Surface Wind-Speed Statistics Modelling: Alternatives to the Weibull Distribution and Performance Evaluation
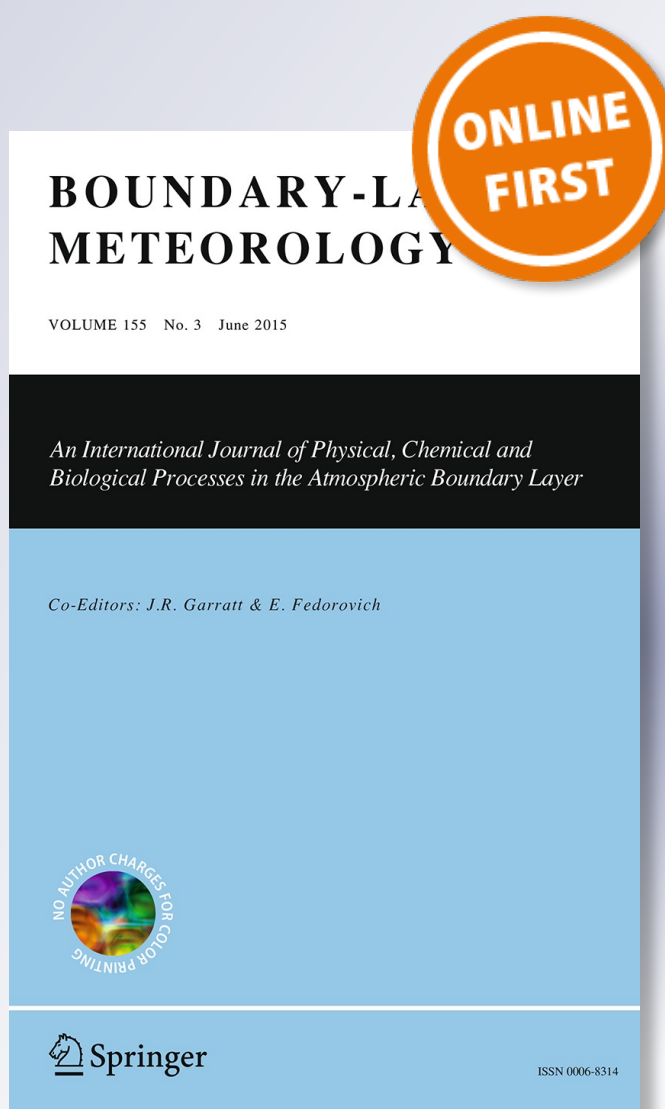
Philippe Drobinski, Corentin Coulais & Bénédicte Jourdier

BOUNDARY-LAYER METEOROLOGY

VOLUME 155   No. 3   June 2015

An International Journal of Physical, Chemical and Biological Processes in the Atmospheric Boundary Layer

Co-Editors: J.R. Garratt & E. Fedorovich

NO AUTHOR CHARGES FOR COLOR PRINTING

ONLINE FIRST

Springer

ISSN 0006-8314

Springer

Springer

CrossMark

ARTICLE

Surface Wind-Speed Statistics Modelling: Alternatives to the Weibull Distribution and Performance Evaluation

**Philippe Drobinski[1] · Corentin Coulais[2] ·
Bénédicte Jourdier[1,3]**

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Wind-speed statistics are generally modelled using the Weibull distribution. However, the Weibull distribution is based on empirical rather than physical justification and might display strong limitations for its applications. Here, we derive wind-speed distributions analytically with different assumptions on the wind components to model wind anisotropy, wind extremes and multiple wind regimes. We quantitatively confront these distributions with an extensive set of meteorological data (89 stations covering various sub-climatic regions in France) to identify distributions that perform best and the reasons for this, and we analyze the sensitivity of the proposed distributions to the diurnal to seasonal variability. We find that local topography, unsteady wind fluctuations as well as persistent wind regimes are determinants for the performances of these distributions, as they induce anisotropy or non-Gaussian fluctuations of the wind components. A Rayleigh–Rice distribution is proposed to model the combination of weak isotropic wind and persistent wind regimes. It outperforms all other tested distributions (Weibull, elliptical and non-Gaussian) and is the only proposed distribution able to catch accurately the diurnal and seasonal variability.

1 Introduction

Understanding and modelling wind-speed statistics is key to a better understanding of atmospheric turbulence and diffusion, and at stake in practical applications such as air qual-

✉ Philippe Drobinski
philippe.drobinski@lmd.polytechnique.fr

1   Laboratoire de Météorologie Dynamique - Institut Pierre Simon Laplace, CNRS and Ecole Polytechnique, Palaiseau, France

2   Huygens-Kamerlingh Onnes Lab, Universiteit Leiden, PO Box 9504, 2300 RA Leiden, The Netherlands

3   French Environment and Energy Management Agency, 20 avenue du Grésillé, BP 90406, 49004 Angers Cedex 01, France

ity and pollution transport modelling, estimation of wind loads on buildings, prediction of atmospheric or space probe and missile trajectory, and wind-power analysis. This is done using the distribution of wind speed M. The Weibull distribution is an extremely commonly used paradigm used to model wind-speed statistics (e.g. Justus et al. 1976, 1978; Seguro and Lambert 2000; Cook 2001; Weisser 2003; Ramírez and Carta 2005; Gryning et al. 2014), and is defined as

$$p(M) = \frac{k}{\lambda} \left(\frac{M}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{M}{\lambda}\right)^k\right], \tag{1}$$

where M is the wind speed, $k > 0$ is the shape parameter, and $\lambda > 0$ is the scale parameter of the distribution. In wind-energy research the Weibull distribution is used to obtain additional flexibility in order to fit an observed wind-speed histogram. A practical advantage over use of the histogram is that the distribution is given by two parameters only, k and λ (Drobinski 2012).

To produce e.g. a wind atlas, the best parameter estimates are obtained using the method of moments, which ensures that the energy content of the fitted Weibull distribution equals the energy content of the observed histogram (Troen and Petersen 1989). The method of moments is a method of estimation of distribution parameters introduced by e.g. Pearson (1894, 1902a, b, 1936), and one begins with deriving equations that relate the moments of the distribution (i.e., the expected values of powers) to the parameters of interest. Then a sample is drawn and the moments are estimated from the sample, with the equations then solved for the parameters, using the sample moments. This results in estimates of those parameters. The method of moments ensures the best estimation of wind-energy potential but does not ensure the maximum likelihood with the observed histograms. This can lead to large errors when considering only a fraction of the wind distribution, between the cut-in and cut-out wind speeds of a specific wind-turbine power curve for instance.

For some design applications including wind loads and structural safety, it is also necessary to have information on the distribution of the complete population of wind speed at a site. Estimation of fatigue damage must account for damage accumulation over a range of extreme winds, the distribution of which is usually fitted with a distribution of the Weibull type (Davenport 1966). In chemistry-transport modelling, the Weibull distribution is used to represent the subgrid-scale variability of the wind speed. This allows improvement of the simulation of aerosol saltation at the surface and emission fluxes into the atmosphere that are triggered by a threshold in the wind speed (Menut 2008). In such contexts, maximizing the likelihood of fitted distributions to observed wind-speed histograms is a major issue.

However, it has been long known that the Weibull distribution is only an approximation and may fit poorly the wind-speed statistics, especially in the case of non-circular (i.e. non-isotropic) or non-normal (i.e. non-Gaussian wind components) distributions (Tuller and Brett 1984). The wide use of the Weibull distribution is purely empirical and there is a lack of physical background justifying the use of the Weibull distribution to model wind statistics. Many previous studies have considered the limitations of the Weibull distribution for modelling wind speeds (e.g. Bauer 1996; Erickson and Taylor 1989; Li and Li 2005; Carta et al. 2009; He et al. 2010; Morrissey and Greene 2012). Bauer (1996), Erickson and Taylor (1989) and He et al. (2010) quantified the deviation of the surface wind-speed distribution from the Weibull distribution from in situ, remotely sensed and modelled wind speeds. In situ and modelled oceanic surface wind speeds from extratropical latitudes are reasonably well simulated by the Weibull distribution (Bauer 1996). About 30–35 % of modelled surface wind-speed frequency distributions are found to be non-Weibull (Erickson and Taylor 1989), and slightly less over the ocean (30–32 %) than over land (30–35 %) with seasonal variations.

Conversely, the remotely sensed wind speeds agree poorly with the corresponding empirical distributions. At a more regional scale, He et al. (2010) showed that measured surface wind-speed frequency distributions in North America are sensitive to the underlying land-surface types, seasonal and diurnal cycles, and the departure from a Weibull distribution is larger at nighttime.

Possible better suited surface wind-speed frequency distributions have been investigated (e.g. Carta et al. 2009, for a review). Some wind-speed distributions were based on the use of the bivariate normal distribution with wind speed and direction as variables (Smith 1971; McWilliams et al. 1979; McWilliams and Sprevak 1980; Colin et al. 1987; Weber 1991, 1997), some on more ad hoc distributions based on the fact that wind speed is defined on the positive real line $[0, +\infty[$, which generally have an exponential-like distribution (Bryukhan and Diab 1993; Li and Li 2005; Morrissey and Greene 2012). The Weibull distribution parameters present a series of advantages with respect to other distributions (e.g. flexibility, dependence on only two parameters, simplicity of the estimation of its parameters) but cannot represent all the wind regimes (e.g. those with high percentages of null wind speeds, bimodal distributions, …). Therefore, its generalized use cannot be justified. Other distributions based on an expansion of orthogonal polynomials (Morrissey and Greene 2012) or the maximum entropy principle (Li and Li 2005) can produce more accurate estimates of the wind-speed distribution than the Weibull function, and can represent a wider range of data types as well. Such distribution functions have been compared to each other (Gamma, Rayleigh, Weibull and Weibull mixture, Beta, log-normal, inverse Gaussian distributions) and the mixture distributions have provided the highest values of coefficient of determination (Carta et al. 2009). In general, the other tested wind-speed distributions displayed lower performance.

In this study, we propose to derive wind-speed distributions based on the use of the bivariate distribution with the two wind components. Indeed, by contrast with the wind-speed modulus, wind components obey the momentum conservation equations (e.g. Salameh et al. 2009), which provide great physical insight for the derivation of the wind-speed distribution. As stated above, several authors proposed the use of the bivariate normal distribution with wind speed and direction as variables. For instance, the isotropic Gaussian model of McWilliams et al. (1979) and McWilliams and Sprevak (1980) was derived from the assumptions that the wind-speed component along the prevailing wind direction is normally distributed with non-zero mean and a given variance, while the wind-speed component along the orthogonal direction is independent and normally distributed with zero mean and the same variance. The anisotropic Gaussian model of Weber (1991, 1997) is a generalization of the model of McWilliams et al. (1979). In Weber (1991), no restrictions are imposed on the standard deviations of the longitudinal and lateral fluctuations. The marginal wind-speed distributions are obtained after integration over the direction variable (Colin et al. 1987; Carta et al. 2008). Here, we go one step further, and use wind records at 89 locations in France to:

– compute alternative distributions analytically with different assumptions on the wind components: (1) normal distributions with different variances to model wind anisotropy; (2) non-Gaussian distributions, from the super-statistics theory, to better model wind extremes; (3) a mixture of normal distributions to model multiple wind regimes.
– perform in-depth verification of these distributions against observations covering various sub-climatic regions in France to identify those distributions that perform optimally, and the reasons for this, such as terrain complexity and dominant weather regimes.
– analyze the sensitivity of the proposed distributions to the diurnal to seasonal variability.

Section 2 presents the data and the methodology, while Sect. 3 discusses the performance of the Weibull distribution. Section 4 presents three alternative wind-speed distributions and how they are derived from the wind components. Finally Sect. 5 presents how these distributions fit the observations, analyzes their sensitivity to diurnal cycle and seasonal variability and summarizes their performances.

2 Observations and Methodology

2.1 Wind Measurements

We use wind measurements at a height of 10 m above the ground from the NOAA ISD-Lite database (Smith et al. 2011). This database has long time records, but we only use measurements made between January 1 2010 and December 31 2013 because the accuracy is better over these recent years: the wind speed M is binned with intervals of 1 knot $(0.514\,\mathrm{m\,s^{-1}})$[1] instead of 2 knots for previous years, and there are less gaps in the records. The wind direction D is binned with intervals of 10°. The 10-min averaged wind speeds and directions are recorded every hour. Among all the stations present in the ISD-Lite database, we select stations in France where the accuracy is 1 knot over the four years and where the data availability is greater than 97 % over each one of the four years, with a minimum of 85 % for each month. We therefore retain 89 stations. At each station we have about 3.4×10^4 measurements so that the extremes should be well described. However, on average, the wind-speed records are correlated over a 10-h time period leaving thus about 3×10^3 independent samples. Calm conditions ($M = 0$) have been removed before all calculations since they are not taken into account in the Weibull distribution nor in the other distributions that we plan to use. The calm conditions represent 5.2 % of the entire dataset with of course disparities among stations. Most are around 3 % but 14 stations are above 10 %. When wind components are used instead of wind speed, the zonal (west–east) and meridional (south–north) wind components (u and v) are derived from the wind speeds and directions. The wind components are often correlated so we use the procedure described by Crutcher and Baer (1962) to reduce component correlation to zero. The change of variable $u \rightarrow u'$, $v \rightarrow v'$, is obtained by a simple rotation of an angle ψ defined by

$$\tan(2\psi) = 2\frac{\rho_{uv}}{\sigma_u^2 - \sigma_v^2},\qquad(2)$$

where ρ_{uv} is the correlation coefficient and σ_u^2 and σ_v^2 are the component variances. The procedure relies on the property that any Gaussian joint probability distribution is characterized by a covariance matrix, which is positive and definite. Therefore, the matrix is fully characterized by its eigenvalues (which we call σ_u and σ_v) and its eigenvectors are orthonormal. In the new base (which has been rotated by an angle ψ calculated over the entire dataset), the matrix is diagonal and has no cross-terms, corresponding to the correlation. We verified at all stations that u' and v' have no correlation. Considering correlation free components for the wind field allows the manipulation of simpler expressions for the distributions. The final results are not affected if fit to the real meridional/zonal components instead of the rotated ones. For the sake of simplicity thereafter u and v notations are used for the rotated wind components u' and v'.

---

[1] The data were recorded in knots and not in m s^{-1}.

2.2 Wind Regimes at the Studied Stations

The topography of France is presented in Fig. 1, with the locations of the 89 weather stations. The stations located in the northern and western parts of France are in rather flat terrain. The stations located in the south-east part of France are in a much more complex environment with elevated mountain ridges: the Pyrénées to the south-west (highest elevation 3404 m), the Alps to the east (highest elevation 4807 m) and the Massif Central in between (highest elevation 1886 m). The Rhône valley separates the Alps from the Massif Central by a gap 200 km long and 60 km wide, and the Aude valley separates the Pyrénées from the Massif Central.

In the northern and western regions, all wind directions are experienced, even though this part of France is located in the storm track so that strong winds are often from the west, from the Atlantic Ocean (Vautard 1990; Plaut and Vautard 1994; Simonnet and Plaut 2001). Indeed, over the north-west Atlantic cyclones originate, travel eastwards and affect the European continent. In the southern region, frequent channelled flow can persist for several days. The strongest and most frequent channelled valley flow is the mistral, which derives from the north/north-west in the Rhône valley (Drobinski et al. 2005), and occurring when a synoptic northerly flow impinges on the Alpine range. As the flow experiences channeling, it is substantially accelerated and can extend offshore over horizontal ranges exceeding few hundreds of kilometres (Salameh et al. 2007; Lebeaupin Brossier and Drobinski 2009). The mistral occurs all year long but exhibits a small seasonal variability either in speed and direction, or in its spatial distribution (Orieux and Pouget 1984; Guénard et al. 2005, 2006).
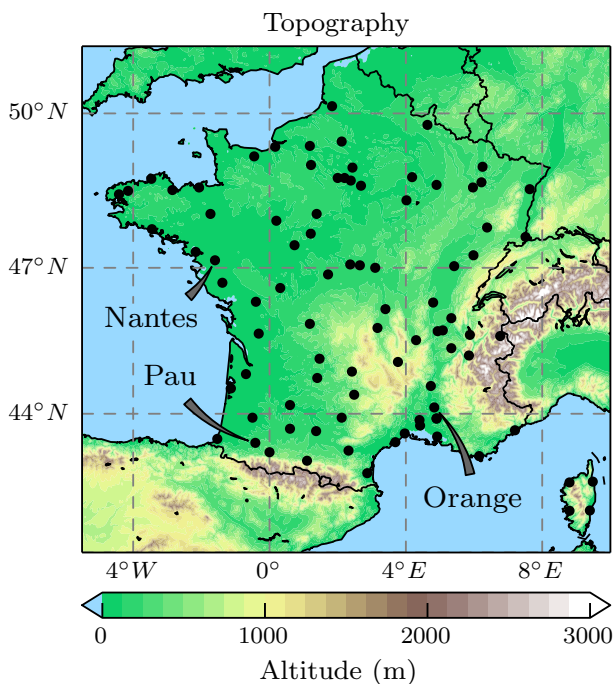

Fig. 1 Map of France's topography with the locations of the 89 weather stations used in this study, including the three stations used as examples: Nantes, Pau and Orange

The mistral shares its occurrence with a northerly land breeze and southerly sea breeze (e.g. Bastin and Drobinski 2005, 2006; Drobinski et al. 2006, 2007), which can also be channelled in the nearby valleys (Bastin et al. 2005b) or interact with the mistral (Bastin et al. 2005a, 2006). In such a region, accounting for such persistent wind systems for modelling the wind-speed statistics is thus mandatory.

We will often refer to three stations as examples among the 89 stations: Nantes is in flat terrain, such as most stations in north-western France; Pau is in a more complex topographic region, close to the Pyrénées mountains, and Orange is in the Rhône valley and influenced by the mistral channelled flow.

2.3 Goodness-of-fit of the Distributions

We introduce several wind-speed distributions and analyze how they fit the observational data. The comparison is made on the cumulative density function (CDF) for wind speed, and we compute a distance score between the CDF of the tested distribution (F) and the observed empirical CDF ($\hat{F}_n$). It must be noted that the Weibull distribution leads to an analytic expression for the CDF, but not so for the other distributions that will be derived hereafter. Therefore the CDF (F) are numerically computed from their probability density function (PDF) expressions. To be consistent among distributions, even the Weibull CDF is numerically computed from its PDF.

The empirical CDF ($\hat{F}_n$) for a set of observations ($x_1, \ldots, x_n$), of length n is defined by

$$\hat{F}_n(t) = \frac{\text{number of elements} \leq t}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{x_i \leq t\}. \tag{3}$$

Because of the coarse resolution of the observations (1 knot), this empirical CDF is a step function so we add a small random noise to the wind-speed data in order to smooth the CDF. The added noise is a continuous uniform distribution between -0.5 knot and $+0.5$ knot, and ensures a continuous empirical CDF. Figure 2 shows examples of the CDFs with or without smoothing. Adding this noise does not affect the fitting of the distributions and it is not necessary to compute the distance scores. But the steps in the CDF lead to overestimation of the distances between $\hat{F}_n$ and F, especially at the stations with low mean wind speeds, such as at Pau, where we see large steps. So smoothing the empirical CDF enables us to better quantify the performances of the distributions and to make comparison between different stations easier.

The goodness-of-fit scores used herein are Cramer–von Mises and modified Anderson–Darling statistics. A general equation for these scores is

$$\Delta_n^2 = n \int_{-\infty}^{\infty} \left(F(x) - \hat{F}_n(x) \right)^2 \omega(x)\, \mathrm{d}F(x), \tag{4}$$

where $\omega(x)$ is a weighting function. Various expressions of $\omega(x)$ are given in Table 1, and details for numerical computations of the solution to this equation are given in Appendix 1.

In the case of the Cramer–von Mises (CvM) statistic (W_n^2), the weight is constant ($\omega(x) = 1$) so that the centre of the distribution actually dominates the equation. Here, the centre of the distribution is not one single point but the region around the median, mean or maximum of the distribution (where the PDF is above, say, 0.1). The Anderson–Darling score (A_n^2) puts weight on the tails of the distributions, where the tail corresponds to the part of the distribution that exceeds the 90th centile. In order to analyze the upper tail corresponding to strong winds, we can use the modified right-tail Anderson–Darling (ADR) statistic (R_n^2)
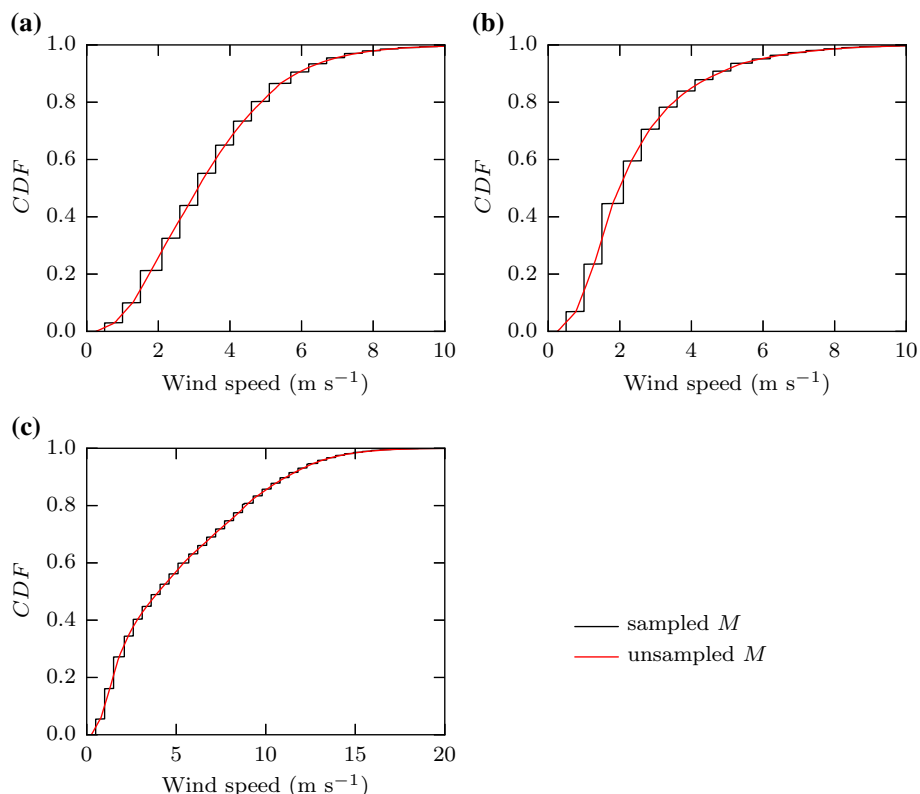
Fig. 2 Comparison of the discrete "step-function" empirical CDF (*black*) and its continuous version after adding a small random noise to the wind-speed data (*red*), at three stations: **a** Nantes, **b** Pau, **c** Orange

Table 1 Goodness-of-fit statistics

Name	Δ_n^2	$\omega(x)$
Cramer–von Mises (CvM)	W_n^2	1
Anderson–Darling (AD)	A_n^2	$[F(x)(1 - F(x))]^{-1}$
Right-tail AD (ADR)	R_n^2	$[1 - F(x)]^{-1}$
Right-tail AD of second degree (AD2R)	r_n^2	$[1 - F(x)]^{-2}$

(Sinclair et al. 1990) or, for even greater weight placed on the tail, the modified right-tail ADR of second degree (AD2R) statistic (r_n^2) as defined by Luceño (2006).

We will use the W_n^2 and r_n^2 scores to assess the goodness-of-fit on the centre and tail of the distributions. As in any goodness-of-fit test, there are thresholds for rejection of the null hypothesis (i.e. the hypothesis that the observed data are drawn from F) for different significance levels. But these thresholds depend on the distribution F, on the autocorrelation in the data and could only be estimated by simulations (e.g. Ahmad et al. 1988). Therefore the question of limit values for the metrics is very complex and beyond the scope of the present work. Rather than defining thresholds we will simply compare the scores of different fits at each station.
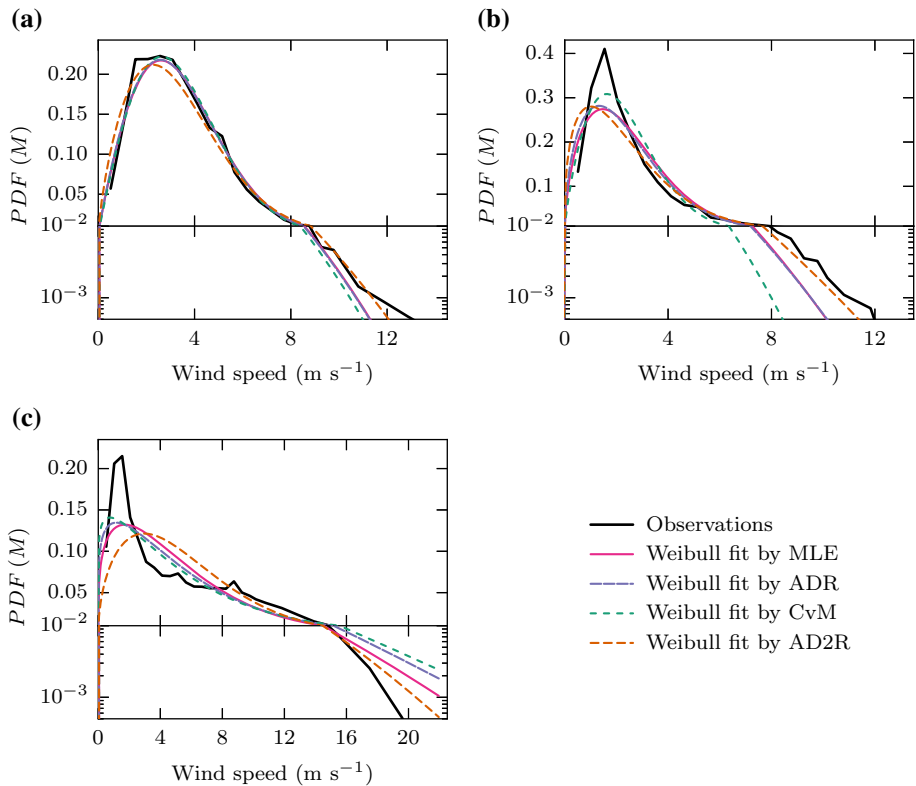
Fig. 3 Wind-speed distributions at three stations: **a** Nantes, **b** Pau, **c** Orange. *Black* observed distributions. *Colours* Weibull distributions fitted by four different methods: MLE (*pink*) or minimizing ADR (*blue*), CvM (*green*) or AD2R statistics (*orange*). The *y*-axis is divided into a linear axis and a logarithmic axis (to better resolve the tail)

3 Performance of Weibull Distribution

As a reference before introducing other distributions, we consider the Weibull distribution for the wind speed M (Eq. 1). First we benchmark four different methods for fitting the Weibull distribution because there are several possible fitting procedures. Popular methods include the method of moments, often used in a wind atlas, or by the usual maximum likelihood estimate (MLE). The principle of the MLE, originally developed by Fisher (1912, 1922), states that the desired probability distribution is the one that makes the observed data "most likely", which means that one must seek the value of the distribution parameters that maximize the likelihood. We compare the MLE method and three minimization algorithms based on CvM, ADR or AD2R statistics. The ADR minimization means that the algorithm finds the best parameters to minimize the R_n^2 score measuring the distance between the empirical CDF and the fitted CDF. Examples of the Weibull fits from the four fitting methods are given in Fig. 3 at three stations. While good fits have a similar shape regardless of the fitting methods in Nantes (Fig. 3a), spurious fits in Pau or Orange show a significant spread (Fig. 3b, c). CvM and AD2R represent bounds for the fits, favouring on the one hand the centre of the distribution (CvM), and on the other hand the tail (AD2R). ADR minimization is a good

compromise favouring the tail but not overly so, and also yields results similar to the MLE. For these reasons we adopt the ADR minimization for all fits throughout the study.

In the flat terrain of northern France, as at Nantes (Fig. 3a), the Weibull distribution describes well the centre of the distribution but it tends to underestimate the tails of the distributions. In wind energy, the tail of the wind distribution is not important for estimating the wind resource but it is of high importance when addressing wind loading and damage fatigue, pollutant transport or the impact of wind storms. In more complex terrain, the fit to the Weibull distribution is less accurate. For example, at Pau in the Pyrénées, the distribution is more peaked than for the Weibull approximation (Fig. 3b), and is worse in the southern valleys such as at Orange (Fig. 3c). The wind-speed distribution exhibits a peak at low wind speeds (about 2 m s^{-1}) followed by a "shoulder", with a more concave shape between 6 and 12 m s^{-1} due to the channelled flow. Figure 3c shows that the Weibull distribution cannot fit such a complex shape, and we estimate that, in these particular cases, using a Weibull distribution leads to errors exceeding 10 % on occasions regarding the wind energy.

For a more quantitative analysis, we use the Cramer–von Mises (W_n^2) and AD2R scores (r_n^2) to quantify the distance between the empirical and the predicted CDFs. The quantities W_n^2 and r_n^2 measure the distance between the empirical CDF ($\hat{F}_n$) and the fitted distribution's CDF, with a weight on the values in the centre or at the right tail of the distribution, respectively. For example at Nantes, Pau and Orange, the Cramer–von Mises (W_n^2) scores for the Weibull distribution are 2, 44 and 24 respectively, and the AD2R scores (r_n^2) are 840, 1560 and 1460. We need to be careful when comparing stations, for example we see a lower W_n^2 score at Orange than at Pau whereas the Weibull fit appears poorer. Indeed the score values depend on the function $\hat{F}_n$, different at each station and dependent on the number of observations n. So we cannot compare one station to another but we can compare the scores of several fits at a unique station (see Sect. 5). Nevertheless, it is important to provide an estimate of the fit quality. Based on our observations of all fits, we consider that a Cramer–von Mises score (W_n^2) < 2 indicates a good fit for the centre of the distribution, and an AD2R score (r_n^2) < 100 indicates a good fit for the tail of the distribution. This is consistent with that we observed at the three stations where only the fit at the centre of the distribution at Nantes is excellent.

Figure 4 shows a map of the W_n^2 and r_n^2 scores for the Weibull distribution. It mainly shows that the Weibull distribution is suited to model the centre of the wind-speed distribution in northern France where the W_n^2 score <2. In southern France, the measured wind-speed histograms deviate significantly from the Weibull distribution. As expected, this figure also shows the low performance of the Weibull distribution for the highest wind-speed values, with r_n^2 score exceeding 100 nearly everywhere. Note that the systematic deviation from the Weibull distribution in the southern region might be related to the complex topography. In the following, we investigate alternative distributions, where topography-induced effects such as wind anisotropy and the existence of persistent wind regimes are captured.

4 Alternative Wind-Speed Statistics Models

For a deeper insight into the differences between observed wind-speed distributions and the commonly used Weibull distribution, we now consider bivariate distributions of the two wind components to take into account the wind-field anisotropy. At many stations in the southern regions, where the Weibull distribution does not model the wind statistics well, the wind field is very anisotropic. Indeed, we see in Fig. 5 that at Pau and Orange the wind components u and v have very different statistics, whereas at Nantes u and v have not identical but close PDFs. To evaluate the wind anisotropy at all stations we use the variance ratio σ_u^2/σ_v^2 (or the
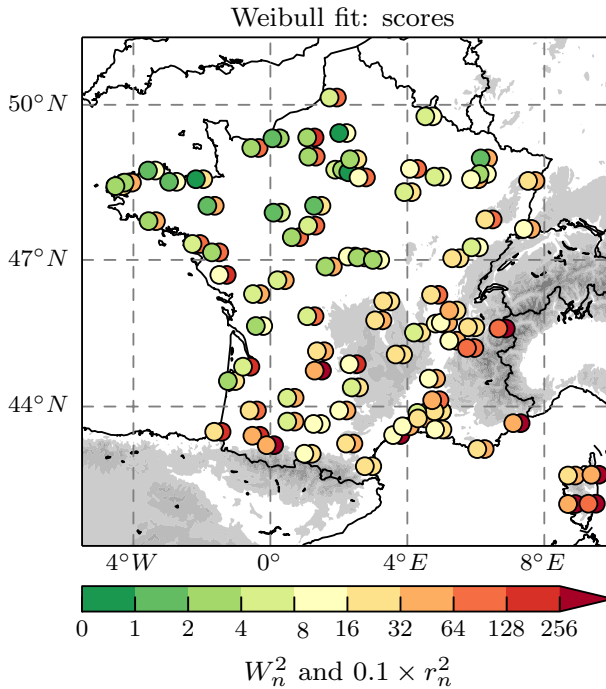
Fig. 4 Weibull distribution fit by minimizing the ADR score. At each station, the foreground circle gives the W_n^2 score (value <2 for a good fit at the centre of the distribution). The *background circle* gives the r_n^2 score (value < 100 for a good fit at the tail of the distribution) multiplied by 0.1 to have a common colour axis

inverse ratio in the case of $\sigma_v > \sigma_u$). Figure 6a shows this ratio, and we see that the wind field is nearly isotropic in the north-western region (i.e. variance ratio $\approx$1) but becomes very anisotropic in the south-eastern region where the flow is channelled by the mountains and the variance ratio can exceed 3.

4.1 Elliptical Distribution

The very first approach to model anisotropy is to consider a bivariate distribution of the Gaussian wind components u and v with variances σ_u^2 and σ_v^2. We recall that u and v are not correlated, and in the case of zero means, $\mu_u = \mu_v = 0$, the joint PDF is

$$p(u, v; \sigma_u^2, \sigma_v^2) = \frac{1}{2\pi \sigma_u \sigma_v} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right). \tag{5}$$

Applying the usual transformation from Cartesian to polar coordinates (M, ϕ), and integrating over the angle ϕ (Chew and Boyce 1962), the joint PDF for the wind speed M is

$$P_{ELL}(M; \sigma_u^2, \sigma_v^2) = \frac{M}{\sigma_u \sigma_v} \exp\left(-a\, M^2\right)\, I_0\left(b\, M^2\right). \tag{6}$$

with $a = (\sigma_u^2 + \sigma_v^2)/(2\sigma_u\sigma_v)^2$, $b = (\sigma_u^2 - \sigma_v^2)/(2\sigma_u\sigma_v)^2$, and $I_0(x)$ is the modified Bessel function of the first kind and zero order. This particular bivariate normal distribution will be called elliptical hereafter.
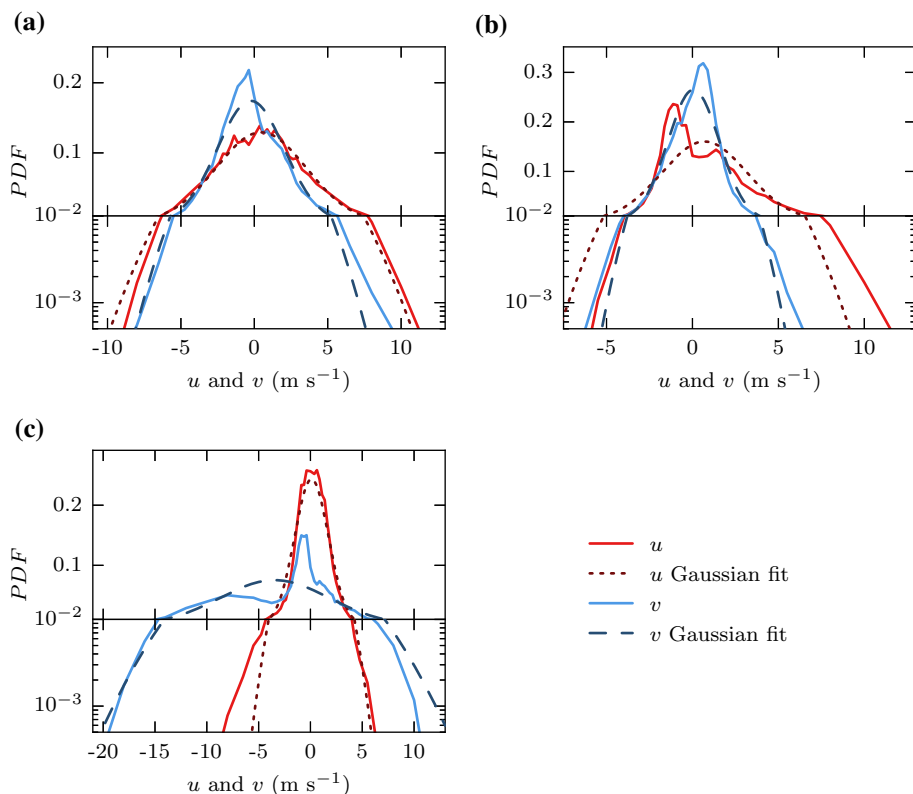
Fig. 5 Probability density functions for the wind components at three stations: **a** Nantes, **b** Pau and **c** Orange. *Plain curves* observed distributions for *u* (*red*) and *v* (*blue*). *Dashed curves* their fits using a Gaussian distribution

4.2 Non-Gaussian Distribution

With the previous elliptical distribution, we assumed the wind components to follow a Gaussian distribution. However, this assumption is not always valid, since Gaussian curves sometimes fail to describe the histograms—see Fig. 5. We can evaluate the departure of each component u and v from a Gaussian shape by computing the ADR score for the two components, i.e. $A_n^2(u)$ and $A_n^2(v)$; Fig. 6b shows the sum $A_n^2(u) + A_n^2(v)$. Theoretically, the strict Gaussianity is reached when the sum equals zero, however, from visual inspection, a value around 20 can still be considered as reasonably Gaussian. In the flat terrain of north-western France, the scores are not too high, indicating a good fit to a Gaussian. This is however not the case in southern and eastern France. In the following, we use super-statistics defined by Beck and Cohen (2003) to address such a deviation from Gaussianity. This approach consists in representing the long-term stationary state by a superposition of different states that are weighted with a certain probability density.

For the sake of brevity, we focus on one component, u, throughout the following. The large tails of the wind-component distributions originate from the transient nature of the wind field. Meteorological conditions indeed change on a range of time scales (day, anticyclonic duration, season). For instance, the strongest winds are typically recorded in winter, whereas
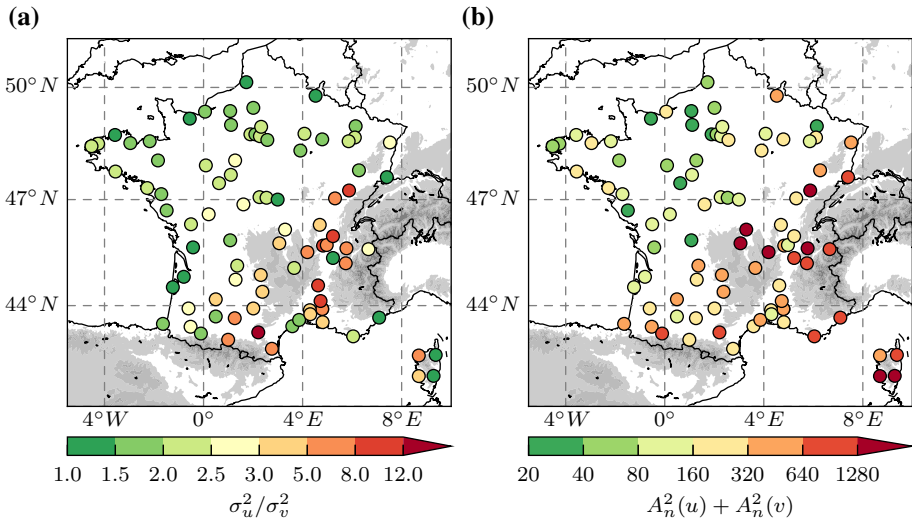
(a) **(b)**

Fig. 6 **a** Measurement of the wind anisotropy at each station: ratio of the wind-component variances σ_u^2/σ_v^2 (or the inverse ratio in the case $\sigma_v > \sigma_u$). **b** Measurement of departure from a Gaussian shape at each station: sum of the Anderson–Darling statistics for a Gaussian distribution of each wind component: $A_n^2(u) + A_n^2(v)$

in summer, long-lasting anticyclonic conditions relate to lower wind speeds near the surface. This induces a change in the statistical properties (e.g. wind-component variance) at several time scales. Here, we model this by assuming a Gaussian shape to the wind-component distribution, but with a fluctuating standard deviation σ. The super-statistics of the wind component u can be derived as follows (see similar methods in Beck and Cohen 2003; Rizzo and Rapisarda 2005),

$$p(u) = \int f(\beta)\, p(u; \beta)\, \mathrm{d}\beta, \tag{7}$$

where f is the probability distribution of the fluctuating variable $\beta = 1/(2\sigma^2)$, and $p(u; \beta)$ is the probability distribution of the wind component u, depending on β. Assuming a Gaussian shape for the wind component at short time scales gives,

$$p(u; \beta) = \sqrt{\frac{\beta}{\pi}} \exp\left(-\beta u^2\right). \tag{8}$$

We observe that the distribution of β is often consistent with a Gamma distribution (see Fig. 7), which gives,

$$f(\beta) = \frac{1}{b\, \Gamma(c)} \left(\frac{\beta}{b}\right)^{c-1} \exp\left(-\frac{\beta}{b}\right), \tag{9}$$

where Γ is the Gamma function, c is the shape parameter, and b is the scale parameter of the distribution. Combining Eqs. 8 and 9 into Eq. 7 (see "Distribution of Wind Components" section in Appendix 2 for computational details), we obtain,

$$p(u) = \sqrt{\frac{b}{\pi}} \frac{\Gamma(c + \frac{1}{2})}{\Gamma(c)} \left(1 + b\, u^2\right)^{-(c + \frac{1}{2})}. \tag{10}$$

This distribution, a generalized Boltzmann factor, has been obtained when attempting to generalize the entropy definition (Tsallis 1988; Beck and Cohen 2003). Interestingly, it turns
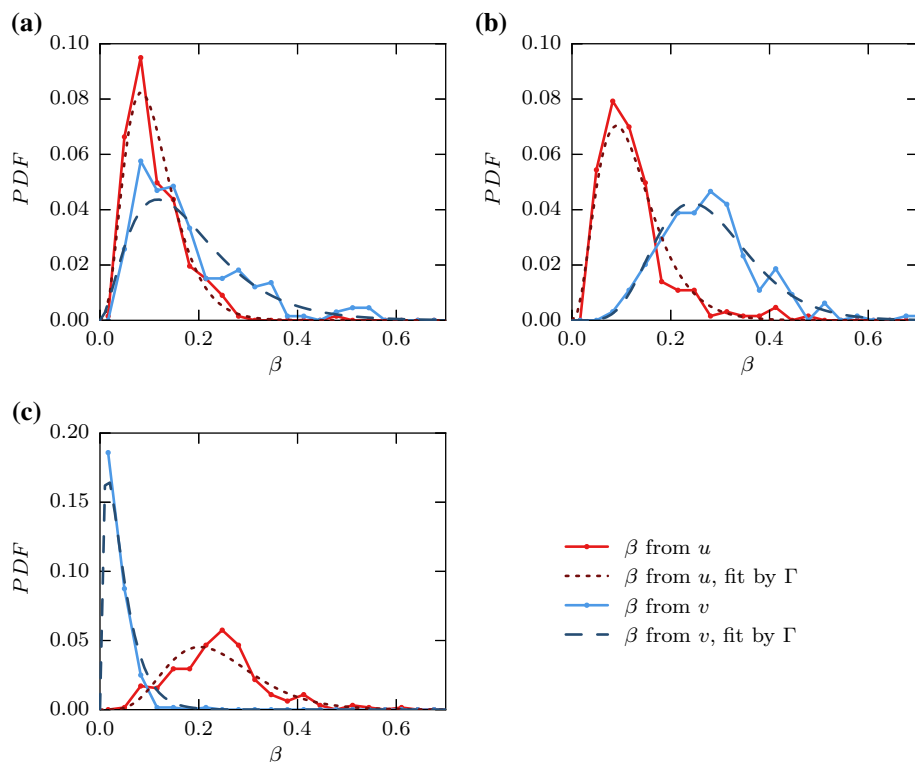
Fig. 7 PDF of $\beta = [2\sigma^2]^{-1}$ where σ^2 is the variance of u (*red*) or v (*blue*) wind component, divided into time intervals of 1 week. *Dashed curves* fit of β by a Gamma distribution. At three stations: **a** Nantes, **b** Pau and **c** Orange

out that the distribution in Eq. 10 is also the stationary solution of a first-order stochastic differential equation with multiplicative noise and additive noise terms. These noise terms, whose strengths are related to parameters b and c, can be physically interpreted as an interplay between turbulence, chaotic atmospheric variability and the mean wind speed (Sura and Gille 2003; Bernardin et al. 2009).

We now turn to the wind-speed distribution, which is at stake in this study. We assume that the components u and v are statistically independent (we recall that there is no correlation between u and v). For the sake of simplicity, and in order to compute the joint distribution, we also assume that u and v have similar statistics, i.e. both components are described by the same b and c. This is a debatable assumption, which is not generically true, but allows considerable simplifications for the derivation of the analytic distribution. Therefore the wind-speed distribution is obtained by computing the joint distribution in radial coordinates and integrating over the wind direction (see "Distribution of Wind Speed" section in Appendix 2 for computational details). We obtain the wind-speed PDF for this non-Gaussian (NG) distribution,

$$P_{NG}(M; b, c) = 2b \, \frac{\Gamma^2(c + \frac{1}{2})}{\Gamma^2(c)} \, M \, \left(1 + b \, M^2\right)^{-(c+\frac{1}{2})}$$
$$\times F\left(c + \frac{1}{2}, \frac{1}{2}; 1; -\frac{b^2 \, M^4}{4(1 + b \, M^2)}\right) \tag{11}$$

where F is the ordinary hypergeometric function. To our knowledge, such a wind-speed distribution has never been proposed in the literature.

4.3 The Rayleigh–Rice Distribution

Another simpler way of using super-statistics is to superpose only two different local dynamics. For one wind component, u (idem for v),

$$p(u) = \int \alpha_u(\mu) p(u, \mu) \mathrm{d}\mu, \tag{12}$$

where $p(u, \mu)$ is the probability distribution of the wind component u for a mean wind speed μ and $\alpha_u(\mu)$ is a weighting function depending on μ. In the case of the two regimes scenario, $\alpha_u(\mu)$ is bimodal,

$$\alpha_u(\mu) = (1 - \alpha)\, \delta(\mu) + \alpha \delta(\mu - \mu_u), \tag{13}$$

where δ is the Dirac function. Such a method has been used to include the contribution of zero wind speed in wind statistics modelling (Takle and Brown 1978; Tuller and Brett 1984).

We introduce here the Rayleigh–Rice distribution, based on the observations that, in the valleys in southern France, there are two wind regimes, both of which can be described by a particular bivariate normal distribution:

(i) random flow: the wind components have zero means and similar variances. This wind-speed statistic is well described by a Rayleigh distribution: equal variances $\sigma_u^2 = \sigma_v^2 = \sigma_1^2$ and zero means $\mu_u = \mu_v = 0$. The Rayleigh distribution is a particular case of a Weibull distribution with shape parameter $k = 2$, and is a particular case of the previously introduced elliptical distribution (Eq. 6 with equal variances: $b = 0$ so $I_0(b M^2) = 1$),

$$P_{\text{Rayleigh}}(M; \sigma_1^2) = \frac{M}{\sigma_1^2} \exp\left(-\frac{M^2}{2\sigma_1^2}\right) \tag{14}$$

(ii) channelled flow: the wind components have different means. The Rice distribution describes well this wind-speed statistic: equal variances $\sigma_u^2 = \sigma_v^2 = \sigma_2^2$ and non zero means $\mu_u \neq \mu_v$,

$$P_{\text{Rice}}(M; \mu, \sigma_2^2) = \frac{M}{\sigma_2^2} \exp\left(-\frac{M^2 + \mu^2}{2\sigma_2^2}\right) I_0\left(\frac{M\mu}{\sigma_2^2}\right), \tag{15}$$

where $\mu = \sqrt{\mu_u^2 + \mu_v^2}$ and I_0 is the modified Bessel function of the first kind of order zero.

The resulting distribution is a sum of the distributions (i) and (ii) conditioned to the absence and presence of the channelled-flow occurrence, α is here the weight corresponding to the occurrence of channelled-flow events. We obtain the Rayleigh–Rice distribution for the wind speed,

$$P_{RR}(M; \alpha, \sigma_1^2, \mu, \sigma_2^2) = \alpha \frac{M}{\sigma_2^2} \exp\left(-\frac{M^2 + \mu^2}{2\sigma_2^2}\right) I_0\left(\frac{M\mu}{\sigma_2^2}\right)$$
$$+ (1 - \alpha) \frac{M}{\sigma_1^2} \exp\left(-\frac{M^2}{2\sigma_1^2}\right). \tag{16}$$

This model can be applied also to the combination of a weak isotropic wind regime and a sustained prevailing flow, as is the case in northern France with prevailing strong westerlies. It can also be seen as an extension of the model proposed by McWilliams et al. (1979) and McWilliams and Sprevak (1980), as it allows consideration of two types of flow regimes with different probabilities of occurrence.

5 Performance of the Alternative Distributions

The three distributions introduced in the previous section are fitted to the observations, using a minimization algorithm on the ADR statistic, such as decided with the Weibull distribution in Sect. 3. This means that the two or four parameters of each distribution are adjusted in order to minimize the R_n^2 distance between the CDF and the observed empirical distribution. For the elliptical distribution, the fit determines the best parameters σ_u^2 and σ_v^2 of Eq. 6. For the non-Gaussian distribution the PDF is more complex due to the hypergeometric term in Eq. 11, making it more difficult to fit than a Weibull distribution. The Rayleigh–Rice distribution for wind speed in Eq. 16 is more difficult to fit because it has four parameters, especially because of the non-linear effect of the α parameter that modulates the respective weights of the Rayleigh and Rice distributions. To overcome this difficulty, we first fit the distribution for only three parameters and a fixed value of α, repeat this for a series of different α values, and choose the best of all fits. This best fit is then used as a first estimate to fit with four parameters and it rapidly converges.

Now we discuss the performances of the three distributions, first at stations Nantes, Pau and Orange and afterwards in a more systematic fashion at all 89 stations.

5.1 Examples of Fits at Three Stations

Figure 8 shows the fits for the four distributions at three example stations, and additionally we give the values of the W_n^2 and r_n^2 scores for these fits in Table 2.

At Nantes (Fig. 8a), in a flat area, the four distributions give very good results, all very close except on the tail where they differ a little. For the centre of the distribution indeed all distributions have $W_n^2 < 2$. We can see that the non-Gaussian and Rayleigh–Rice distributions give better fits on the tail, in accordance with r_n^2 scores < 100 in Table 2. In contrast, the Weibull and elliptical distributions that underestimate the tail have higher r_n^2 scores. At Pau (Fig. 8b), the wind distribution is very peaked and this peak is largely missed by the Weibull. We see that the other distributions are closer to the actual peak with, in order, elliptical, non-Gaussian and Rayleigh–Rice, but never perfectly modelling it. This is seen in Table 2 with $W_n^2 = 44$ (Weibull) versus 22 (elliptical), 12 (non-Gaussian) and 7 (Rayleigh–Rice), but still above 2. On the tail the Rayleigh–Rice is the only distribution that fits well to the observations, and the only one with $r_n^2 < 100$. At Orange (Fig. 8c), we have a shouldered histogram where the Weibull is not well-suited; neither are the elliptical and non-Gaussian distributions. Only the Rayleigh–Rice distribution is capable of fitting the two peaks. It is the only distribution with low scores at Orange in Table 2: 1 for W_n^2 and 60 for r_n^2.

5.2 Diurnal and Seasonal Variability

Several studies have pointed out that the deviation of the Weibull distribution from the observed wind-speed distribution displays diurnal to seasonal variations (e.g. Erickson and Taylor 1989; He et al. 2010). Table 3 gives the values of the W_n^2 and r_n^2 scores for the fits of
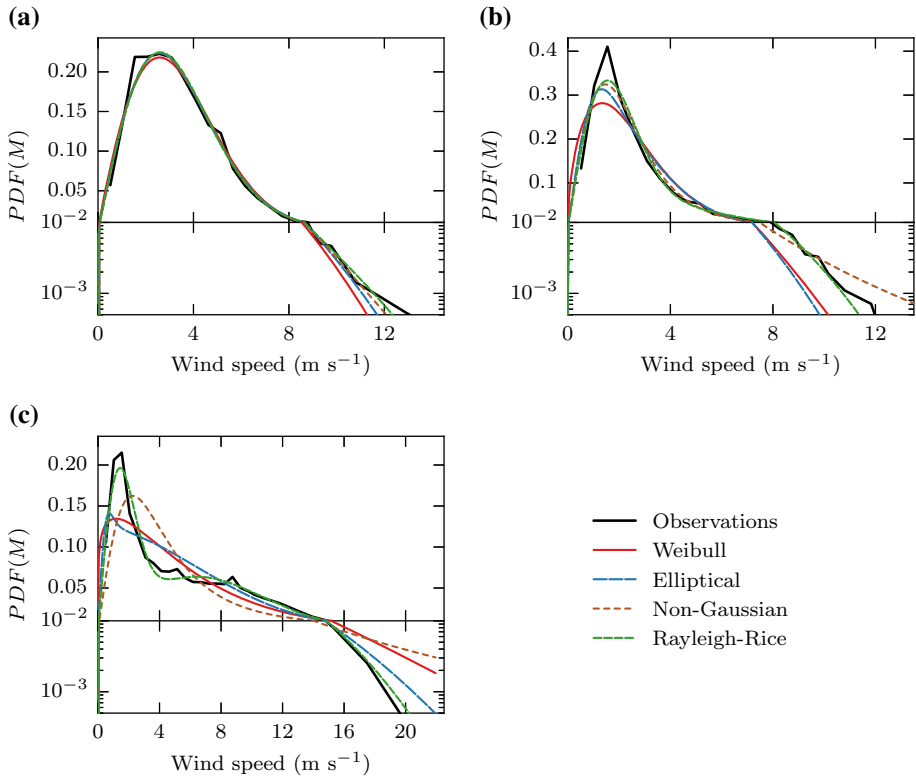
Fig. 8 Wind-speed distributions at three stations: **a** Nantes, **b** Pau, **c** Orange. *Black* observed distributions. *Colours* fit for Weibull (*red*), elliptical (*blue*), Rayleigh–Rice (*green*) and non-Gaussian (*brown*) distributions

Table 2 Goodness-of-fit scores of the distributions in Fig. 8. The quantities W_n^2 (Cramer–von Mises) and r_n^2 (right-tail ADR of second degree) measure the distance between the empirical and fitted distributions with focus on the centre and right tail of the distribution, respectively. A lower value indicates a better fit

	Nantes		Pau		Orange	
	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2
Weibull	2.0	839	43.9	1563	24.2	1458
Elliptical	1.0	392	21.9	4067	43.6	486
Non-Gaussian	0.9	82	11.5	292	91.0	3076
Rayleigh–Rice	0.9	57	7.4	89	1.1	61

the Weibull, elliptical, non-Gaussian and Rayleigh–Rice distributions at the three stations, Nantes, Pau and Orange, and at night (0000 UTC) and day (1200 UTC). Consistently with He et al. (2010), the two-parameter Weibull distribution fits better daytime than nighttime wind-speed distributions at the three stations for both the centre and tail of the distribution. Quantitatively, the number of data in each fit is divided by 24 so the scores are much lower than previously. In practice, they should be multiplied by 24 to be comparable to the scores of

Table 3 Same as Table 2 for daytime (observations only at 1200 UTC) and nighttime (observations only at 0000 UTC)

| | NIGHT (0000 UTC) | | | | | | DAY (1200 UTC) | | | | | |
| | Nantes | | Pau | | Orange | | Nantes | | Pau | | Orange | |
	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2
Weibull	0.4	156	1.9	320	2.0	75	0.1	60	1.5	47	0.8	62
Elliptical	0.2	171	1.1	713	5.8	31	2.0	29	0.7	62	0.8	35
Non-Gaussian	0.1	16	0.9	15	3.0	156	2.0	30	0.5	12	2.8	119
Rayleigh–Rice	0.1	49	0.5	3	0.1	25	0.1	3	0.3	11	0.1	10

The number of data in each fit is divided by 24 so the scores are much lower than previously

Table 4 Same as Table 2 for extended winter (October–March) and summer (April–September)

| | Winter | | | | | | Summer | | | | | |
| | Nantes | | Pau | | Orange | | Nantes | | Pau | | Orange | |
	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2	W_n^2	r_n^2
Weibull	1.0	974	30.7	1508	20.9	953	1.5	49	15.7	251	7.3	604
Elliptical	0.5	498	15.9	6234	40.6	355	0.9	16	6.9	344	8.9	270
Non-Gaussian	0.5	132	8.5	144	64.1	1854	0.9	16	6.8	180	32.4	1272
Rayleigh–Rice	0.2	25	1.6	141	0.4	29	0.9	79	3.2	21	0.7	94

Table 2. So for a "perfect fit" the values of W_n^2 and r_n^2 scores should be lower than about 0.1 and 4, respectively. As also shown by He et al. (2010), Table 3 shows that the daytime scores for the Weibull fit are systematically lower than those for nighttime, with values of 0.1 against 0.4 at Nantes, 1.5 against 1.9 at Pau and 0.8 against 2.0 at Orange; this behaviour occurs for all stations (not shown). The higher ability of the other distributions to fit the observations during daytime remains for the Rayleigh–Rice distribution but it is not systematically the case for the elliptical and non-Gaussian distributions. Table 3 shows that at Nantes, W_n^2 and r_n^2 scores for the elliptical and non-Gaussian distributions are 2.0 during daytime against 0.2 and 0.1 at night, respectively. Indeed, the main peak is narrower at night than during the daytime (lower near-surface wind speed) whereas the tail is not significantly different. This creates a longer tail at night that cannot be captured by the elliptical and non-Gaussian distributions.

The impact of the seasonal cycle on the wind-speed statistics has also been investigated. Erickson and Taylor (1989) showed that overland 35 % of the wind-speed distributions are judged to be non-Weibull in January versus 30 % in July. Table 4 gives the values of the W_n^2 and r_n^2 scores for the fits of the Weibull, elliptical, non-Gaussian and Rayleigh–Rice distributions at the three stations, Nantes, Pau and Orange in winter (October–March) and summer (April to September) (the scores should be multiplied by 2 to be comparable to the scores of Table 2. For "perfect fit" the values of W_n^2 and r_n^2 scores should be lower than about 1 and 50, respectively). Table 4 shows a less clear behaviour. At Nantes, all distribution fits give larger W_n^2 score in summer than in winter suggesting a better fit of the centre of the distribution in summer. It is however the reverse for the tail of the distribution, except

for the Rayleigh–Rice distribution. At the other stations, the Weibull distribution, as well as the elliptical and non-Gaussian distributions, better fit the observations in summer for both the centre and the tail. The Rayleigh–Rice distribution displays in general the opposite behaviour, generally performing better during winter. This can easily be explained by the higher probability of persistent strong winds over France, which produce the secondary peak or shoulder at higher wind speeds, enabling a more accurate and reliable fit of the Rayleigh–Rice distribution. However, in any case, in absolute value the Rayleigh–Rice distribution generally outperforms the other distributions.

5.3 Systematic Quantification of Performances

We now generalize the findings from the three example stations and make a systematic comparison of the performances of each distribution against the Weibull. At each station, we compute the W_n^2 and r_n^2 scores of each distribution, such as we did for the Weibull (Fig. 4). Then the comparison of the W_n^2 (respectively r_n^2) scores indicates which distribution performs best on the centre (respectively the tail) of the distribution. We consider that two distributions are similar when the difference in W_n^2 scores is <2 (<100 for the r_n^2 scores).

The performances of the elliptical distribution are summarized in Fig. 9. The W_n^2 and r_n^2 scores at each station are given in Fig. 9a and their comparisons to the Weibull scores are given in Fig. 9b. In the north-western region the fits are in general good and similar to those of the Weibull (white dots in Fig. 9b). Elsewhere the elliptical distribution is in general better at describing the centre of the distribution (blue foreground dots in Fig. 9b). But even if the distribution performs better than the Weibull, Fig. 9a shows that the fits are
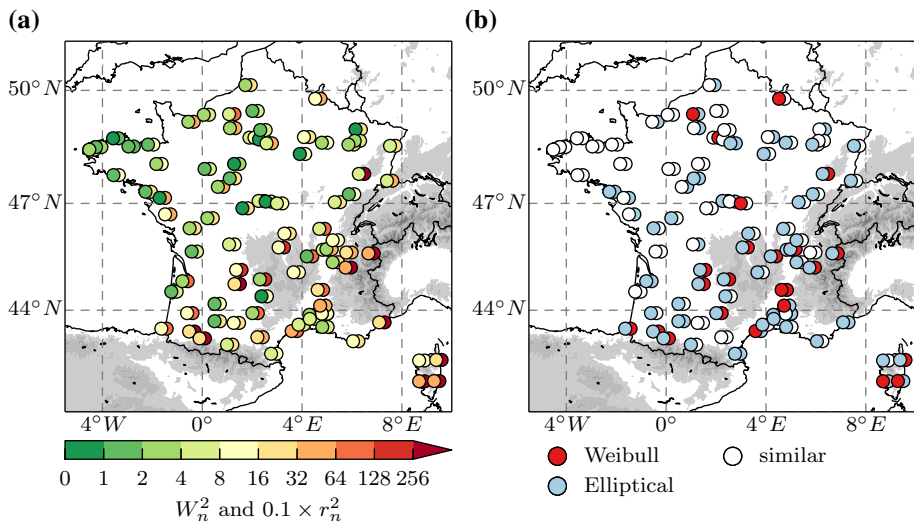

Fig. 9 **a** Elliptical distribution fit by minimizing the ADR score. At each station, the foreground circle gives the W_n^2 score (value <2 for a good fit on the centre of the distribution). The *background circle* gives the r_n^2 score (value <100 for a good fit on the tail of the distribution) multiplied by 0.1 to have a common colour axis. **b** Best distribution between Weibull (*red*) and elliptical (*blue*). At each station, the *foreground circle* gives the best distribution for the centre of the distribution according to W_n^2 scores. The *background circle* gives the best distribution for the tail of the distribution according to r_n^2 scores. *White dots* corresponds to stations where the scores are equal or almost, i.e. W_n^2 (respectively r_n^2) scores differing by less than 2 (respectively 100)
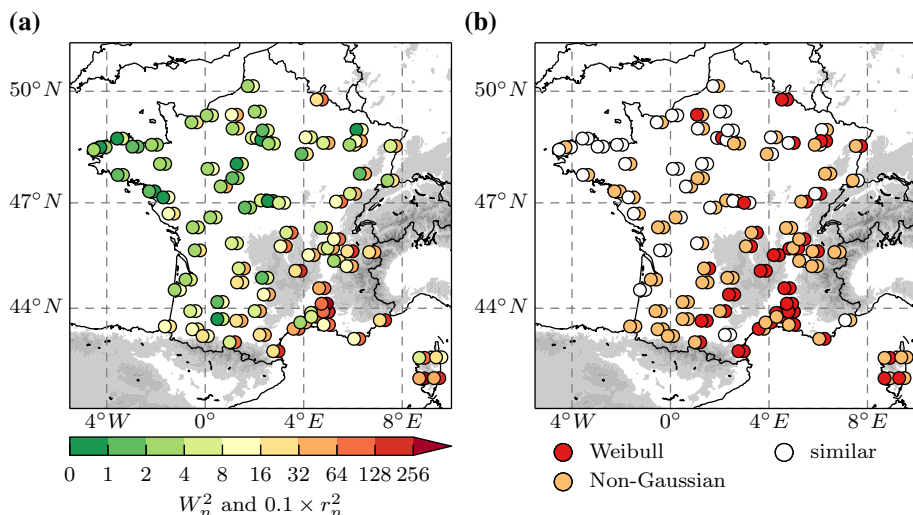
Fig. 10 Same as Fig. 9 for the non-Gaussian distribution (in *orange*) compared to the Weibull distribution (in *red*)

not very good: we still have high W_n^2 values in the southern region. Concerning the tail of the distribution (background dots), we can see that the elliptical distribution is often not better than the Weibull. This is due to the Gaussian assumption and the reason why we introduced the non-Gaussian distribution.

The performances of the non-Gaussian distribution are summarized in Fig. 10. In general in the north-western region this distribution gives similar results as the Weibull for the centre of the distribution but it improves the representation of the tail, where the non-Gaussian character of the wind components is better taken into account. In the southern region, this distribution performs better than the Weibull, except in the most complex regions (see the red dots in the south-eastern region in Fig. 10b). These stations correspond to high anisotropy of the wind components, the variance ratio $\sigma_u^2/\sigma_v^2 > 3$ (see Fig. 6a), which explains why the non-Gaussian distribution is not appropriate. Indeed we assumed similar parameters b and c for both components even if it is rarely the case (see Fig. 7). This hypothesis is necessary in order to compute an analytic expression for the wind-speed distribution, but it is sometimes too strong, especially in those very complex orographic environments.

Figure 11 summarizes the performances of the Rayleigh–Rice distribution. Figure 11b shows that it is doing similar or better than the Weibull on the centre of distribution at all stations, and better or similar on the tail at 73 over 89 stations. It does not only outperform the Weibull but Fig. 11a also shows that the fits are very good: $W_n^2 \approx 0$ at almost all stations. The Pau site, where the peak is not well fitted by the Rayleigh–Rice, is actually an exception since it is among the five worst W_n^2 scores. The Rayleigh–Rice distribution is designed for regions of flows channelled in valleys, or where sustained wind field prevails. Indeed, it performs very well in the southern region at stations where channelled flows create shouldered distributions, such as at Orange. The Rayleigh–Rice is capable of fitting the two peaks, so it improves the representation of the wind-speed statistics in these complex areas. Surprisingly, even in other areas without bimodal distribution, the Rayleigh–Rice distribution brings some improvement, so superposing two regimes
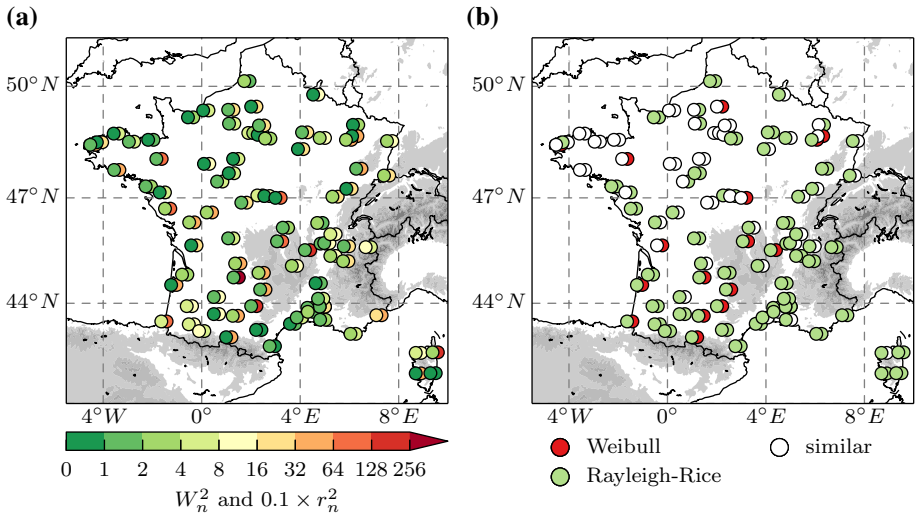
Fig. 11 Same as Fig. 9 for the Rayleigh–Rice distribution (in *green*) compared to the Weibull distribution (in *red*)

enables to better represent the shape of the wind speed statistics. This is consistent with the review of Carta et al. (2009) regarding mixture distributions involving the Weibull distribution.

The Rayleigh–Rice distribution is not easily fitted because it has four parameters and its PDF expression is quite complex. Under the assumption $\sigma_1 = \sigma_2 = \sigma$, we reduce to three parameters and Eq. 16 reduces to the following simpler form,

$$P_{RR}(M; \mu, \sigma^2, \alpha) = \frac{M}{\sigma^2} \exp\left(-\frac{M^2}{2\sigma^2}\right)\left[(1 - \alpha) + \alpha \exp\left(-\frac{\mu^2}{2\sigma^2}\right) I_0\left(\frac{M\mu}{\sigma^2}\right)\right]. \quad (17)$$

This three-parameter equation performs well in the flat areas but not in the valleys where it is important to assume different variances for the channelled and isotropic wind field in order to have a good fit on both peaks.

For the purpose of completeness, let us mention that a similar study compares Rice-like and Weibull distributions without accounting for the superposition of different weather regimes (Baïle et al. 2011). They also report a better description of the tails of the distributions by the Rice-like distribution, although it is less determinant in flat regions. This is one reason for which a Rayleigh–Rice distribution is proposed instead of a mixture of Weibull distributions (Carta and Ramírez 2007; Carta et al. 2009). Thus, taking into account the existence of persistent wind regimes is a good approach to quantify wind-speed statistics.

Figure 12 gives a visual summary of the comparison of the four distributions. In the left panel (Fig. 12a), we compare only the Weibull, elliptical and non-Gaussian distributions which all depend on two parameters only, whereas the right panel (Fig. 12b) also includes the Rayleigh–Rice four-parameter distribution. The white dots in Fig. 12a correspond to stations in north-western France where the wind components are close to Gaussian shape and without too much anisotropy (see Fig. 6), such as Nantes. The fits of all four distributions are quite accurate and very close, except on the tail where the Weibull and elliptical distributions tend to underestimate the probability of strong winds. In the other areas, we saw that the
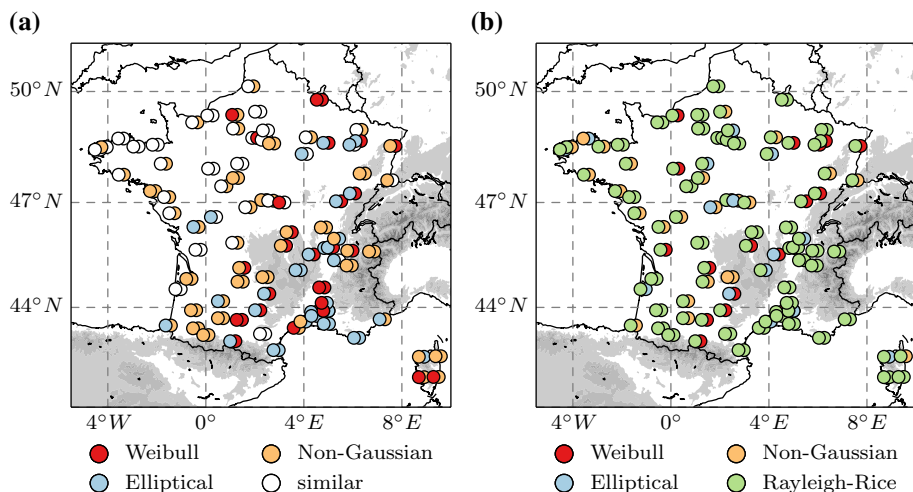
Fig. 12 **a** Best distribution between Weibull (*red*), elliptical (*blue*) and non-Gaussian (*orange*). *White dots* when all three distributions give similar results; **b** Best distribution between the same three and Rayleigh–Rice (*green*). At each station, the foreground (respectively background) circle gives the best distribution for the centre (respectively tail) of the distribution, based on CvM statistic W_n^2 (resp. the AD2R statistic r_n^2)

distributions are less accurate and either one of the three is the best fit, depending on the wind characteristics.

Finally, Fig. 12b shows the benefit of a mixed distribution such as the Rayleigh–Rice to model wind-speed statistics for a wide range of environments. This new distribution outperforms the other three almost everywhere. One can note that the Rayleigh–Rice distribution performs best even where the anisotropy ratio is much larger than 1 and/or where the wind components are not Gaussian. This could be seen as contradictory with the fact that the Rayleigh–Rice distribution is the mixture of two normal distributions. This suggests that the non-Gaussianity of the observed wind-speed distribution, which can be partly reproduced by our non-Gaussian distribution, is probably dominated by the bimodal nature of the distribution. Regarding anisotropy, the good behavior of the Rayleigh–Rice distribution suggests that the anisotropic nature of the wind-speed distribution is most probably carried by the existence of a sustained prevailing flow rather than different wind-component variances as proposed in McWilliams et al. (1979), McWilliams and Sprevak (1980) and Weber (1997). It also explains why the elliptical distribution performs worse than the Rayleigh–Rice distribution.

Other quantitative comprehensive evaluations could be used. A global performance index could be for instance the power of the distribution, namely the third moment of the distribution, which is maybe of more practical importance and allows comparisons between stations. We did not use this indicator since it does not ensure the best good fit of the distributions to the observations, which is a key aspect of this study. However the analysis of such index (not shown) confirms the analysis using the CvM and AD2R scores. The Rayleigh–Rice outperforms the other distributions with on average less than 2 % relative error with respect to the observations. The non-Gaussian is the less efficient with relative error often exceeding 20 % (half of the stations). The Weibull and elliptical distributions display similar performance (slightly better for the elliptical distribution) with relative errors ranging between 5 and 20 % in most stations.

6 Conclusion

The use of the Weibull distribution for wind statistics modelling is a convenient and powerful approach. It is however based on empirical rather than physical justification and might display at times strong limitations for its application. Based on wind measurements collected at 89 locations throughout France, for a wide range of environments, from flat to complex orography with different weather regimes, we compared the Weibull distribution and two other two-parameter probability density functions for the wind speed, here called elliptical and non-Gaussian. We therefore provide greater physical insight into the validity domain of the Weibull distribution, depending on the wind characteristics, mainly the fluctuations and anisotropy. The elliptical distribution assumes a Gaussian shape for the wind components but takes into account the anisotropy by assuming different variances for each component. The non-Gaussian distribution is based on the recently developed super-statistics theory. It assumes fluctuating variances of the two wind components, which are eventually modelled by a Gaussian distribution over a short time interval. But for analytic calculation purpose, the proposed wind-speed distribution does not take into account the anisotropy. Where the wind components are close to Gaussian shape and without too much anisotropy, such as at most stations in north-western France, all fits are quite close and rather accurate, except on the tail of the distribution where only the non-Gaussian distribution does not underestimate the strong wind probability. In more complex regions, close to the mountains, in southern and eastern France, the wind field can present anisotropy and/or departure from Gaussian shape, and either the elliptical or the non-Gaussian distribution can be better suited than the Weibull to represent the wind statistics. We also introduced a Rayleigh–Rice four-parameter distribution as a combination of a Rayleigh distribution to model the isotropic wind field and a Rice distribution to model persistent wind regimes. This gives excellent results, especially for the weather stations located in the Rhône or Aude valleys (where the mistral and tramontane channelled flows accur, respectively) where the Weibull or other two-parameter distributions, are not able to reproduce the observed shouldered distributions. Combining Rayleigh and Rice distributions is another way of applying the super-statistics theory, which models the wind system as the superposition of local dynamics at different intervals with different mean wind speeds.

Finally, this study points out the limits of using a unique analytic expression to model the wind statistics, since the wind field and its statistical distribution can greatly vary spatially. The more sophisticated distributions obviously fit more complex wind regimes better but with less simple estimation of their parameters. This is the case for our Rayleigh–Rice distribution that by far outperforms the other distributions at most stations. One use of parametric distributions, especially the Weibull distribution, is the statistical downscaling of near-surface wind speed to produce regional wind-speed climatologies (Pryor et al. 2005). We showed that a number of analytical distributions can represent wind speed distributions. Knowing properties such as surrounding topography, anisotropy, existence of persistent wind regimes can help in determining which distribution performs optimally. However, we also advocate non-parametric statistical methods, based on the wind-speed cumulative distribution function, or percentiles that would not be sensitive to the complexity of the observed wind-speed distribution (e.g. Michelangeli et al. 2009; Salameh et al. 2009; Lavaysse et al. 2012; Vrac et al. 2012).

Appendix 1: Computation of the Goodness-of-fit Statistics

The Cramer–von Mises and Anderson–Darling statistics have computing relations equivalent to Eq. 4 and better suited for computing (Luceño 2006). The sample of observations $(x_1, \ldots, x_n)$, of length n, is sorted so that $x_1 \leq \cdots \leq x_n$. Here, F is the CDF of the distribution, and we use the notation $z_i = F(x_i)$.

Cramer–von Mises relation:

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(z_i - \frac{i - 1/2}{n} \right)^2 \tag{18}$$

Anderson–Darling relation:

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left(\ln(z_i) + \ln(1 - z_{n+1-i}) \right) \tag{19}$$

Right-tail Anderson–Darling relation:

$$R_n^2 = \frac{n}{2} + 2 \sum_{i=1}^{n} z_i - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \ln(1 - z_{n+1-i}) \tag{20}$$

Right-tail Anderson–Darling of second-order relation:

$$r_n^2 = 2 \sum_{i=1}^{n} \ln(1 - z_i) + \frac{1}{n} \sum_{i=1}^{n} \frac{2i - 1}{1 - z_{n+1-i}} \tag{21}$$

Appendix 2: Calculations of the Non-Gaussian Distribution

The purpose here is to provide details of the analytic calculations of the non-Gaussian distribution defined in Sect. 4.2.

Distribution of Wind Components

Combining the distributions of Eqs. 8 and 9 into Eq. 7 gives:

$$p(u) = \frac{b^{-c}}{\sqrt{\pi} \, \Gamma(c)} \int \beta^{c - \frac{1}{2}} \exp \left[-\beta \left(u^2 + \frac{1}{b} \right) \right] d\beta. \tag{22}$$

The change of variable $t = \beta \left(u^2 + \frac{1}{b} \right)$ gives:

$$p(u) = \frac{b^{-c} \left(u^2 + \frac{1}{b} \right)^{-c+\frac{1}{2}} - 1}{\sqrt{\pi} \, \Gamma(c)} \int t^{c-\frac{1}{2}} e^{-t} \, dt. \tag{23}$$

We recognize in the integral term the definition of $\Gamma(c + \frac{1}{2})$, hence giving Eq. 10.

Distribution of Wind Speed

From Eq. 10, which gives the PDF of one wind component $p(u)$, we derive the PDF for the wind-speed M, under the assumption that both wind components u and v are represented by the same parameters b and c. Under the hypothesis of independence of u and v, $p(u, v) = p(u) p(v)$. We perform the change of variables $u = M \cos \phi$ and $v = M \sin \phi$ (transformation whose Jacobian is M), obtaining,

$$p(M, \phi) = \frac{M}{Z^2} \left(1 + b M^2 + \frac{b^2 M^4}{4} \sin^2 2\phi \right)^{-c-\frac{1}{2}}. \tag{24}$$

In order to obtain the distribution of M, we integrate $p(M, \phi)$ with respect to ϕ, between 0 and 2π. For this purpose, we use the following hypergeometric integral (see e.g. Erdélyi et al. 1953, p. 115 (7)),

$$\int_0^{\frac{\pi}{2}} \left(1 - z \sin^2 t \right)^{-a} dt = \frac{\pi}{2} \, F \left(a, \frac{1}{2}; 1; z \right). \tag{25}$$

Hence, we obtain Eq. 11:

$$p(M) = 2\pi \frac{M}{Z^2} \left(1 + bM^2 \right)^{-c-\frac{1}{2}} F \left(c + \frac{1}{2}, \frac{1}{2}; 1; \frac{-b^2 M^4}{4(1 + b M^2)} \right), \tag{26}$$

where F is the ordinary hypergeometric function.

References

Ahmad M, Sinclair C, Spurr B (1988) Assessment of flood frequency models using empirical distribution function statistics. Water Resour Res 24:1323–1328

Baïle R, Muzy JF, Poggi P (2011) An M-Rice wind speed frequency distribution. Wind Energy 14(6):735–748. doi:10.1002/we.454

Bastin S, Drobinski P (2005) Temperature and wind velocity oscillations along a gentle slope during sea-breeze events. Boundary-Layer Meteorol 114(3):573–594. doi:10.1007/s10546-004-1237-6

Bastin S, Drobinski P (2006) Sea-breeze-induced mass transport over complex terrain in south-eastern France: a case-study. Q J R Meteorol Soc 132(615):405–423. doi:10.1256/qj.04.111

Bastin S, Champollion C, Bock O, Drobinski P, Masson F (2005a) On the use of gps tomography to investigate water vapor variability during a mistral/sea breeze event in southeastern france. Geophys Res Let 32(L05808). doi:10.1029/2004GL021907

Bastin S, Drobinski P, Dabas A, Delville P, Reitebuch O, Werner C (2005b) Impact of the Rhône and Durance valleys on sea-breeze circulation in the Marseille area. Atmos Res 74(1–4):303–328. doi:10.1016/j.atmosres.2004.04.014

Bastin S, Drobinski P, Guénard V, Caccia JL, Campistron B, Dabas AM, Delville P, Reitebuch O, Werner C (2006) On the interaction between sea breeze and summer mistral at the exit of the Rhône valley. Mon Weather Rev 134(6):1647–1668. doi:10.1175/MWR3116.1

Bauer E (1996) Characteristic frequency distributions of remotely sensed in situ and modelled wind speeds. Int J Climatol 16:1087–1102

Beck C, Cohen EGD (2003) Superstatistics. Physica A Stat Mech Appl 322:267–275. doi:10.1016/S0378-4371(03)00019-0

Bernardin F, Bossy M, Chauvin C, Drobinski P, Rousseau A, Salameh T (2009) Stochastic downscaling method: application to wind refinement. Stoch Environ Res Risk Assess 23(6):851–859. doi:10.1007/s00477-008-0276-9

Bryukhan F, Diab R (1993) Decomposition of empirical wind speed distributions by laguerre polynomials. Wind Eng 17:147–151

Carta J, Ramírez P (2007) Analysis of two-component mixture weibull statistics for estimation of wind speed distributions. Renew Energy 32:518–531

Carta J, Ramírez P, Bueno C (2008) Considerations of the effects of winds on the drift of oil slicks at sea: statistical and temporal aspects of wind velocity, direction and persistence. Energy Convers Manag 49:1309–1320

Carta JA, Ramírez P, Velázquez S (2009) A review of wind speed probability distributions used in wind energy analysis: case studies in the canary islands. Renew Sustain Energy Rev 13(5):933–955. doi:10.1016/j.rser.2008.05.005

Chew V, Boyce R (1962) Distribution of radial error in the bivariate elliptical normal distribution. Technometrics 4:138–139. doi:10.2307/1266181

Colin B, Coupal B, Frayce D (1987) Considerations of the effects of winds on the drift of oil slicks at sea: statistical and temporal aspects of wind velocity, direction and persistence. Wind Eng 11:51–65

Cook NJ (2001) "Discussion on modern estimation of the parameters of the Weibull wind speed distribution for wind speed energy analysis" by J.V. seguro, T.W. lambert. J Wind Eng Ind Aerodyn 89(10):867–869. doi:10.1016/S0167-6105(00)00088-X

Crutcher HL, Baer L (1962) Computations from elliptical wind distribution statistics. J Appl Meteorol 1(4):522–530. doi:10.1175/1520-0450(1962)001<0522:CFEWDS>2.0.CO;2

Davenport AG (1966) The treatment of wind loading on tall buildings. In Proceedings of the symposium on tall buildings. University of Southampton, Pergamon Press, London

Drobinski P (2012) Wind and solar renewable energy potential resources estimation. Addison-Wesley, Reading, MA

Drobinski P, Bastin S, Guénard V, Caccia JL, Dabas A, Delville P, Protat A, Reitebuch O, Werner C (2005) Summer mistral at the exit of the Rhône valley. Q J R Meteorol Soc 131(605):353–375. doi:10.1256/qj.04.63

Drobinski P, Bastin S, Dabas A, Delville P, Reitebuch O (2006) Variability of three-dimensional sea breeze structure in southern France: observations and evaluation of empirical scaling laws. Ann Geophys 24(7):1783–1799

Drobinski P, Saïd F, Ancellet G, Arteta J, Augustin P, Bastin S, Brut A, Caccia JL, Campistron B, Cautenet S, Colette A, Coll I, Corsmeier U, Cros B, Dabas A, Delbarre H, Dufour A, Durand P, Guénard V, Hasel M, Kalthoff N, Kottmeier C, Lasry F, Lemonsu A, Lohou F, Masson V, Menut L, Moppert C, Peuch VH, Puygrenier V, Reitebuch O, Vautard R (2007) Regional transport and dilution during high-pollution episodes in southern france: summary of findings from the field experiment to constraint models of atmospheric pollution and emissions transport (ESCOMPTE). J Geophys Res 112:D13105. doi:10.1029/2006JD007494

Erdélyi A, Magnus W, Oberhettinger F, Tricomi FG, Bateman H (1953) Higher transcendental functions, vol 1. McGraw-Hill, New York, 302 pp

Erickson D, Taylor J (1989) Non-weibull behavior observed in a model-generated global surface wind field frequency distribution. J Geophys Res 94:12,693–12,698

Fisher R (1912) On an absolute criterion for fitting frequency curves. Messenger Math 41:155–160

Fisher R (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond Ser A 222:309–368

Gryning SE, Batchvarova E, Floors R, Peña A, Brümmer B, Hahmann AN, Mikkelsen T (2014) Long-term profiles of wind and weibull distribution parameters up to 600 m in a rural coastal and an inland suburban area. Boundary-Layer Meteorol 150:167–184. doi:10.1007/s10546-013-9857-3

Guénard V, Drobinski P, Caccia JL, Campistron B, Bench B (2005) An observational study of the mesoscale mistral dynamics. Boundary-Layer Meteorol 115(2):263–288. doi:10.1007/s10546-004-3406-z

Guénard V, Drobinski P, Caccia JL, Tedeschi G, Currier P (2006) Dynamics of the MAP IOP 15 severe mistral event: observations and high-resolution numerical simulations. Q J R Meteorol Soc 132(616):757–777. doi:10.1256/qj.05.59

He Y, Monahan A, Jones C, Dai A, Biner S, Caya D, Winger K (2010) Land surface wind speed probability distributions in North America: observations, theory, and regional climate model simulations. J Geophys Res 115(D04):103. doi:10.1029/2008JD010708

Justus CG, Hargraves WR, Yalcin A (1976) Nationwide assessment of potential output from wind-powered generators. J Appl Meteorol 15(7):673–678. doi:10.1175/1520-0450(1976)015<0673:NAOPOF>2.0.CO;2

Justus CG, Hargraves WR, Mikhail A, Graber D (1978) Methods for estimating wind speed frequency distributions. J Appl Meteorol 17(3):350–353. doi:10.1175/1520-0450(1978)017<0350:MFEWSF>2.0.CO;2

Lavaysse C, Vrac M, Drobinski P, Lengaigne M, Vischel T (2012) Statistical downscaling of the French Mediterranean climate: assessment for present and projection in an anthropogenic scenario. Nat Hazards Earth Syst Sci 12(3):651–670. doi:10.5194/nhess-12-651-2012

Lebeaupin Brossier C, Drobinski P (2009) Numerical high-resolution air–sea coupling over the Gulf of Lions during two tramontane/mistral events. J Geophys Res 114:D10110. doi:10.1029/2008JD011601

Li M, Li X (2005) Mep-type distribution function: a better alternative to weibull function for wind speed distributions. Renew Energy 30:1221–1240

Luceño A (2006) Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. Comput Stat Data Anal 51(2):904–917. doi:10.1016/j.csda.2005.09.011

McWilliams B, Sprevak D (1980) The estimation of the parameters of the distribution of wind speed and direction. Wind Eng 4:227–238

McWilliams B, Newmann M, Sprevak D (1979) The probability distribution of wind velocity and direction. Wind Eng 3:269–273

Menut L (2008) Sensitivity of hourly Saharan dust emissions to NCEP and ECMWF modeled wind speed. J Geophys Res 113:D16201. doi:10.1029/2007JD009522

Michelangeli PA, Vrac M, Loukos H (2009) Probabilistic downscaling approaches: application to wind cumulative distribution functions. Geophys Res Lett 36:L11708. doi:10.1029/2009GL038401

Morrissey M, Greene J (2012) Tractable analytic expressions for the wind speed probability density functions using expansions of orthogonal polynomials. J Appl Meteorol Clim 51:1310–1320

Orieux A, Pouget E (1984) Le mistral: contribution à l'étude de ses aspects synoptiques et régionaux. Monographie 5, Direction de la Météorologie

Pearson K (1894) Contribution to the mathematical theory of evolution. Philos Trans R Soc Lond Ser A 195:71–110

Pearson K (1902a) On the systematic fitting of curves to observations and measurements, part I. Biometrika 1:265–303

Pearson K (1902b) On the systematic fitting of curves to observations and measurements, part II. Biometrika 2:1–23

Pearson K (1936) Method of moments and method of maximum likelihood. Biometrika 28:34–59

Plaut G, Vautard R (1994) Spells of low-frequency oscillations and weather regimes in the northern hemisphere. J Atmos Sci 51(2):210–236. doi:10.1175/1520-0469(1994)051<0210:SOLFOA>2.0.CO;2

Pryor SC, Schoof JT, Barthelmie RJ (2005) Empirical downscaling of wind speed probability distributions. J Geophys Res 110:D19109. doi:10.1029/2005JD005899

Ramírez P, Carta JA (2005) 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 Manag 46(15–16):2419–2438. doi:10.1016/j.enconman.2004.11.004

Rizzo S, Rapisarda A (2005) Application of superstatistics to atmospheric turbulence. In: Beck C, Benedek G, Rapisarda A, Tsallis C (eds) Complexity, metastability and nonextensivity: proceedings of the 31st workshop of the international school of solid state physics. World Scientific Publishing Company, Incorporated, pp 246–254

Salameh T, Drobinski P, Menut L, Bessagnet B, Flamant C, Hodzic A, Vautard R (2007) Aerosol distribution over the western Mediterranean basin during a Tramontane/Mistral event. Ann Geophys 25:2271–2291

Salameh T, Drobinski P, Vrac M, Naveau P (2009) Statistical downscaling of near-surface wind over complex terrain in southern france. Meteorol Atmos Phys 103(1–4):253–265. doi:10.1007/s00703-008-0330-7

Seguro JV, Lambert TW (2000) Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J Wind Eng Ind Aerodyn 85(1):75–84. doi:10.1016/S0167-6105(99)00122-1

Simonnet E, Plaut G (2001) Space-time analysis of geopotential height and SLP, intraseasonal oscillations, weather regimes, and local climates over the North Atlantic and Europe. Clim Res 17(3):325–342. doi:10.3354/cr017325

Sinclair C, Spurr B, Ahmad M (1990) Modified Anderson Darling test. Commun Stat Theory Methods 19(10):3677–3686. doi:10.1080/03610929008830405

Smith A, Lott N, Vose R (2011) The integrated surface database: recent developments and partnerships. Bull Am Meteorol Soc 92(6):704–708. doi:10.1175/2011BAMS3015.1

Smith O (1971) An application of distributions derived from the bivariate normal density function. In: Proceedings of the international symposium on probability and statistics in the atmospheric sciences, pp 162–168

Sura P, Gille ST (2003) Interpreting wind-driven southern ocean variability in a stochastic framework. J Mar Res 61(3):313–334. doi:10.1357/002224003322201214

Takle ES, Brown JM (1978) Note on the use of Weibull statistics to characterize wind-speed data. J Appl Meteorol 17(4):556–559. doi:10.1175/1520-0450(1978)017<0556:NOTUOW>2.0.CO;2

Troen I, Petersen EL (1989) European Wind Atlas. Risø National Laboratory, Roskilde

Tsallis C (1988) Possible generalization of Boltzmann–Gibbs statistics. J Stat Phys 52:479–487. doi:10.1007/BF01016429

Tuller SE, Brett AC (1984) The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. J Appl Meteorol Clim 23(1):124–134. doi:10.1175/1520-0450(1984)023<0124:TCOWVT>2.0.CO;2

Vautard R (1990) Multiple weather regimes over the North Atlantic: analysis of precursors and successors. Mon Weather Rev 118(10):2056–2081. doi:10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2

Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L, Somot S (2012) Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. Nat Hazards Earth Syst Sci 12(9):2769–2784. doi:10.5194/nhess-12-2769-2012

Weber R (1991) Estimator for the standard deviation of wind direction based on moments of the Cartesian components. J Appl Meteorol 30:1341–1352

Weber R (1997) Estimators for the standard deviation of horizontal wind direction. J Appl Meteorol 36:1407–1415

Weisser D (2003) A wind energy analysis of Grenada: an estimation using the 'Weibull' density function. Renew Energy 28(11):1803–1812. doi:10.1016/S0960-1481(03)00016-8