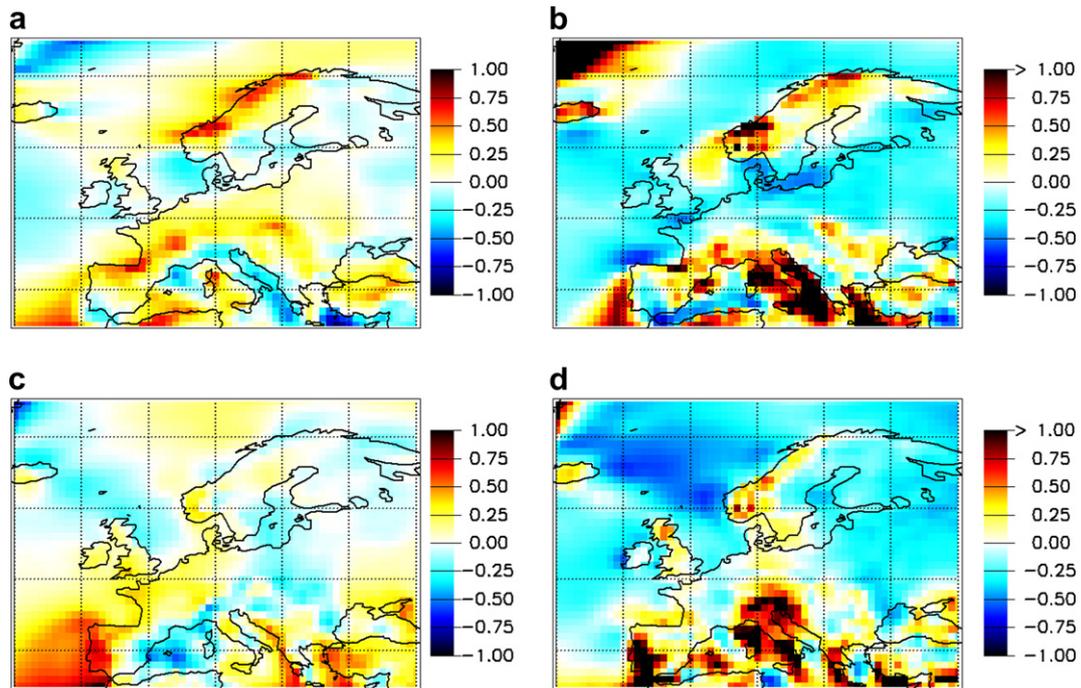


Fig. 2. Geographical distribution of empirical skewness Sk and kurtosis K (see Eq. 2) for the wind components u and v . (a) $Sk(u)$, (b) $K(u)$, (c) $Sk(v)$ and (d) $K(v)$. White color emphasizes $Sk = 0$ and $K = 0$ levels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

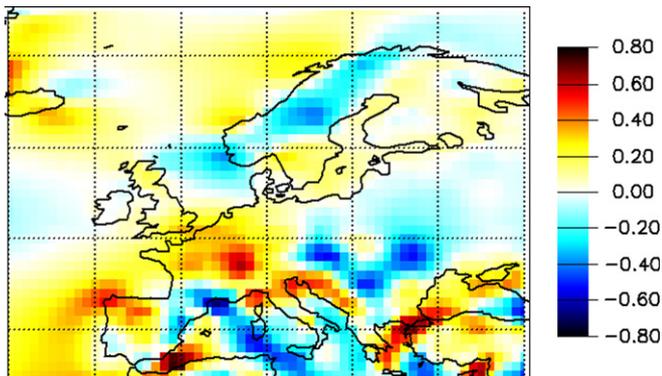


Fig. 3. Correlation coefficient r_{uv} (see Eq. 3) for the u and v wind velocity components. White color emphasizes $r_{uv} = 0$ level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

given calendar day. (For further discussion see Section 4). We tested both cases.

Fig. 4 illustrates the quality of the bivariate Gaussian model fits for the two methods of standardization. The regions of relatively low values of the unexplained percentage variance $100(1 - R^2)$ indicate where the model performs well. We note that most of the deviations between the model and data are due to a higher empirical probability of large wind speeds with respect to a Gaussian distribution. At several locations, the exponential tails $P(s) \propto e^{-s}$ for large values of s provide a better fit than the Gaussian decay $P(s) \propto e^{-s^2}$. Higher values of kurtosis in Fig. 2 refer to this feature as well.

In view of the above results, we should not expect the Rayleigh model to provide an adequate universal description for the empirical wind speed distributions.

3.2. Weibull distribution

The most widely accepted model for wind speed probabilities is the two parameter Weibull distribution [24]:

fluctuations around the annual periodic background signal, which is determined by the average value for a given synoptic hour and

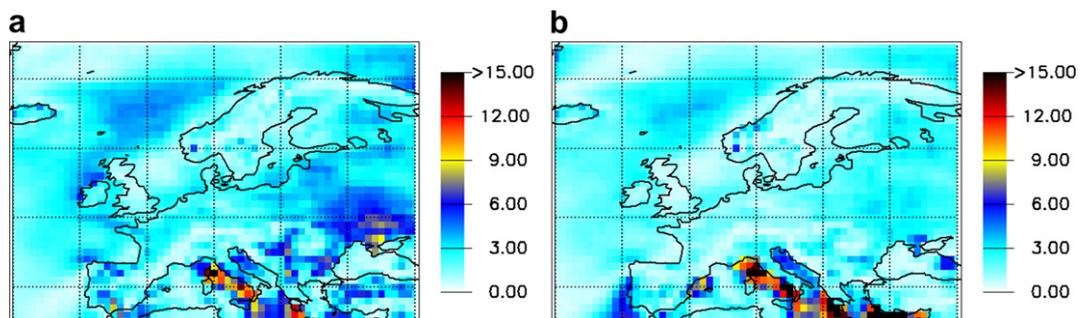


Fig. 4. Unexplained percentage variance $100(1 - R^2)$ of joint Gaussian PDF fits by Eq. (4). (a) Full time average values removed. (b) Daily and annual periodic cycle removed.

$$P_W(s; s_0, k) = \frac{k}{s_0} \left(\frac{s}{s_0}\right)^{k-1} \exp\left[-\left(\frac{s}{s_0}\right)^k\right], \quad (5)$$

where s_0 and k denote the scale and the shape parameters, respectively. It is easy to see that the Rayleigh distribution Eq. (1) is a special case of the Weibull distribution with $k \equiv 2$, while $k \equiv 1$ gives the simple exponential distribution. The Weibull PDF Eq. (5) can be derived theoretically as a form of extreme value distributions [25], and it is the most popular model for failure rate distributions [21,24]. In the context of wind speed histograms, we can rather consider the functional form Eq. (5) as a generalization of the Rayleigh distribution, which provides an increased flexibility to fit empirical data. The properties of the Weibull distribution have been thoroughly studied and numerous studies show that it works well for modeling wind speeds at several locations [2–9,26,13,27–32].

Examples of Weibull fits are shown in Fig. 5, where the geographical locations were chosen to illustrate particularly good and poor fits. In general, the Weibull PDF is a reasonably good model over the ocean and seas, however, histograms for large areas over land can be fitted with rather large errors. Fig. 5b illustrates that neither the central part nor the large speed tail of the empirical histograms are close enough to a Weibull distribution. The latter is more problematic in practice because wind energy is only produced above the cut in speed of wind turbines (typical values are 3–5 m/s), and therefore, an adequate fit of the large speed tail is crucial for wind energy estimates.

The geographical distribution of fitted Weibull parameters is shown in Fig. 6. The scale parameter s_0 shows almost the same pattern as the mean wind speed \bar{s} in Fig. 1a. This is plausible, because the mean value of a Weibull distribution is given by $s_0 \Gamma(1 + 1/k)$,

where the correction factor of the Gamma function changes in the interval $\Gamma \in [0.888, 0.911]$ for the fitted shape parameter regime $k \in [1.4, 2.6]$. The shape parameter k exhibits a more interesting spatial pattern in Fig. 6b. Lower numerical values indicate a slower decay of the large speed tail, and such shape is characteristic of areas of very low average speed \bar{s} or scale parameter s_0 ; however, the relationship is not entirely strict.

A straightforward generalization of the two parameter Weibull PDF is the three parameter Weibull distribution

$$P_W(s; \mu, s_0, k) = \frac{k}{s_0} \left(\frac{s-\mu}{s_0}\right)^{k-1} \exp\left[-\left(\frac{s-\mu}{s_0}\right)^k\right], \quad (6)$$

where the third (location) parameter μ shifts the Weibull peak horizontally. This helps to model wind speed histograms at calm local climates, where $s = 0$ has a significant positive probability. However, such places are exceptional, and therefore, the general improvement of fits by Eq. (6) is marginal.

Since the Weibull distribution does not provide a universally good description of wind speeds everywhere, we tested other empirical distributions too. For example, the lognormal distribution ($\ln(s)$ exhibits Gaussian PDF) was extensively used to fit wind speed histograms over land [12,30]. Our results show that the lognormal distribution is an inappropriate model over most of the areas examined. In general, a lognormal distribution decays much slower towards high wind speeds than the empirical data. Nevertheless, at a few gridpoints (e.g. northern Germany), we found that the lognormal model fits better than the Weibull, see Fig. 7. A typical histogram at these gridpoints is characterized by a sharp peak and an exponential like right tail, which is closer to the lognormal behavior (Fig. 7).

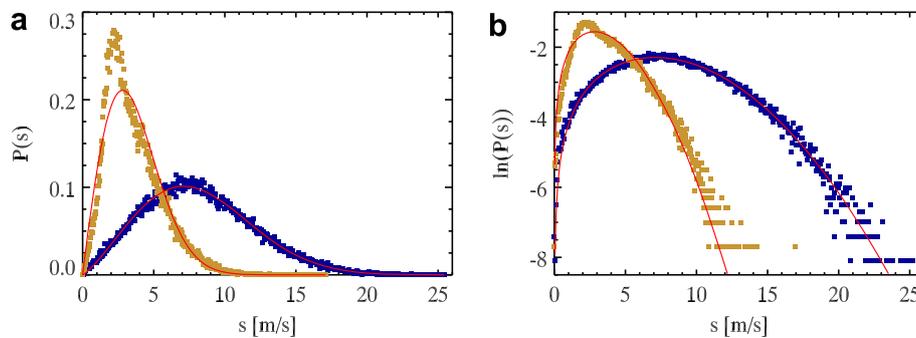


Fig. 5. Normalized histogram of wind speed s at two distinct locations: Atlantic Ocean (66°N 4°W, dark blue), and northern Germany (52°N 11°E, light brown). Maximum likelihood Weibull fits are shown with straight lines. (a) Linear scale, and (b) semi-log scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

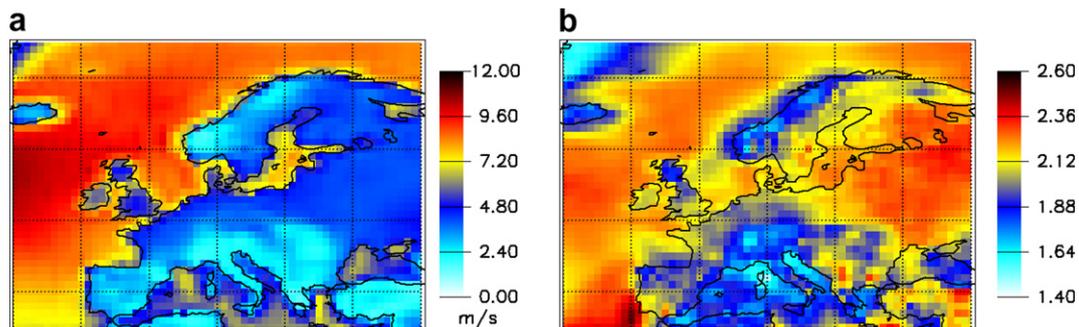


Fig. 6. Geographical distributions of the (a) scale s_0 and (b) shape k parameters of maximum likelihood Weibull fits (see Eq. (5)).

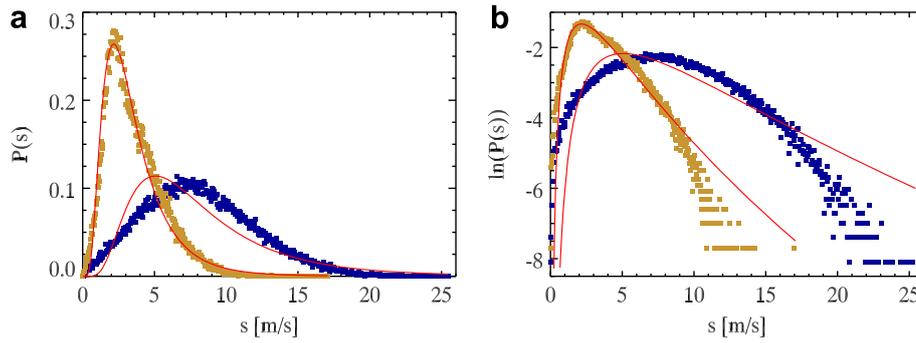


Fig. 7. Lognormal fits for the same normalized histograms as shown in Fig. 5. (a) Linear scale and (b) semi-log scale.

3.3. Generalized gamma distribution

In order to overcome the constraint that both tails of a Weibull peak are described by a single shape parameter k , we can further generalize Eq. (5) to get the generalized gamma (GG) distribution, also suggested by Auwera et al. [11]:

$$P_{GG}(s; s_0, k, \epsilon) = \frac{k}{s_0 \Gamma(\epsilon)} \left(\frac{s}{s_0}\right)^{\epsilon k - 1} \exp\left[-\left(\frac{s}{s_0}\right)^k\right], \quad (7)$$

where the Gamma correction $\Gamma(\epsilon)$ is required for normalization, and the new parameter ϵ improves the shape flexibility. Special cases are $\epsilon \equiv 1$ (the original Weibull distribution), and $k \equiv 1$ which gives the gamma distribution. The lognormal distribution can also be obtained as a limiting distribution when $\epsilon \rightarrow \infty$ [33]. The fixed parameter value $k \equiv 2$ defines a subfamily of GG distributions, which is known as the generalized normal (GN) distribution. The GN distribution is itself a flexible family and includes the Half normal ($\epsilon = 1/2$), Rayleigh ($\epsilon = 1$), Maxwell–Boltzmann ($\epsilon = 3/2$) and chi ($\epsilon = n/2; n = 1, 2, 3, \dots$) distributions.

The function P_{GG} (Eq. (7)) has a single maximum (mode m) at

$$m = s_0 \left(\epsilon - \frac{1}{k}\right)^{\frac{1}{k}}. \quad (8)$$

The effects of the two shape parameters k and ϵ cannot be fully separated, but nevertheless, the asymptotic behavior on the left side is a power law with exponent $q = \epsilon k - 1$, while it is a stretched exponential on the right:

$$\begin{aligned} P_{GG}(s) &\propto s^q & 0 \leq s \ll m, \\ P_{GG}(s) &\propto e^{-(s/s_0)^k} & s \gg m. \end{aligned} \quad (9)$$

The α -th non-central moment of the GG distribution can be obtained easily:

$$\langle s^\alpha \rangle = s_0^\alpha \frac{\Gamma(\epsilon + \alpha/k)}{\Gamma(\epsilon)}. \quad (10)$$

An important property of GG distributions is that the family is closed under power transformations. That is, if $s > 0$ obeys $P_{GG}(s; s_0, k, \epsilon)$, then the corresponding PDF for $z = s^p$ has the form $P_{GG}(z; s_0^p, k/p, \epsilon)$ [33]. This property can be immediately exploited for estimating potential wind power, since it is proportional to the cube of the wind speed s^3 [2,4].

The GG parameters were obtained for the empirical data by the maximum likelihood method, as before. The corresponding set of nonlinear equations were solved by the Newton–Raphson algorithm [34] with initial guesses for the parameters provided by Weibull fits.

Our results show that the GG PDF Eq. (7) provides a significantly improved fit compared to the Weibull model Eq. (5). The improvement is especially spectacular at the right tail (high wind speeds), which is correctly fitted in each case (Fig. 8). Nevertheless, even the generalized gamma distribution cannot capture all the features of the measured histograms over a few geographical areas. Such a problematic region is the northern coast of the Black sea, especially around the Crimean peninsula. Further details and a possible climatological explanation are given in Section 4.

The values of the fitted parameters at different locations can be seen in Fig. 9. The peak maximum m given by Eq. (8) exhibits practically the same pattern as the average wind speed \bar{s} (Fig. 1a) or the Weibull scale parameter s_0 (Fig. 6a), so we do not show the map again. The left tail shape parameter $q = \epsilon k - 1$ (Fig. 9a) is approximately 1 over the seas, which implies a close to linear increase of probabilities for low wind speeds. The corresponding characteristic values are definitely larger over land, typically around $q \approx 2$. The geographical pattern for the right tail shape parameter k (Fig. 9b) has the opposite tendency: typical values around 2 are characteristic over most of the offshore areas, whereas smaller values closer to

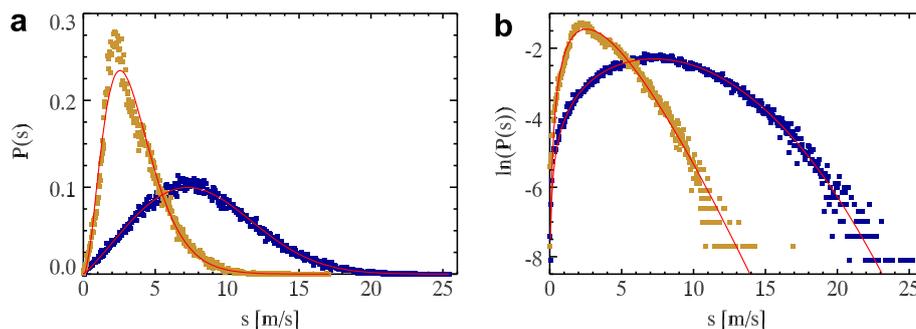


Fig. 8. Maximum likelihood GG fits (see Eq. (7)) for the same normalized histograms as shown in Figs. 5 and 7. (a) Linear scale and (b) semi-log scale.

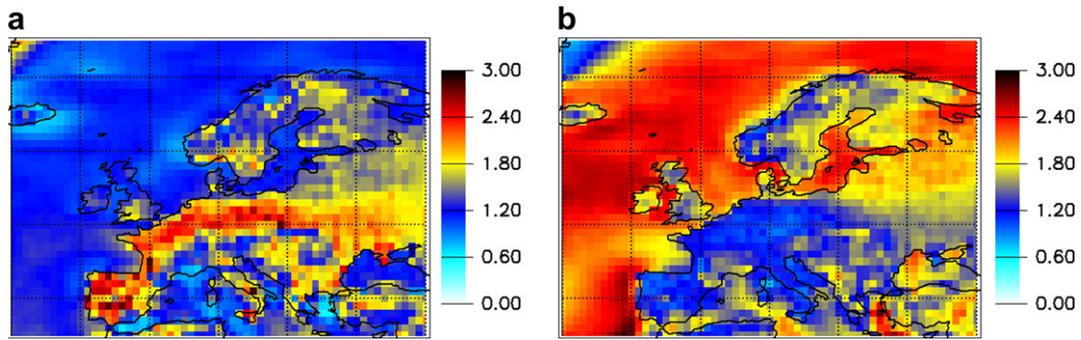


Fig. 9. Geographical distribution of fitted GG parameters (see Eq. (7)). (a) Left tail parameter $q = \epsilon k - 1$ and (b) right tail shape parameter k .

1 over land indicate a slower decay of large wind speed probabilities.

A possible way of confronting the goodness of fit for Weibull and generalized gamma distributions is to compare the spatial patterns of an error parameter, such as the residual (or unexplained) percentage variance $100(1 - R^2)$. The results shown in Fig. 10 convincingly demonstrate that the generalized gamma distribution provides a better tool to fit wind speed histograms over various surface conditions. Note that the range of color scale in Fig. 10 is identical to the scale in Fig. 4, which illustrates that both the Weibull and GG models are superior compared to any Gaussian description that assumes linear correlation between u and v wind velocity components.

The maps of the GG parameters distribution in Fig. 9 showing large coherent geographical areas suggest that the different shapes

for the wind speed histograms have a physical origin. Further insight can be gained by inspecting the correlation plots of the fitted GG parameters $q = \epsilon k - 1$, k and m ; the results are shown in Fig. 11.

The first inference to be drawn in Fig. 11a is that only a minority of histograms coincide with the constrained distributions of Weibull ($k = q + 1$), gamma ($k = 1$) or Rayleigh ($q = 1, k = 2$) forms, the rest have significantly different shape parameters for the left and right tails. Furthermore, the separation of clusters in the parameter space corresponding to histograms over sea surface and over land is remarkable. The overlap between the shape parameters is characteristic for relatively small islands (e.g. Ireland) and restricted maritime regions (e.g. Adriatic sea). Fig. 11b illustrates the interdependence between the left tail shape parameter q and mode m ; the separation for sites over sea and land is clear again. From the point of view of wind power generation,

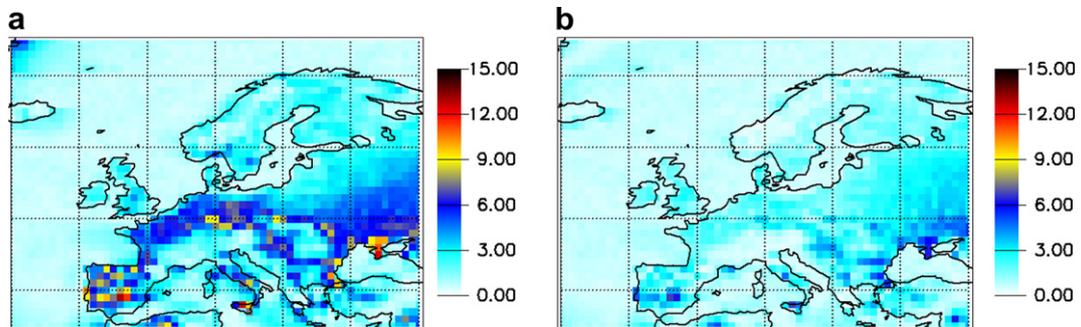


Fig. 10. Unexplained percentage variance $100(1 - R^2)$ for (a) Weibull [Eq. (5)] and (b) GG [Eq. (7)] fits.

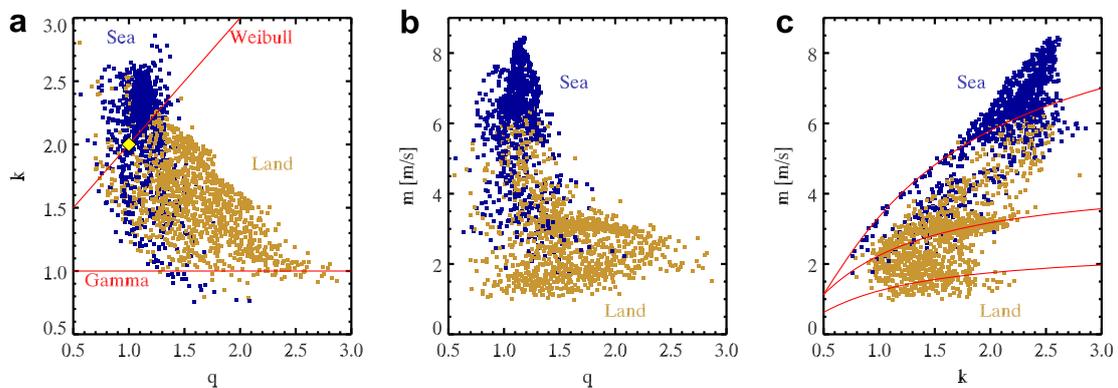


Fig. 11. Correlation plots for the fitted parameters of GG distribution (see Eq. (7)): (a) $q = \epsilon k - 1$ vs. k , straight lines denote the special cases of gamma and Weibull distributions, and a diamond symbol denotes the Rayleigh distribution; (b) q vs. m ; and (c) k vs. m . For an explanation of the continuous lines in (c) see the text. Land and sea surface areas are distinguished by different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

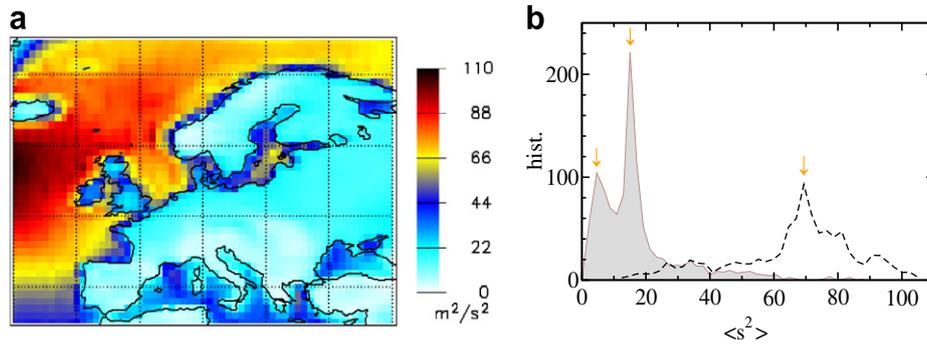


Fig. 12. The second non-central moment $\langle s^2 \rangle$ calculated from Eq. (10) using GG fitted parameters. (a) Geographical distribution. (b) Unnormalized histograms for maritime (dashed line) and terrestrial (gray shading) locations. Characteristic values $\langle s^2 \rangle$ of the peaks are indicated with arrows (see text). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the correlation plot in Fig. 11c is the most interesting. This illustrates the relationship between the right tail shape parameter k and the mode m (essentially the mean wind speed). The optimal parameter regime is located somewhere in the upper right quarter of the diagram Fig. 11c, where the mean wind speed is well over the cut in value of 5 m/s, and the decay of probabilities of very large speeds is fast. The latter property is beneficial because of the smaller frequency of high loads on the wind turbine's tower.

For a further analysis of Fig. 11c, let us suppose that the probability of wind speeds obeys the GG distribution Eq. (7) with a constant second (non-central) moment $\langle s^2 \rangle = C$, characterizing large coherent geographical regions. An invariant second moment means that the average momentum flow rate is conserved. (The instantaneous mass flow rate is given by $\dot{m} = \rho A s_i$, where ρ is the density, A is the cross sectional area and s_i is the instantaneous flow speed. The momentum flow rate is simply $\dot{m} s_i$; its average value over a given time interval is clearly related to $\langle s^2 \rangle$.) In order to check the behavior of the second moment, we determined its geographical distribution (Fig. 12a) and the corresponding histograms separately over land and seas (Fig. 12b). The histograms in Fig. 12b reveal a characteristic second moment $\langle s^2 \rangle \approx 67 \text{ m}^2/\text{s}^2$ for maritime locations (with much lower values around the coastlines), and a bimodal distribution for land with two local maxima at $\sim 5 \text{ m}^2/\text{s}^2$ for high altitudes and $\sim 15 \text{ m}^2/\text{s}^2$ for the rest of the continent.

Next, we return to Fig. 11b. As it is already noticed, the left tail shape parameter $q = \epsilon k - 1$ is close to 1 for most maritime locations, irrespective of the other parameter values. Much larger variations are observed over land; nevertheless, we can fix a typical value somewhere around $q = 2$. This permits a consistency check of the results: when the second moment $\langle s^2 \rangle$ (see Fig. 12b) and shape parameter q are fixed, Eqs. (8) and (10) provide a relationship to express the mode m as a function of k . The thin curves in Fig. 11c illustrate the expected behavior with the fixed parameter values for sea and land; the agreement is quite satisfactory. Note that this procedure does not favor the second moment of $\alpha = 2$ in Eq. (10); practically any of them would work. However, we found that the histogram of the second moments (Fig. 12b) exhibits the best separation of peaks, and therefore, it is suitable to estimate characteristic values.

4. Temporal behavior of wind speed records

The probability density distribution of wind speed represents aggregated statistical information; however, it contains no information about possible temporal patterns, which is equally important for practical purposes. The standard method to reveal such patterns is Fourier spectral analysis. Since the ERA-40 records

contain no missing intervals, we could exploit the speed of the FFT (fast Fourier transform) procedure [34] in our analysis.

The power spectra of wind records exhibit a fairly simple structure; an example is shown in Fig. 13. Only two characteristic cycles could be extracted from all of the data: annual and daily periods. The wide continuum background can be fitted by a Lorentzian spectrum (solid line in Fig. 13)

$$S(f) = \frac{2\tau}{1 + (2\pi f)^2 \tau^2}, \quad (11)$$

where τ is a characteristic time related to an exponential decay of the autocorrelation function as

$$A(\Delta t) = \langle s(t)s(t + \Delta t) \rangle \propto \exp\left(-\frac{\Delta t}{\tau}\right). \quad (12)$$

Typical values for the exponential decay time τ are between 1.5 and 4.0 days.

As for the identified periodicities, the daily cycle is missing from most of the records, and even the annual course can not be resolved at several places. In order to visualize the strength of the periodicities, we determined the partial power of the periodic components by integrating the area under the peaks shown in Fig. 13 for normalized power spectra. The geographical distributions are illustrated in Fig. 14. Numerical values close to zero mean that the given peak fades into the background. However, the maps in Fig. 14a should be considered with care because the low sampling

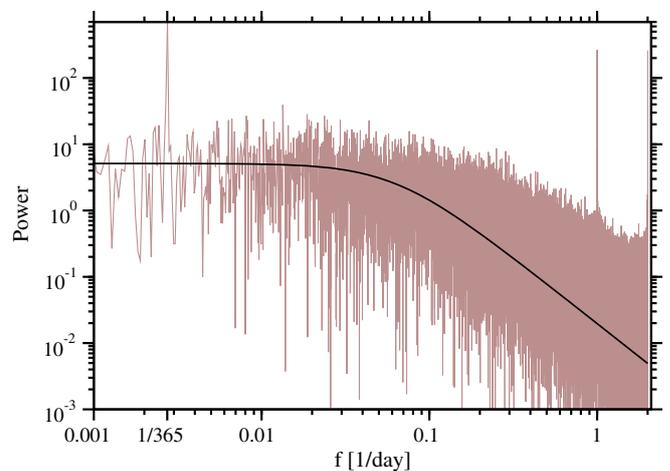


Fig. 13. Normalized power density spectrum of the wind speed record over northern Germany (52°N, 11°E). The solid line shows a Lorentz fit [Eq. (11)].

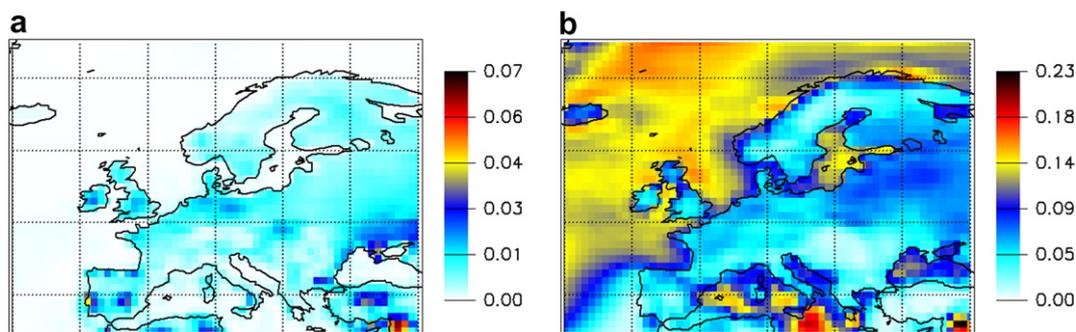


Fig. 14. Partial power for the (a) daily and (b) annual periodicities extracted from normalized FFT power spectra of wind speed records. Note the different color scales. (By definition, the total integrated power has unit value). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

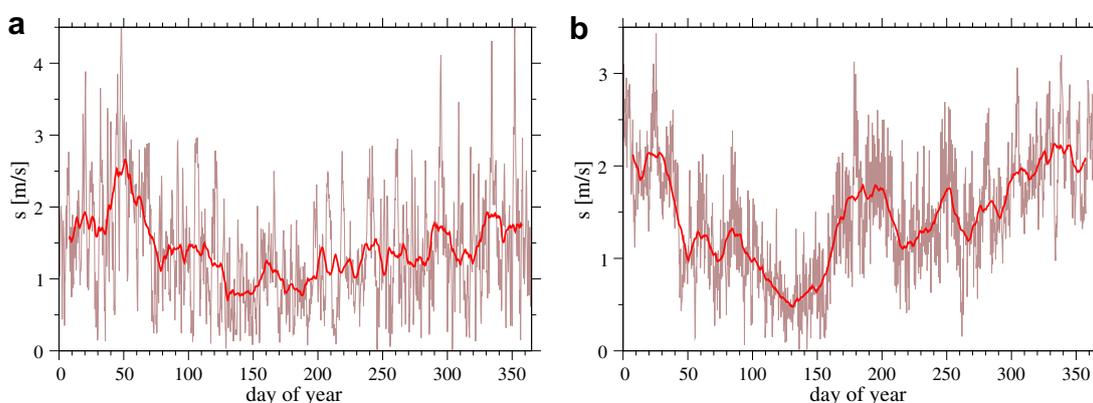


Fig. 15. Annual wind speed cycles (a) over the Atlantic Ocean (66°N, 4°W) and (b) over northern Germany (52°N, 11°E). Brown lines represent the 44 year average for a given synoptic hour and calendar day and a thick red line shows the 15 day moving average. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

frequency and the low spatial resolution can obviously hinder resolution of the daily cycle at many places close to coastlines.

Fig. 15 illustrates typical annual cycles computed over the 44 years. The most remarkable features are that the periodicity is very weak (the amplitude of the 15 day moving average is less than half the mean wind speed \bar{s} for the given locations); furthermore, it is loaded with extremely large day to day fluctuations. Such large fluctuations mean that the annual cycle is present only as a weak change of the boundary conditions determining the main atmospheric flow features, and it is almost fully masked by the short range natural fluctuations of turbulent airflow.

In view of the above results, we can consider the time series of wind speeds as a superposition of three signals related to different physical mechanisms. The slowest component is a weak annual periodic background determined by the changing global insolation. Far the strongest component is determined by synoptic scale meteorological features (extratropical cyclones and anticyclones) and is loaded with short range turbulent fluctuations excited by surface roughness and thermal convection. The third component, representing the daily periodicities, is characteristic of coastal locations where the well known land–sea temperature contrast often induces daily periodic winds.

This picture helps to explain the apparent breaks in many histograms, clearly visible also in the near coast record in Figs. 5, 7 and 8 at $s = 4.5$ m/s. Such histograms can be perfectly decomposed into a superposition of a narrow Gaussian (daily coastal wind), and a wide GG distribution representing the large scale flow. Since this decomposition includes 6 free parameters to be fitted, its usefulness is limited in the practice.

Nevertheless, it is worth checking the effects of removing the background periodicities, thus considering wind speed fluctuations s' around various trends. The first procedure we implemented removes constant average values \bar{u} and \bar{v} separately from the velocity components and produces the fluctuation series

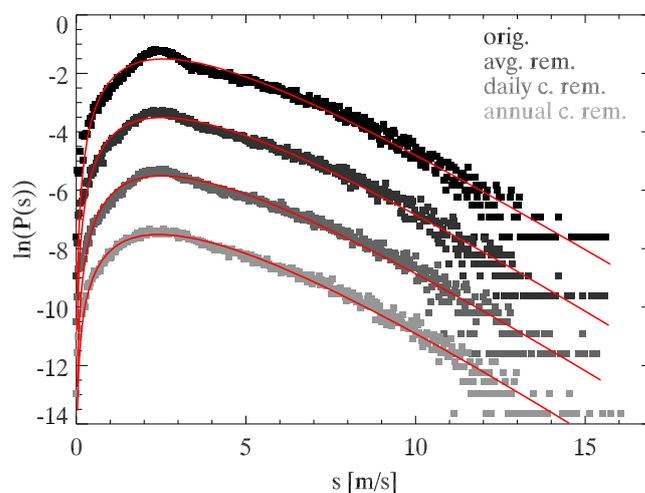


Fig. 16. Semi-log scale wind speed histograms at the Crimean peninsula (45°N, 34°E). From top to bottom: unfiltered data, constant removed [Eq. (13)], daily cycle removed [Eq. (14)], and daily and annual cycle removed [Eq. (15)]. Continuous lines represent generalized gamma fits. The curves are vertically shifted by a value of 2 with respect to each other. The correct scale corresponds to the uppermost curve.

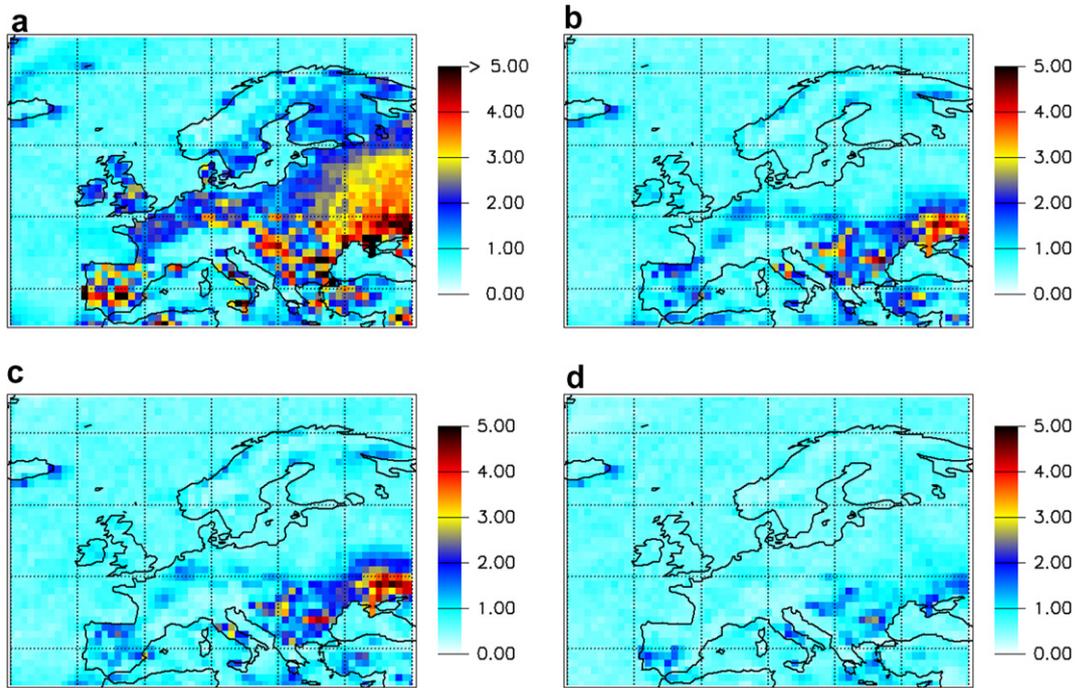


Fig. 17. Residual percentage variance $100(1 - R^2)$ of generalized gamma fits. (a) unfiltered wind speed s , (b) constant trend removed s'_c [Eq. (13)], (c) daily cycle removed s'_d [Eq. (14)] and (d) daily and annual cycle removed s'_g [Eq. (15)]. Note that the color scale is much narrower than in Fig. 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$s'_c(t) = \sqrt{[u(t) - \bar{u}]^2 + [v(t) - \bar{v}]^2}. \quad (13)$$

Note that the results shown in Fig. 4a were obtained by the same method [Eq. (13)]. Next, the daily cycle can be removed by computing the average values for each synoptic hours separately as $\bar{u}(h) = 44^{-1} \cdot 365^{-1} \sum_{y,d} u(y, d, h)$ and $\bar{v}(h)$ similarly and obtaining

$$s'_d(y, d, h) = \sqrt{[u(y, d, h) - \bar{u}(h)]^2 + [v(y, d, h) - \bar{v}(h)]^2}, \quad (14)$$

where the time variable t is replaced by the full calendar indices $y = 1 \dots 44$ (year), $d = 1 \dots 365$ (day) and $h = 1 \dots 4$ (hour). Finally, the daily and annual periodicity can be removed by obtaining average values for the velocity components as $\bar{u}(d, h) = 44^{-1} \sum_y u(y, d, h)$ and $\bar{v}(d, h)$ similarly for a given day and hour in a year. The fluctuations around the daily and annual cycles are given as

$$s'_g(y, d, h) = \sqrt{[u(y, d, h) - \bar{u}(d, h)]^2 + [v(y, d, h) - \bar{v}(d, h)]^2}. \quad (15)$$

Note that the results shown in Fig. 4b were obtained by this method [Eq. (15)].

The effects of gradually removing a constant, daily, and annual cycles from the original records is demonstrated in Fig. 16. It is obvious that wind speed fluctuations are much closer to a single generalized gamma distribution than the original wind speed s . This indicates that the main component of wind, determined by the synoptic scale atmospheric patterns, obeys a GG probability density distribution indeed. The improvement works almost globally, as illustrated in Fig. 17.

The spatial distribution of fitted parameters and the correlation plots between them are very similar to the results obtained without the removal of averages (see Figs. 9 and 11), and therefore, the corresponding figures for wind speed fluctuations are not shown.

5. Conclusions

As a first step of an extensive assessment of wind energy potential over Europe, we investigated the wind speed distribution based

on the 44 years long ERA-40 data base. Our aim was to give a uniform description of wind speed as a random variable and to compare the performance of the widely used analytical PDFs. The advantage of these model PDFs is that a few parameters (typically 1–3) can capture most features of the observed wind speed histograms. We estimated parameters using the maximum likelihood method.

The simplest physically plausible model predicts wind speed to follow a Rayleigh distribution. Since we have simultaneous data of the wind velocity components, it was possible to show that the assumptions of the Rayleigh model are, in general, not fulfilled by real winds. Neither do the components follow Gaussian distribution nor are they uncorrelated. By dropping the strict assumptions of the Rayleigh model but still assuming Gaussian wind velocity components, the resulting model is still unable to give a general description of wind speed over Europe.

As a next step, we tested the most commonly used two parameter Weibull PDF. It is known to perform well in describing real wind speed distributions at individual locations due to its shape flexibility. We found that the Weibull model can indeed adequately characterize wind histograms over the seas and over some parts of land areas as well. However, it fits poorly over major inland areas, such as northern Germany for example, which is an extremely important area concerning wind energy production. The popularity of this distribution is probably due to its simplicity, and histograms of short records might contain so much noise that there is no need for a more accurate description with alternative PDFs.

Part of the Weibull model's limitations can be overcome by a simple generalization leading to the generalized gamma (GG) distribution. This model has three parameters, two of which are shape parameters accounting for the increased flexibility of the PDF. We found that the GG model fits well almost everywhere, especially if we focus on the high speed tail of the distribution, which is essential in wind power estimations. A closer look at "anomalous" histograms and a frequency domain analysis revealed that wind records are sort of a superposition of three signals (besides the high fre-

quency components of turbulence, which are inaccessible due to the low sampling frequency of the data). Two of these signals are periodic: a weak annual cycle can be observed almost everywhere and a daily cycle over coastal and inland areas. After removing these periodicities, we get the main wind component. These are wind speed fluctuations mainly due to synoptic scale meteorological phenomena. We showed that these fluctuations are almost perfectly described by the generalized gamma model.

Acknowledgements

We thank the Hungarian Meteorological Service for access to the ERA-40 data bank. This work was supported by the European Commission's DG RTD NEST Programme "Tackling Complexity in Science" (Contract No. 043363). IMJ is thankful for a János Bolyai research scholarship of the Hungarian Academy of Sciences.

References

- [1] The European Wind Energy Association; 2007 <<http://www.ewea.org>>.
- [2] Burton T, Sharpe D, Jenkins N, Bossanyi E. Wind energy handbook. Chichester: John Wiley & Sons; 2001.
- [3] Patel MR. Wind and solar power systems: design, analysis, and operation. 2nd ed. Oxford: Taylor & Francis; 2005.
- [4] Hau E. Wind turbines: fundamentals, technologies, application, economics. 2nd ed. Berlin: Springer; 2005.
- [5] Conradsen K, Nielsen LB. Review of Weibull statistics for estimation of wind speed distributions. *J Appl Meteor* 1984;23:1173–83.
- [6] Seguro JV, Lambert TW. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *J Wind Eng Ind Aerodyn* 2000;85:75–84.
- [7] Ramirez P, Carta JA. Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution. A case study. *Energy Convers Manage* 2005;46:2419–38.
- [8] Monahan AH. The probability distribution of sea surface wind speeds. Part I: Theory and sea winds observations. *J Climate* 2006;19:497–520.
- [9] Monahan AH. The probability distribution of sea surface wind speeds. Part II: Dataset intercomparison and seasonal variability. *J Climate* 2006;19:521–34.
- [10] Erickson DJ, Taylor JA. Non-Weibull behavior observed in a model-generated global surface wind field frequency distribution. *J Geophys Res C* 1989;94:12693–8.
- [11] Van der Auwera L, De Meyer F, Malet LM. The use of the Weibull three parameter model for estimating mean wind power densities. *J Appl Meteor* 1980;19:819–25.
- [12] Luna RE, Church HW. Estimation of long term concentrations using a "universal" wind speed distribution. *J Appl Meteor* 1974;13:910–6.
- [13] Justus CG, Hargraves WR, Mikhail A, Graber D. Methods for estimating wind speed frequency distributions. *J Appl Meteor* 1978;17:350–3.
- [14] Brown BG, Katz RW, Murphy AH. Time series models to simulate and forecast wind speed and wind power. *J Appl Meteor* 1984;23:1184–95.
- [15] Bardsley WE. Note on the use of the inverse Gaussian distribution for wind energy applications. *J Appl Meteor* 1980;19:1126–30.
- [16] Bauer E. Characteristic frequency distributions of remotely sensed in situ and modelled wind speeds. *Int J Climatol* 1996;16:1087–102.
- [17] Li M, Li X. MEP type distribution function: a better alternative to Weibull function for wind speed distributions. *Renew Energy* 2005;30:1221–40.
- [18] Akpinar S, Akpinar EK. Wind energy analysis based on maximum entropy principle (MEP)-type distribution function. *Energy Convers Manage* 2007;48:1140–9.
- [19] Uppala SM, Källberg PW, Simmons AJ, Andrae U, Da Costa Bechtold U, Fiorino M, et al. The ERA-40 re-analysis. *Quarter J Roy Meteorol Soc* 2005;131:2961–3012.
- [20] Ghosh A. A FORTRAN program for fitting Weibull distribution and generating samples. *Comput Geosci* 1999;25:729–38.
- [21] Cohen AC, Whitten BJ. Parameter estimation in reliability and life span models. Boca Raton: CRC Press; 1988.
- [22] Hantel M, editor. Observed global climate. Berlin: Springer; 2005.
- [23] <<http://www.ecmwf.int/research/era/>> ERA-40 Atlas/docs/section_B/parameter_10mwwi.html (ERA-40 Atlas) 2007.
- [24] Dodson B. The Weibull analysis handbook. 2nd ed. Milwaukee: ASQ Quality Press; 2006.
- [25] Coles S. An introduction to statistical modeling of extreme values. New York: Springer; 2001.
- [26] Justus CG, Hargraves WR, Yalcin A. Nationwide assessment of potential output from wind powered generators. *J Appl Meteor* 1976;15:673–8.
- [27] Takle ES, Brown JM. Note on the use of Weibull statistics to characterize wind speed data. *J Appl Meteor* 1978;17:556–9.
- [28] Stewart DA, Essenwanger OM. Frequency distribution of wind speed near the surface. *J Appl Meteor* 1978;17:1633–42.
- [29] Tuller SE, Brett AC. The characteristics of wind velocity that favour the fitting of Weibull distribution in wind speed analysis. *J Appl Meteor* 1984;23:124–34.
- [30] Garcia A, Torres JL, Prieto E, De Francisco A. Fitting wind speed distributions: a case study. *Solar Energy* 1998;62:139–44.
- [31] Celik AN. Assessing the suitability of wind speed probability distribution functions based on wind power density. *Renew Energy* 2003;28:1563–74.
- [32] Chang TJ, Wua YT, Hsua HY, Chub CR, Liao CM. Assessment of wind characteristics and wind turbine characteristics in Taiwan. *Renew Energy* 2003;28:851–71.
- [33] Dadpay A, Soofi ES, Soyer R. Information measures for generalized gamma family. *J Econom* 2007;138:568–85.
- [34] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes in C. 2nd ed. Cambridge: Cambridge University Press; 1992.