energies

MDPI

Review

A Review on Wind Speed Extreme Values Modeling and Estimation for Wind Power Plant Design and Construction

Elio Chiodo [1], Bassel Diban [2], Giovanni Mazzanti [2,*] and Fabio De Angelis [3]

1 Department of Industrial Engineering, University of Naples Federico II, 80125 Napoli, Italy; chiodo@unina.it
2 Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, 40136 Bologna, Italy; bassel.diban2@unibo.it
3 Department of Structures for Engineering and Architecture, University of Naples Federico II, 80125 Napoli, Italy; fabio.deangelis@unina.it
* Correspondence: giovanni.mazzanti@unibo.it

Abstract: Rapid growth of the use of wind energy calls for a more careful representation of wind speed probability distribution, both for identification and estimation purposes. In particular, a key point of the above identification and estimation aspects is representing the extreme values of wind speed probability distributions, which are of great interest both for wind energy applications and structural tower reliability analysis. The paper reviews the most adopted probability distribution models and estimation methods. In particular, for reasons which are properly discussed, attention is focused on the evaluation of an opportune "safety index" related to extreme values of wind speeds or gusts. This topic has gained increasing attention in recent years in both wind energy generation assessment and also in risk and structural reliability and safety analysis. With regard to wind energy generation, there is great sensitivity in the relationship between wind speed extreme upper quantiles and the corresponding wind energy quantiles. Concerning the risk and reliability analysis of structures, extreme wind speed value characterization is useful for a proper understanding of the destructive wind forces that may affect structural tower reliability analysis and, consequently, the proper choice of the cut off wind speed value; therefore, the above two kinds of analyses are somewhat related to each other. The focus is on the applications of the Bayesian inference technique for estimating the above safety index due to its effectiveness and usefulness.

Keywords: Bayesian estimation; extreme wind speeds; safety analysis; peak over threshold; wind turbines; design and construction; structural reliability

Citation: Chiodo, E.; Diban, B.; Mazzanti, G.; De Angelis, F. A Review on Wind Speed Extreme Values Modeling and Estimation for Wind Power Plant Design and Construction. *Energies* **2023**, *16*, 5456. https://doi.org/10.3390/en16145456

Academic Editor: Davide Astolfi

Received: 23 May 2023
Revised: 3 July 2023
Accepted: 12 July 2023
Published: 18 July 2023

1. Introduction

Wind energy is one of the faster-spreading renewable energy sources (RES) that have been developed as an effective response to growth and global economic development over decades [1–3]. This has been led by the continuous penetration of renewables, including wind turbines, in the electrical power system. However, the penetration of renewable generation in the electrical grid has raised many questions related to the reliability of the power system and the viability of wind farms [4,5]. This assessment is essential for investigating the feasibility, energy costs, and economic benefits of wind farm projects, which justifies the importance of wind speed probability distributions in this process to reduce the uncertainty of wind energy estimation [6,7].

Wind turbines are designed to operate at optimum wind speeds. Higher winds lead to cutting off power generation and, in extremes, lead to structural damage to blades and towers. Thus, wind energy production strongly depends on wind speed probability distribution. For these reasons, extreme wind distributions provide information on rare but critical events applied to wind turbines [8,9].

Extreme value (EV) theory provides a parametric description and modeling of extreme and rare events, i.e., tails in probability distributions. It is usually used in risk assessment, management, and prevention. It has a wide range of applications, e.g., structural engineering, finance, earth science, traffic prediction, and geological engineering, including hydrology and wind extremes estimation. This is an important statistical theory in engineering and applied science [10]. EV theory has also been introduced as a good fit for the upper quantiles of wind speeds, which have a small sample size compared to lower wind speeds [4]. It was introduced by Fisher and Tippet [11], then developed by Gumbel [12]. Subsequently, it was progressively and extensively used to estimate wind extreme values in many research articles [13–17] and review articles [1,6,10,18,19].

Nonparametric models are also used to estimate wind speed extreme values [20]. Peak Over Threshold (POT) is considered one of the most common nonparametric methods whereby large individual events can be included. Probabilistic modeling of POT was introduced in [21–23] and statistically developed in [24–26], whereby the generalized Pareto distribution is fit to the extreme values of a variable.

Parameter estimators for parametric models have always been a research topic of high importance [27]. The aim is to estimate a proper *safety index* in the absence of large extreme wind data as an alternative to classical statistical analysis [4]. Indeed, extreme value models depend on the extrapolation of statistical models beyond the main observed data [28], while Bayesian analysis provides a method to update beliefs about an unknown quantity of interest based on the occurrence likelihood, which is a function of parameters of a given statistical model [29]. However, meaningful prior information is necessary to improve the quality of the Bayesian analysis, as demonstrated in [28,30].

This review aims to investigate the most used models in the literature to estimate wind speed extreme values, which impact both wind farm design and construction, as well as the structural reliability [31–33] and reliability of the electrical power delivered to the power grid. The structure of the paper is as follows. After the present Introduction, in Section 2, the paper reviews the extreme value theory and the most adopted models of such theory; both parametric and nonparametric models are investigated in this review. This paves the way for the subsequent Sections 3 and 4 on Bayes inference methods for extreme wind speed; for the sake of brevity, such Bayes inference methods are only exemplified in the POT framework, but the same methodology can be applied to all other previously reviewed models. Conclusions are drawn in Section 5, and the basics of some useful probability distributions used in the framework of extreme wind speed estimation are recalled.

2. Extreme Value Theory

Extreme value theory (EVT) has undergone fast development and has become a mature and significant probabilistic theory [34]. Extreme values are either very small or very large values in a given set of random variables. Extreme value theory is successfully used in the most diverse fields, such as structural reliability, biometrics, finance, insurance, and risk theory, while its appearance in the field of renewable energy, namely to describe extreme wind speed (EWS) statistics, is relatively recent [1]. Its statistical aspects have recently undergone considerable development due to the fact that rare events can have catastrophic consequences [6]. Some examples are earthquakes and other environmental events, such as floods, precipitation, and other high-risk, low-probability situations [35].

Let us deal with a sequence set of random variables $(X_1, X_2, \ldots, X_n)$, assumed as statistically independent and identically distributed (IID), and our interest is in their extremes [27]:

$$\begin{aligned} W_n &= min(X_1,\ X_2\ , \ldots, X_n \ldots) \\ Z_n &= max\ (X_1,\ X_2\ , \ldots, X_n \ldots) \end{aligned} \quad (1)$$

In this paper, being interested in the peak values of wind speed (WS), our interest is mainly focused on random variables (RVs) such as Z_n.

Being related to values that, by their nature, rarely occur, EVT studies modeling events have a very small probability of occurrence, and this requires adequate statistical methods for their inference. That is why the adoption of Bayesian methods in this field has experienced rapid growth, but it is still not so widespread in the field of renewable energy;

therefore, an account of such methods will also be illustrated in this review paper. For more traditional or Maximum Likelihood methods, the reader is referred to authoritative books such as [36].

There are several methods for modeling extreme values depending on how the sampling process is carried out. Sampling schemes in extreme value modeling can be classified into time-based and event-based approaches. In the case of a time-based sampling scheme, a series of observations are blocked into fixed intervals over time, and the block maxima (or minima) are dealt with as extremes, while in the event-based sampling scheme, the observations are treated as extremes if a threshold is exceeded. The resulting distributions from these two sampling schemes, if applied properly, are independent of the true underlying distribution, typically unknown in practice [13].

Once the sampling strategy is made, the attention of EVT is typically focused on the limiting behavior of the sequence of maximum values of the given set of RV. A brief reference is made here to Gumbel's block methodology, while in the next section, the POT methodology is illustrated [37].

Assuming that a linear normalization of the sequence of maximum values is possible, Z_n, so that a non-degenerate limit is attainable for the sequence $(Z_n - b_n)/a_n$, with $a_n > 0$ and $b_n \in R$. Then, according to Gnedenko's theorem, the RV has a cumulative distribution function (CDF) of the type of the extreme value distribution (EVD), given—for any $x \in$ R—by:

$$\begin{aligned} F_{EV}(x) &= exp(-exp(-x)) \quad if\ d = 0 \\ F_{EV}(x) &= exp\left(-(1 + d\,x)^{1/d}\right) \quad if\ d \neq 0 \end{aligned} \tag{2}$$

and d is the so-called extreme value index, a key parameter in EVT. We then say that F—the common CDF of random variables ($X_1, X_2, \ldots X_N$)—is in the max-domain of attraction of EVD in (1) [27].

If $d < 0$, the right tail is "short", i.e., sup(x: $F(x) < 1$), the right endpoint of F, is finite. This class is called the Weibull class and contains, among others, the Uniform and the Inverse Burr CDFs.

If $d > 0$, the right tail is "heavy", of a negative polynomial type, and F has an infinite right endpoint. Examples in this class (the so-called "Frechet class") are the Pareto, Burr, Student's, and Log-Gamma CDFs. If $d = 0$, the right tail is of an exponential type, and the right endpoint can then be either finite or infinite. This class (the so-called "Gumbel class") encompasses the Exponential, Normal, Lognormal, Gamma, and classical Weibull CDFs, with an infinite right endpoint, but also (less frequently) some models with a finite right endpoint [35].

In most of the many applications of EVT, the EVD in Equation (2) is often rewritten on the three domains of attractions as follows:

Weibull:

$$F_W(x;\ d) = exp\left(-(-x)^{-1/d}\right) \qquad (d < 0) \tag{3}$$

Frechet or "Inverse Weibull" (IW):

$$F_{IW}(x;\ d) = exp\left(-x^{-1/d}\right) \qquad (d > 0) \tag{4}$$

Gumbel:

$$F_G(x;\ d) = exp(-\exp(-x)) \qquad (d = 0) \tag{5}$$

It is recalled that such models are suitable when the data consist of a set of maxima.

The CDFs of (2) through (4) will be further discussed when appropriate while illustrating the various EWS models in Section 3.

3. Extreme Value Models

Wind Power Density (WPD) is the available wind power per square meter of the swept area of a turbine. It has a cubic function with respect to the wind speed:

$$WPD = \frac{P}{A} = \frac{1}{2}\rho\vartheta^3 \quad (6)$$

where ϑ is the wind speed (m/s), ρ is the air density (kg/m^3), P is the total power available at the area swept by the rotor's blades (W), A is the area swept by the rotor's blades (m^2). Equation (6) shows that the p-quantiles of the wind speed are transformed into p-quantiles of the WPD, especially the upper quantiles (e.g., 0.95 or 0.99), which correspond to high WPD, hence high wind speed. In this case, the small variations in wind speed may be translated into great variations in WPD. The latter highlights the importance of the accurate assessment of wind speeds [4]. Many parametric and nonparametric models for the estimation of extreme wind speeds can be found in the literature, as follows.

(a) parametric extreme value distribution models:

(I) Gumbel distribution (Type I)

First introduced by Emil Gumbel [12]. This distribution is widely used to model extreme events.

PDF:

$$f(x,\chi,\delta) = \frac{1}{\delta}exp\left(-\frac{(x-\chi)}{\delta} - exp\left(-\frac{(x-\chi)}{\delta}\right)\right) \quad (7)$$

where x is the random variable (RV), χ is the location parameter, and δ is the scale parameter ($\delta > 0$).

CDF:

$$F(x,\chi,\delta) = exp\left(-exp\left(-\frac{(x-\chi)}{\delta}\right)\right) \quad (8)$$

Gumbel distribution fits extreme wind speeds in many studies. In [6], Gumbel distribution is the best fit for extreme wind speeds at 143 stations in the USA. However, the infinite upper tail of the Gumbel distribution seems to be physically unrealistic to delineate the bounded nature of the peak extreme wind speeds [6].

(II) Inverse Weibull distribution (Fréchet distribution Type II)

This distribution was used by Maurice Fréchet in his paper [38] to fit extreme events. Two-parameter IW distribution with a scale parameter α and a shape parameter β can be written as follows:

PDF

$$f(x,\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{\alpha}{x}\right)^{\beta+1} exp\left(-\left(\frac{\alpha}{x}\right)^{\beta}\right) \quad (9)$$

CDF

$$F(x,\alpha,\beta) = 1 - exp\left(-\left(\frac{\alpha}{x}\right)^{\beta}\right) \quad (10)$$

While three-parameter IW distribution with an additional location parameter μ can be written as follows:

PDF

$$f(x,\alpha,\beta,\mu) = \frac{\beta}{\alpha}\left(\frac{\alpha}{x-\mu}\right)^{\beta+1} exp\left(-\left(\frac{\alpha}{x-\mu}\right)^{\beta}\right) \quad (11)$$

CDF

$$F(x,\alpha,\beta,\mu) = 1 - exp\left(-\left(\frac{\alpha}{x-\mu}\right)^{\beta}\right) \quad (12)$$

Sarkar et al. obtained the best fit to extreme wind speeds when Fréchet distribution is used [39]. Inverse Weibull distribution is used by Chiodo et al. for extreme wind speed estimation showing a very good fitting [4].

(III) Weibull Distribution (Type III):

This distribution was first introduced by Weibull to fit extreme events. Two-parameter Weibull distribution with a scale parameter α and a shape parameter β can be written as follows:

PDF

$$f(x,\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) \tag{13}$$

CDF

$$F(x,\alpha,\beta) = 1 - exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) \tag{14}$$

While three-parameter Weibull distribution with an additional location parameter μ can be written as follows:

PDF

$$f(x,\alpha,\beta,\mu) = \frac{\beta}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{\beta-1} exp\left(-\left(\frac{x-\mu}{\alpha}\right)^{\beta}\right) \tag{15}$$

CDF

$$F(x,\alpha,\beta,\mu) = 1 - exp\left(-\left(\frac{x-\mu}{\alpha}\right)^{\beta}\right) \tag{16}$$

Two-parameter Weibull distribution is widely used in various fields, including the estimation of wind speeds [40]. On the other hand, three-parameter Weibull distribution showed a high accuracy in estimating low or null wind speeds since the location parameter shifts the Weibull peak horizontally, which helps to model low wind speeds [41,42]. For instance, many studies are found in the literature that used two-parameter Weibull distribution to estimate wind energy and characterize wind speed [40,43,44]. However, Weibull distribution is not applicable for extreme values of wind speed that have little influence on the parameters of Weibull distribution. Therefore, beyond a certain threshold, other extreme values distributions must be used [13]. Perrin et al. also showed that the Weibull distribution generates an incorrect estimation of the tails of the distribution; additionally, confidence bounds are not provided in this case [13].

(IV) The generalized extreme value distribution

The latter three distributions can be combined in one generalized extreme value (GEV) distribution which is widely used in the case of extreme events [45]. The CDF of the generalized value distribution is described as follows:

$$F(x,\alpha,\beta,\mu) = exp\left(-\left(1+\beta\left(\frac{x-\mu}{\alpha}\right)^{-1/\beta}\right)\right) \tag{17}$$

where $\alpha > 0$ while the tail-length or shape parameter (β) and the location parameter (μ) are real values. The GEV distribution is versatile, whereby α has a significant impact on the skewness and the kurtosis [46].

The GEV distribution is a heavy right tail distribution compared to the Weibull distribution [47]. The latter allows estimating extreme wind values above a certain threshold with higher accuracy. However, the estimation of the threshold of extreme wind speeds is challenging [48]. In [49], the GEV model is truncated to zero at the left tail to avoid predicting negative wind speeds [49]. Pinheiro et al. revealed that GEV distribution is more flexible in fitting extreme wind speeds using Monte Carlo simulations and data sets of wind speeds in Florida, USA [46].

(V) The generalized Pareto distribution

The generalized Pareto distribution (GPD) is also used in the application of extreme wind speed estimation. Holmes et al. found that GPD makes use of all relevant data on extreme wind speeds, not only the annual maxima [50]. Brabson et al. compared GPD to GEV and found that GPD reduces the uncertainty in the quantile variances. However, the choice of the threshold becomes more challenging [51].

PDF:

$$f(x,\alpha,\beta) = \frac{1}{\alpha}\left(1 - \frac{x-\mu}{\alpha}\right)^{\frac{1-\beta}{\beta}} \tag{18}$$

CDF

$$F(x,\alpha,\beta) = 1 - \left(1 - \frac{\beta}{\alpha}(x-\mu)\right)^{\frac{1}{\beta}} \tag{19}$$

where $\beta > 0$ when $\mu < x < \mu + \alpha/\beta$ and $\beta \leq 0$ when $x > \mu$.

(VI) Gamma Distribution

Gamma distribution is one of the most used distributions in extreme wind speed modeling. It is used to overcome the constraint at both tails in the Weibull distribution.

PDF:

$$f(x,\alpha,\beta) = \frac{\alpha^{\beta}}{\Gamma(\beta)} x^{\beta-1} exp(-\alpha x) \tag{20}$$

where α is the scale parameter and β is the shape parameter.

The main disadvantage of Gamma distribution is that it, as inferred from Section 2 on extreme value theory, does not process the theoretical properties which make it suitable to represent extreme values (properties which are instead possessed by the Gumbel, Inverse Weibull, and other models here illustrated). Therefore, extreme wind speed estimation using Gamma distribution is challenging compared to other methods [52]. Kiss et al. found that Gamma distribution provides an adequate and unified description of a wide range of wind speeds of the ERA-40 database covering 44 years in Europe [42].

(VII) Mixed Distributions:

In some wind climates, it has been found that extremes are drawn from two or more distributions, making it necessary to consider mixed probability distributions. Mixed distributions are of two types: the first includes two distributions having the same type but with different parameters, while the second type includes two different types of extreme values distributions. According to Raynal et al., the mixed distributions might be formed as follows [53]:

$$F(x) = p\ F_1(x) + (1-p)\ F_2(x) \tag{21}$$

where x is the RV, $F_1(x)$, $F_2(x)$ the two mixed distributions with their parameters, p is the association parameter $0 < p < 1$. Mixed General Extreme Value (MGEV) distributions include two GEV distributions $F_1(x)$ and $F_2(x)$ and can be formed in a similar way as MGED [54]. Similarly, Escalante et al. applied the Mixed Reverse Weibull (MRW) distribution [55]. Mixed Gumbel-Reverse Weibull Distribution (G-RW) and Mixed Gumbel-General Extreme Values Distribution are also discussed in [56]. Rossi et al. used two-component Extreme Value Distribution [57]. However, it is still challenging to choose the type and the number of the different distributions to be mixed, as the fitting is significantly affected by the types of the mixed distributions [58,59], not to mention the complexity of the parameters estimation process of the mixed distributions, which might lead to overparameterization [1]. Multi-parameter probability distributions, whereby more than three parameters are included, have been used to pursue higher fitting accuracy, such as Burr distribution, Johnson SB distribution, Kappa distribution, and five-parameter Wakeby distribution [1]. The Compound Inverse Rayleigh (CIR) distribution is used by Chiodo et al. in [60] and compared with two other distributions having the same median, i.e., the Inverse Rayleigh and the exponential distributions. The results showed that CIR has an 18% greater 0.95-quantile compared to the Rayleigh distribution. This model showed high efficiency in extreme wind speed estimation [60].

(b) Nonparametric Distribution Models:

In nonparametric models, no specific assumptions on the data distribution are made a priori. In this case, there is no bias linked to specific models, and there are no parameters to estimate. The performance of nonparametric models is better and more robust than that of the parametric models, as it is challenging to find a qualified parametric distribution that describes the actual wind speeds [1]. The Kernel Density Estimator shows a higher fitting accuracy of wind speeds in [61]. The Maximum Entropy Principle has also been used in [8,62] for low wind speeds and power density estimation, demonstrating more accuracy than parametric distributions. Furthermore, the Maximum Entropy Principle is also used for extreme values estimation in [63].

(c) Peak Over Threshold (POT)

A different approach to the probabilistic modeling of wind speed extreme values is the stochastic evaluation, which gives origin to the so-called "peak over threshold" (POT) approach [19,64]. Such a dynamic approach is adopted, in particular, in the framework of assessing adequate tower safety margins. Towers are constructed to endure for their operational lifetime, which may last many years. In this case, the designers need to estimate the extreme values of wind speed (WS), i.e., the maximum wind gust amplitude over a prefixed time. Hereinafter, wind gust amplitude over a given time will be simply denoted by "gust". Let us also denote the stochastic process of WS values over time by $S = S(t)$, and let z be the threshold that is a sufficiently high value of WS in such a way that every value of WS higher than z can be considered a gust [65,66]. This value is related to the tower structure and is typically utilized to define the cut-off value of the WS [67,68]. Then, let $C_z(t)$ represent the stochastic counting process of the WS values crossing the threshold z, resulting in the number of peaks over threshold. Introducing mild assumptions, such as that the mean duration of each gust τ_k is much smaller than the mean time between the successive crossings u_k, in addition to the assumption that the barrier level is high enough, the $C_z(t)$ process can be described, as deduced in the extremal processes books [69], by the well-known Poisson probability law $p(k,t)$ expressing the probability that $C_z(t)$ attains a given integer value k; such probability law is given by:

$$P(h,t) \equiv P[C_z(t) = h] = e^{-\lambda t} \cdot \frac{(\lambda t)^h}{h!} \quad = 0, 1, \ldots, \infty \tag{22}$$

In (22) λ is the mean number of threshold crossings in unit time. The mean (or expected value) and variance of the process $C_z(t)$ are numerically equal and given by:

$$E[C_z(t)] = Var[C_z(t)] = \lambda t \tag{23}$$

Focusing on the amplitude of gust at T_k, which is a random variable, denoted by G_k. A safety index to characterize the extreme values $C = C_z(t)$ is the maximum gust amplitude over a given time. Furthermore, it is an obvious index of the damage to the system caused by the gusts. The latter can be satisfied by associating to the stochastic process $C_z(t)$ and to the random variables G_k ($k = 1, 2, \ldots, C(t)$), the following stochastic process [70,71]:

$$\begin{gathered} M(t) = max[G_1, G_2, \ldots, G_C],\ if\ C(t) \geq 0 \\ M(t) = 0,\ otherwise. \end{gathered} \tag{24}$$

where (as in the following): $C(t) = C_z(t)$. Let $Q(t)$ be the CDF of $M(t)$ at time t:

$$Q(t,m) = P[\ M(t) < m\ ] \tag{25}$$

By assigning a safety level, m^*, the following Unsafety Index (UI), $U(t)$— which is also a stochastic process—can be consequently defined:

$$U(t) = 1 - Q(t, m^*) \tag{26}$$

Indeed, being $Q(t,m^*) = P[M(t) < m^*]$, the probability that m^* is never exceeded over the time horizon $(0,t)$, then $U(t)$ is easily shown to express the probability that m^* is exceeded at least once over $(0,t)$, which justifies its name as "Unsafety index".

In order to express the above UI in a more compact form, let us notice that for each value c of $C(t)$—the relationship becomes:

$$\begin{gathered}[max[G_1,\ G_2, \ldots,\ G_c] < m^*] \\ if\ and\ only\ if:\ [(G_1 < m^*) \cap \ldots \cap (\ G_C < m^*)]\end{gathered} \tag{27}$$

We assume that the RV G_k are statistically independent and identically distributed with the common, time-independent, cumulative distribution function $\varphi(x)$:

$$\varphi(x) = P(G_k \le x), \forall\, k = 1, 2, \ldots \ldots n \tag{28}$$

After trivial manipulations [72], one can obtain the following equation for the function $Q(t{,}m^*)$ under the Poisson hypothesis for $C_z(t)$:

$$Q(t, m^*) = \ exp[-\lambda t(1 - \varphi(m^*))] \tag{29}$$

In the paper, for the sake of estimation, for reasons to be explicit in the following, no explicit formulation will be adopted to model the parent distribution $G(z)$. The *UI* is an Exponential complementary CDF, as shown in the following equation:

$$U(t) = 1 - exp(-\lambda t w) \tag{30}$$

having defined:

$\lambda = \lambda_z$ = mean gust frequency (i.e., expected number of gust occurrences per unit time);

$w = w\ (m^*) = 1 - \varphi(m^*) = P(G_j > m^*)$ = exceedance probability (*EP*) of the value m^* by any single *RV* G_j.

It has to be highlighted that the EP depends neither on the index j nor on time. The function $w = w(m^*)$, i.e., the *EP*—which is related to the gust CDF—may, of course, assume various equations. As will be discussed in the sequel, a Bayesian approach appears to be particularly adequate to develop inference for such a model, as shown in Section 4.

4. On Bayes Inference Methods for Extreme Wind Speed (EWS) and Related Safety Indices Distribution

4.1. Introduction

The classical, and by far the most adopted, procedure for the estimation of the EWS statistics is usually based on the likelihood function (LF). However, this procedure is known to be efficient only for a suitably large amount of data. While this aspect is often not a serious drawback for the most general WS statistical analysis, it may become so for very extreme wind speed values, which rarely occur [73].

Moreover, it is a desirable and time-saving requirement that a rapid decision on the installation of a wind farm can be accomplished even in the presence of a limited amount of data [37]. Furthermore, on more theoretical grounds, the classical procedure does not provide the accuracy of the estimated parameters. In the Bayesian approach, the unknown parameters are considered random variables or uncertain quantities characterized by their probability distributions, and a priori (or prior) information is used to describe the a priori distribution of the parameters. The Bayes theorem supplies the parameters' distribution conditioned to the observed data [74–77].

Bayesian inference is an efficient tool for utilizing experimental field data (data that constitute the only source of knowledge referred to in classical statistical inference) and prior knowledge (which in an engineering environment are always available). By integrating such sources, it is possible to derive an effective estimation procedure and to include available information when there is a possible lack of available data [78]. The present section is particularly devoted to Bayesian inference applications to the (un) safety index's estimation in the framework of the POT methodology, which is more complex, being related to stochastic process theory, than the inference on standard EV Distribution of WS. This latter topic may be deduced from classical books or papers on Bayesian inference [29,65,78–83].

So, the Bayesian estimation method here is focused on the UI in the POT approach. As previously illustrated, the expression of UI, with respect to the time interval (0,t), is given by:

$$U = U(t) = 1 - exp(-\lambda t w) \tag{31}$$

The dependence on t, the time horizon under study, is often omitted since it is a fixed parameter. Moreover, the complementary function $S = S(t) = 1 - U$ (S is obviously denoted "Safety Index", *SI*) is often used as an alternative of U. In addition, since t is a fixed parameter, let it be denoted by τ in the following, so that the following uncertain parameter U is the one to be estimated:

$$U = U(\tau) = 1 - exp(-\lambda \tau w) \tag{32}$$

So that the *SI* and the *UI*, referred to as the interval (0,τ), are expressed as

$$S = \exp(-\lambda w); U = 1 - S \tag{33}$$

In the two following sub-sections, the nonparametric (Section 4.2) and the parametric approach (Section 4.3) to the above POT methodology are illustrated. Typical numerical results obtained by the above Bayes approach are only hinted at in this section, referring to the above papers [4,27,37,60] for numerical details.

Here, it is deemed opportune to give only a brief account of classical statistical estimation of the SI parameters, i.e., λ and w (here, lowercase letters are adopted for such parameters since, in classical statistics, they are regarded as unknown deterministic constants, and not RV as in Bayesian statistics). It is well known that, denoting by ξ' the Maximum Likelihood (ML) estimate of a generic parameter ξ:

- The *ML* estimate of λ is given on the basis of an available random sample of the time between gusts: (T_k: k = 1, . . . , n), by:

$$\lambda' = n / \sum_{j=1}^{n} T_j \tag{34}$$

- The ML estimate of the *EP*, w, once observed the data:
  N = number of gusts in the given interval $(0, \tau)$ and:
  M = number of exceedances in the same interval,
  is simply given by:

$$w' = M/N \tag{35}$$

Such two estimators are consistent, and w' is also unbiased for w. Of course, given the invariance of *ML* estimates, the *ML* estimate for the *SI*, S', is given, in view of (33), by:

$$S' = exp[-w'\lambda'] \tag{36}$$

Of course, the merits of the Bayesian estimation method to be illustrated in the next sub-sections are to be compared with those of such *ML* estimates for the *SI* in order to appreciate their efficiency [37].

4.2. A Bayesian Estimation Method in the Nonparametric Approach

The Bayesian estimation procedure is herein applied for the *SI* or the *UI* without assuming a particular parametric form for the CDF F [77,84]. In the paper [37], a Beta prior PDF is assigned directly to w in the case in which some prior information is given on $\lambda(m^*)$: in other words, here, no functional form is assumed for λ, but some prior knowledge exists on the probability that the WS value m^* is surpassed. For the estimation, the "input data" is a joint prior PDF, denoted as $g(w, \lambda)$, for the parameters of interest w and λ. It is well known that in the Bayesian methodology, the parameters w and λ to be estimated are considered as RV [35]. Accordingly, they are provided with a PDF which can be integrated and updated with field data, herein denoted by D, by the reported Bayes' theorem:

$$g(w,\lambda|D) = \frac{g(w,\lambda)L(D|w,\lambda)}{C} \tag{37}$$

where $L(D\,|\,w,\lambda)$ represents the Likelihood of the data D conditional to the parameters (w,λ), that is the PDF or—for discrete observations as the considered case—the "probability mass" functions; C represents a constant (with respect to the parameter values) given by:

$$C = \int_o^{+\infty}\int_o^{+\infty} g(w,\lambda)L(D|w,\lambda)dwd\lambda \tag{38}$$

The best Bayes estimator depends on the choice of the so-called "Loss Function". By far, the most adopted Loss Function is the "Square Error Loss Function" (SELF), by which the best estimator is the one that minimizes the mean square error. The best SELF Bayes estimator—of a given function $\xi = \xi(w,\lambda)$—i.e., the one which minimizes the SELF is provided by the posterior mean:

$$\xi^\circ = \mathrm{E}[\xi|D] = \int_o^{+\infty}\int_o^{+\infty} \xi(w,\lambda)g(w,\lambda|D)dwd\lambda \tag{39}$$

where ξ° is a Bayes estimate of the generic parameter ξ.

For the examined case, we are estimating UI, so that, for a time t, one can write:

$$U^\circ = \mathrm{E}[U|D] = \int_o^{+\infty}\int_o^{+\infty} (1-exp(-w\lambda))g(w,\lambda|D)dwd\lambda \tag{40}$$

The Bayes inference discussed in [27] expresses the "conjugate" priors for the RV w and λ, i.e., the Beta prior PDF (Appendix A) for w and the Gamma prior PDF (Appendix B) for λ. The two random variables are considered to meet the assumptions of statistical independence.

In the following, the suffix "0" is used to denote the prior PDF parameters, and the suffix "1" to denote the posterior PDF parameters. Accordingly, the prior joint PDF $g(w,\lambda)$ is expressed by:

$$g(w,\lambda) = betapdf(w;r_0,s_0)\cdot gampdf(\lambda;n_0,\delta_0) \tag{41}$$

The values of the parameters $(r_0,s_0,n_0,\ \delta_0)$, are deduced from prior information, i.e., test plant or experts' judgments. Data collection is provided by measuring the number of gusts $N(\tau)$, in the time interval $(0,\tau)$, and the number of gusts M that exceed the fixed value m*. When the time interval τ has been fixed, the number of gusts $n = N(\tau)$ becomes a constant, and the RV M becomes a Binomial RV that represents the number of exceedances in n independent proofs (see Appendix D) [85].

The suitable combination of the prior PDF and the likelihood function, in accordance with Bayes' formula, allows verification of the posterior PDF of w and λ are again Beta and Gamma, respectively, with updated values of the parameters as a function of the measured values n and m, which represent sufficient statistics for the problem [86]. In this regard, the following relation is expressed (see Appendix D):

$$g(w,\lambda|D) = betapdf(w;r_1,s_1)\cdot gampdf(\lambda;n_1,\delta_1) \tag{42}$$

where:

$$r_1 = r_0 + m \tag{43}$$

$$s_1 = s_0 + n - m; \tag{44}$$

$$n_1 = n_0 + n \tag{45}$$

$$\delta_1 = \delta_0/(1 + \tau\delta_0) \tag{46}$$

The expression above entails the posterior conditional independence of w and λ, once the data D are assigned. Moreover, in case the following relation holds among the prior parameters:

$$n_0 = r_0 + s_0 \tag{47}$$

then, the PDF is expressed in a straightforward manner. In such case, the product $w\ \lambda$ has Gamma distribution with parameters $(r_0,\ \delta_0)$ according to the properties of the Beta and Gamma distribution [81]. As a consequence, the expression:

$$V = w\lambda \tag{48}$$

is still Gamma distributed with parameters $(r_0,\ \delta_0)$. Therefore, given that the *SI* is expressed by:

$$S(t) = exp(-w\lambda) = exp(-V) \tag{49}$$

the following relation can be inferred:

$$V = -lnS \tag{50}$$

Which means that since *W* has a Gamma PDF, the SI *S* has a Negative Log Gamma PDF (Appendix C). Consequently, the prior PDF for *S* is:

$$g(s) = nlgampdf(s; n_o, \delta_0) \tag{51}$$

i.e., the prior PDF $p(s)$ is:

$$g(s) = \frac{1}{\Gamma(r_0)\delta_0^{r_0}} s^{\left(\frac{1}{\delta_0}-1\right)}(-logs)^{r_0-1},\ 0 \le s \le 1 \tag{52}$$

S is an RV that falls in the range (0,1), as in accordance with the relation (49). In a similar way used to update the prior PDF g(w,λ) into the posterior PDF $g(w,\lambda|D)$, by recalling the relationship between (w,λ) and the SI, one obtains, for the Safety Index SI = $Q(t)$, the following posterior PDF, for the assumed data collection D:

$$g(s|D) = nlgpdf(s; r_1, \delta_1) \tag{53}$$

where $r_1,\ \delta_1$ are given in (43) and (46), respectively. The procedure allows determining the Bayes estimate of SI according to the properties of the Negative Log Gamma function:

$$S^0 = E[S|D] = \frac{1}{(1+\delta_1)^{r_1}} \tag{54}$$

When the relation (47) is not satisfied, numerical algorithms such as those discussed in [60] can be assumed. It is useful to note that both the posterior PDF $g(s\,|\,D)$ and S° are dependent on time horizon τ, since the SI is a function of τ. As clear from the previous relations, this dependency is represented by δ_1. For the sake of finding evidence of the efficiency of the Bayesian estimator in addition to the illustration of its performances, numerical experiments have been carried out using Monte Carlo simulation in [60], particularly focusing on:

1. Evaluating the Mean Square Error of the Bayes estimator;
2. Comparing Bayesian estimates with the classical ones, particularly with the most used Maximum Likelihood (ML) estimates.

Various sample sizes and input data were considered. SI data were produced from the assumed prior PDF on (w, λ), while:

- Data of the observed number of gusts in a given time interval τ were produced using a Poisson process of mean frequency λ (which is randomly produced from the prior PDF) in the interval $(0, \tau)$;
- Data of the observed exceedance number m were produced by a Binomial RV with parameters (n, w), being also w randomly produced using the prior PDF.

As an example taken from [60], the prior data, supposed to be deduced from past observations in this field and relevant to an extreme WS value m^* = 20 m/s, are chosen in the following way:

- λ has a Gamma PDF with μ = 11.0 and σ = 0.22 (year^{-1});
- w has a Beta PDF with μ = 0.02 and σ = 0.0275 (w is a unitless parameter, being a probability).

The choice of the two parameters (n_0, δ_0) of the prior Gamma PDF, and those (r_0, s_0) of the prior Beta PDF are easily obtained by inverting the relations (see Appendices A–F) between the mean and variance of such priors and the relevant parameters.

For each sample size n, a number of $N = 10^4$ replications has been performed in which the above RV λ and w were generated according to the above PDF, and the Bayes estimate of S was deduced. In particular, the results for small (n = 1, n = 2, n = 5) or moderate (n = 15, n = 20, n = 40) sample sizes are reported in the paper in terms of the classical indices [60]:

MSEB: Mean Square Error of the Bayes estimator;

MSEL: Mean Square Error of the ML estimator;

REFF = MSEL/MSEB.

The "REFF" index represents the relative efficiency of the Bayes estimator with respect to the ML estimator. The Mean Square Errors have been obtained at the end of each simulation as the averages over the N sampled estimator's square errors. The numerical results show that the efficiency of the Bayesian estimation increases when the amount of data is small. Moreover, the Bayesian estimation performs much better than the classic one, as shown in the REFF values, when enough data are used.

4.3. A Bayesian Estimation Method in the Parametric Approach

It is recalled that the following Unsafety index (UI), U(τ,m*)—which is a stochastic process—has been deduced in the POT framework:

$$U = U(\tau, m^*) = 1 - e^{-\lambda\tau[1-\varphi(m^*)]} \quad (55)$$

U(τ) expresses the probability that a limit value m^* is exceeded at least once over $(0,\tau)$ by part of the EWS process.

In the parametric approach, a parametric form is assumed on the CDF φ(x). Here, the illustration of a Bayes methodology reported in [60] and also some results, is performed with reference to an Inverse Rayleigh Distribution (IRD) for the above CDF and PDF reported respectively in (56) and (57):

$$f(x|a) = 2ax^{-3}e^{-\left(\frac{\alpha}{x^2}\right)}, \quad x > 0 \quad (56)$$

$$\varphi(z|a) = e^{-\left(\frac{\alpha}{z^2}\right)} \quad (57)$$

In which a is the scale parameter and z is a generic WS value in which the CDF and the PDF are evaluated. The IRD has been demonstrated to constitute a valid model to fit EWS values [60] but is here adopted only to show the feasibility of the proposed Bayesian estimation method. In Figure 1, as one single (for the sake of brevity) indicative but not exhaustive example, three typical IRD PDF (Figure 1a) and CDF (Figure 1b) curves corresponding to the analytical expressions of (56) and (57) are shown, characterized by three different values of the parameter α (50, 100, and 200); it is easily shown [60] that such values of α correspond to the following median values (50th percentile) of the WS RV: 8.5, 12, and 17 m/s, respectively.

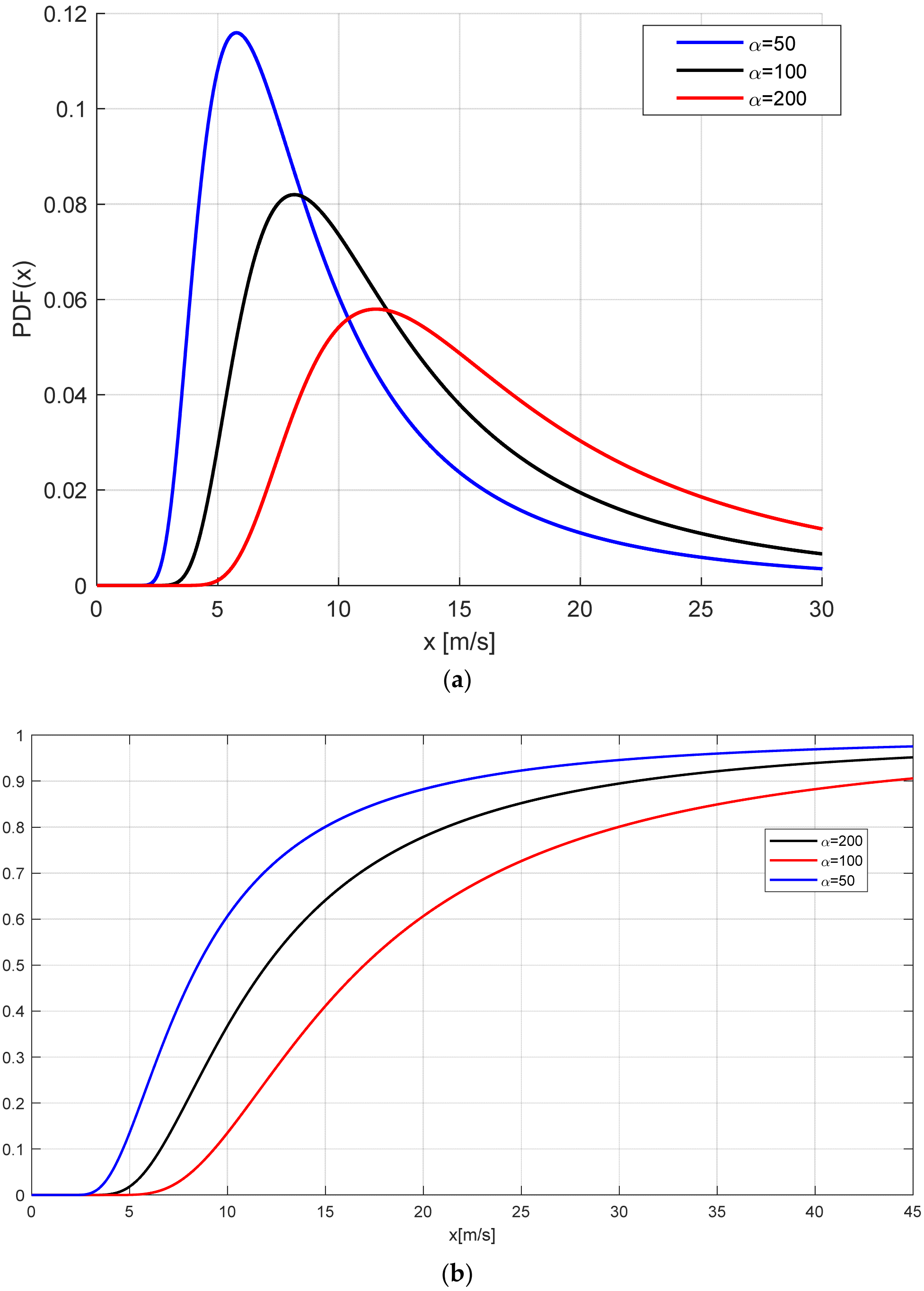

Figure 1. Various inverse Rayleigh PDFs (**a**) and cumulative distribution functions (CDFs) (**b**) with three different values of the parameter α.

For the sake of numerical example, Table 1 lists values of the UI that correspond to various values (φ, α), in correspondence of a time horizon of 1 year and a WS threshold value of 26 m/s. Only a few hints of the classical ML estimation method of the IRD distribution are illustrated with additional discussion on the identification of the model. The MLE φ' of the Poisson frequency φ was already discussed in a previous section. Then, consider the MLE of the CDF $\varphi\ (z)$, which appears in (56) for an IRD model. First, let us consider the MLE of parameter α of the IRD based on a given sample ($X_1, X_2, \ldots\ X_N$) of observed WS values. By assigning a zero value to the partial derivative of the likelihood function, the MLE of this parameter is obtained as follows:

$$\alpha' = \frac{N}{\sum_{i=1}^{N} x_k^{-2}} \tag{58}$$

Table 1. Some values of the UI corresponding to nine couples of values (φ, α), t = 1 year, and m* = 26 m/s.

φ (Year^{-1})	α (m^2/s^2)		
	50	100	200
2	0.133	0.240	0.401
4	0.248	0.423	0.641
8	0.435	0.667	0.871

Then, as the MLE is not variant, the MLE of the CDF $\varphi(z)$ that appears in (57) is given by $e^{-(\frac{\alpha^*}{x^2})}$, so that the ML of the UI is expressed by:

$$U' = 1 - e^{-\lambda' t[1 - e^{-(\frac{\alpha'}{m^{*2}})}]} \tag{59}$$

Then, let us illustrate a possible Bayes Inference Method proposed in [27] for the UI, which depends on the IRD Model. The model implies, apart from the use of the already illustrated Gamma prior PDF for the parameter φ, estimating the parameter α in the expression of the IRD CDF. In this case, two methods were investigated in [27] to set appropriate prior distributions:

(1) Gamma prior conjugate distribution on α (see Appendix F);
(2) Negative Exponential Beta distribution (NEB) (not conjugate distribution) on α (see Appendix E). Its advantage is that it implies a Beta distribution on the already introduced parameter w, so showing such similarity with the previous nonparametric approach.

First, the Gamma prior distribution is discussed here, although briefly (since it is already reported in the literature on the IRD model). The RV $S = 1 - e^{-(\frac{\alpha}{z^2})}$ has a prior PDF which can be easily assessed in terms of the Negative Log-Gamma (NLG) PDF (Appendix C). Indeed, since the RV α has a gamma PDF with parameters (ν_0, δ_0), it can be verified that also the RV

$$\beta = \frac{\alpha}{m^{*2}} \tag{60}$$

has a prior gamma PDF, with parameters ν and δ expressed by $\nu = \nu_0$ and $\delta = \delta_0/m^{*2}$.

Thus, defining the NLG model (Appendix C), U = 1 − S = $e^{-\beta}$ has an NLG prior PDF, hereinafter denoted by g(u):

$$g(u) = \frac{1}{\Gamma(\nu)\delta^{\nu}} u^{(\frac{1}{\delta}-1)} (-log u)^{\nu-1}, \quad 0 \leq u \leq 1 \tag{61}$$

It is worth mentioning that the CDF and other parameters (e.g., quantiles and moments) are also RV, described by appropriate distributions. In [27], it is deduced that prior information available in terms of quantiles (e.g., the median value or 0.50-quantile) can be transformed into prior information on the RV α.

Furthermore, the peculiar method suggested here is that S, an RV that falls in the range (0,1), is assumed to be a Beta RV, implying a Negative Exponential Beta (NEB) PDF of the parameter α, as can be seen from Appendix F, since

$$\alpha = [-log(\varphi)]z^2 \tag{62}$$

Assuming an NEB prior PDF for α implies that the IRD CDF, i.e., the negative exponential function of α in (57), has a Beta distribution. The advantages of the latter assumption have been discussed in the previous section. In this case, the Bayes Method of Inference requires numerical methods which are, as might be expected, very similar to those illustrated in the nonparametric approach [87]. Also, in the parametric approach described here, the numerical evidence of the efficiency of the Bayesian estimators has been proved in [27] by means of large sets of numerical experiments, by means of Monte Carlo simulation in [4], using both Gamma and NEB priors distribution on α. They were conducted for various sample sizes and various input data values. Also, in this case, the results were very satisfactory. In summary, the relative efficiency with respect to the ML estimate—as measured by REFF—is always greater than 1; in particular, as typically happens [88], it is much greater for samples with small sizes, whereby the performance of the ML estimates is definitively better using the Bayes ones.

5. Conclusions

The present review illustrates the modeling and the estimation of the extreme wind speed (EWS), especially related to safety and structural problems, safety being defined as the probability that EWS is lower than a prefixed extreme value which might be dangerous for windfarm structures. The aim is to evaluate both the wind power production and the structural safety of the installations. This is clearly a basic topic to be addressed both in the planning and in the operation stage of a wind farm and has been tackled here using the so-called risk index. Such risk index is deemed to be an important tool in helping a decision-making procedure for assessing if a wind farm is worth being installed and whether its operation will be safe or reliable enough in time. Many parametric extreme wind speeds models have been investigated, including the Weibull, Inverse Weibull, Generalized Extreme Value distribution, and more. Moreover, some hints have been devoted to nonparametric models, and some details on the POT methodology have been illustrated in the framework of stochastic process theory. Subsequently, a detailed review has been presented of the Bayesian methods of wind speed estimation using both the parametric and nonparametric models in the framework of the POT methodology. The Bayesian method showed an efficient estimation even when a large dataset was not available, as required by the classical method. The Bayesian method showed robust estimation with respect to the prior hypotheses of the PDF. The latter method does not require the estimation of many parameters as the more traditionally adopted "Generalized Extreme Value" distribution, but only two parameters.

Author Contributions: Conceptualization, E.C. and G.M.; methodology, B.D. and F.D.A.; resources, E.C. and B.D.; writing—original draft preparation, E.C. and B.D.; writing—review and editing, G.M. and F.D.A.; visualization, B.D. and F.D.A.; supervision, E.C. and G.M.; project administration, G.M.; funding acquisition, G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Beta Distribution

Let Y be an RV assuming values in the (0,1) interval. It is said to possess a Beta Distribution if its PDF, with argument x, with positive parameters (r,s) expressed by:

$$f(y;r,r) = betapdf(y;rs) = \begin{cases} \Gamma(r+s)\frac{y^{r-1}(1-y)^{s-1}}{\Gamma(r)\Gamma(s)}, & 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

The mean value and variance of the Beta PDF are given by:

$$\mu_B = \frac{r}{r+s} \quad \sigma_B^2 = \mu_B^2 \left\{ \frac{s}{r(r+s+1)} \right\}$$

Appendix B. The Gamma Distribution

The Gamma PDF with positive parameters v and δ (shape and scale parameter, respectively) is expressed by:

$$f(x) = \frac{1}{\delta^\nu \Gamma(\nu)} x^{\nu-1} exp\left(-\frac{x}{\delta}\right)$$

where $\Gamma(\nu)$ is the "Euler-Gamma" Function evaluated at ν. The mean value and variance of the distribution are:

$$E[X] = \nu\delta \quad Var[X] = \nu\delta^2$$

Appendix C. The Negative Log-Gamma (NLG) Distribution

Let X be a Gamma RV with parameters v and δ (as in Appendix B). If Y = exp(−Y), then Y is said to possess a Negative-Log Gamma (NLG) distribution, so that the PDF of this NLG RV is:

$$f(y) = \frac{1}{\delta^\nu \Gamma(\nu)} y^{[(\frac{1}{\delta})-1]} (-logy)^{\nu-1},\ 0 \le y \le 1$$

with v and δ being the above Gamma PDF parameters.

The NLG mean value and variance, and are, respectively, equal to the following:

$$\mu_{NLG} = \frac{1}{(1+\delta)^\nu};\ \sigma^2{}_{NLG} = \frac{1}{(1+2\delta)^\nu} - \frac{1}{(1+\delta)^{2\nu}}$$

The kth moment of the NLGD PDF is as follows:

$$\mu_k = E\left[Y^k\right] = \frac{1}{(1+k\delta)^\nu}$$

Appendix D

Assuming that the sampling plan (which produces the observed data from system operation) consists of the number $N(\tau)$ of gusts during a given observation time τ, and the number M of WS, which exceeds the limit value m*. $N(\tau)$ is a Poisson RV with mean $\phi\tau$, and—for the assumed independence of the RV, M is—conditional to the observed number, say n, of Gust Amplitude—a Binomial RV with pr. of "success" (exceedance) w in n independent proofs.

So the Likelihood Function of the data D, once observed n gusts and m exceedances, is given by:

$$L(D|Q,\varphi) = P\left[(N(\tau) = n) \cap (M = m)\right] = P[(N(\tau) = n)]P(M = m)|N(\tau) = n)]$$

with:

$$\text{P}[N(\tau) = n] = \exp(-\varphi\tau) \cdot \frac{(\varphi\tau)^n}{n!} \quad n = 0,\ 1, \ldots\ldots\ldots, \infty$$

$$P[(M = m)|(N(\tau) = n) = \binom{n}{m} w^m (1-w)^{n-r}, \quad m \in \{0,\ 1, \ldots\ldots, n\}$$

The above relations are justified by the Poisson hypothesis for the number of gusts, and the Binomial distribution with parameters (w,m) for the number of exceedances, w being the probability of a single exceedance.

Appendix E. The Gamma as a Prior Conjugate Distribution for the IRD Model

Let $(x_1, x_2, \ldots, x_n)$ be a random sample generated by an IRD with parameter α. Then, each PDF related to the generic value x_j has the following expression:

$$f\left(x_j|\alpha\right) = 2\alpha x_j{}^{-3} e^{-\left(\frac{\alpha}{x_j{}^2}\right)},\ x_j > 0$$

Let the prior PDF on α be a gamma PDF with parameters ν_0 and δ_0:

$$p(\alpha; \nu_0, \delta_0) = \frac{a^{\nu_0 - 1}}{\delta_0{}^{\nu_0} \Gamma(\nu_0)} e^{-\left(\frac{a}{\delta_0}\right)}$$

The likelihood function of the sample, i.e., of the data $D = x$, is as follows:

$$L(D|\alpha) = \prod_{j=1}^{n} f_j\left(x_j\right) = \frac{(2\alpha)^n}{\prod_{j=1}^{n} x_j^3} e^{\left(-\alpha \sum_{j=1}^{n} \frac{1}{x_j^2}\right)}$$

Therefore, denoting by Π and Σ, the following statistics (functions of the data D)

$$\Pi = \prod_{j=1}^{n} x_j^3,\ \ \Sigma = \sum_{j=1}^{n} \frac{1}{x_j^2}$$

and denoting by K, an opportune constant, the posterior PDF of α is expressed by the following equation:

$$q(\alpha|D) = \frac{K 2^n \alpha^{\nu_0 + n - 1}}{\Gamma(\nu_0)\Pi} e^{-\alpha\left(\frac{1}{\delta_0} + \Sigma\right)}$$

Comparing the latter with the expression of a gamma PDF, it is apparent that also the posterior PDF is a gamma PDF:

$$q(\alpha|D) = \text{gammapdf}(\alpha; \nu_1, \delta_1)$$

with updated parameters:

$$\begin{cases} \nu_1 = \nu_0 + n & \text{(Shape parameter)} \\ \delta_1 = \frac{\delta_0}{1 + \delta_0 \Sigma} & \text{(Scale parameter)} \end{cases}$$

Appendix F. The Negative Exponential Beta (NEB) Distributions

As already reported, an RV Y—assuming values in (0,1)—has a Beta distribution if its PDF is expressible as follows:

$$f(y; p, q) = betapdf(y; p, q) = \begin{cases} \Gamma(p+q) \frac{y^{p-1}(1-y)^{q-1}}{\Gamma(p)\Gamma(q)}, & 0 < y < 1 \\ 0 & elsewhere \end{cases}$$

where p and q are positive shape parameters. In particular, if $p = q = 1$, a uniform distribution over (0,1) is obtained. The mean value and the variance of the beta distribution are given by the following:

$$\mu = \frac{p}{p+q};\ \sigma^2 = \mu^2 \frac{q}{p(p+q+1)}$$

It can be noticed that if Y is a beta (p,q) RV, then the following "complementary beta" RV X = 1 − Y is a beta (q,p) RV. Let Y be a beta(a,b) RV. The PDF of the transformed RV X = −log(Y), here denoted as a Negative Exponential Beta (NEB) PDF, by well-known rules of RV transformations, is as follows:

$$g_k(x) = negebpdf(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} e^{-ax}\left(1 - e^{-x}\right)^{(b-1)},\ 0 < x < \infty$$

For another perspective, the RV $Y = e^{-X}$ has a beta(a,b) distribution if and only its inverse function $X = -\log(Y)$ has an NEB (*a*,*b*) PDF. The evaluation of the mean and variance of this model is not easy; however, since $X = -\log(R)$ is a decreasing function of Y, one can calculate the *p*-quantile of X, X_p, using the q-quantile of Y, where $q = 1 - p$:

$$X_p = -\log Y_{(1-p)}$$

References

1. Shi, H.; Dong, Z.; Xiao, N.; Huang, Q. Wind speed distributions used in wind energy assessment: A review. *Front. Energy Res.* **2021**, *9*, 769920. [CrossRef]
2. Manwell, J.F.; McGowan, J.G.; Rogers, A.L. *Wind Energy Explained: Theory, Design and Application*; John Wiley & Sons: Chichester, UK, 2010.
3. Jain, P. *Wind Energy Engineering*; McGraw-Hill: New York, NY, USA, 2011.
4. Chiodo, E.; Mazzanti, G.; Karimian, M. Bayes estimation of Inverse Weibull distribution for extreme wind speed prediction. In Proceedings of the International Conference on Clean Electrical Power (ICCEP), Taormina, Italy, 16–18 June 2015.
5. Pryor, S.C.; Barthelmie, R.J.; Bukovsky, M.S.; Leung, L.R.; Sakaguchi, K. Climate change impacts on wind power generation. *Nat. Rev. Earth Environ.* **2020**, *1*, 627–643. [CrossRef]
6. Cheng, E.; Yeung, C. Generalized extreme gust wind speeds distributions. *J. Wind Eng. Ind. Aerodyn.* **2002**, *90*, 1657–1669. [CrossRef]
7. Letcher, T. *Wind Energy Engineering: A Handbook for Onshore and Offshore Wind Turbines*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2017.
8. Zhang, H.; Yu, Y.-J.; Liu, Z.-Y. Study on the maximum entropy principle applied to the annual wind speed probability distribution: A case study for observations of intertidal zone anemometer towers of rudong in east china sea. *Appl. Energ.* **2014**, *114*, 931–938. [CrossRef]
9. Yan, J.; Liu, Y.; Han, S.; Gu, C.; Li, F. A Robust Probabilistic Wind Power Forecasting Method Considering Wind Scenarios. In Proceedings of the 3rd Renewable Power Generation Conference, Naples, Italy, 24–25 September 2014.
10. Abdulali, B.A.A.; Abu Bakar, M.A.; Ibrahim, K.; Ariff, N.B.M. Extreme value distributions: An overview of estimation and simulation. *J. Probab. Stat.* **2022**, *2022*, 5449751. [CrossRef]
11. Fisher, R.A.; Tippett, L.H.C. Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Philos. Soc.* **1928**, *24*, 180–190. [CrossRef]
12. Gumbel, E.J. *Statistics of Extremes*; Columbia University Press: New York, NY, USA, 1958.
13. Perrin, O.; Rootzén, H.; Taesler, R. A discussion of statistical methods used to estimate extreme wind speeds. *Theor. Appl. Clim.* **2006**, *85*, 203–215. [CrossRef]
14. Castillo, E.; Hadi, A.S.; Balakrishnan, N.; Sarabia, J.M. *Extreme Value and Related Models with Applications in Engineering and Science*; Wiley: Hoboken, NJ, USA, 2005.
15. An, Y.; Pandey, M.D. A comparison of methods of extreme wind speed estimation. *J. Wind Eng. Ind. Aerodyn.* **2005**, *93*, 535–545. [CrossRef]
16. De Oliveira, M.M.F.; Ebecken, N.F.F.; de Oliveira, J.L.F.; Gilleland, E. Generalized extreme wind speed distributions in South America over the Atlantic Ocean region. *J. Theor. Appl. Climatol.* **2011**, *104*, 377–385. [CrossRef]
17. Kunz, M.; Mohr, S.; Rauthe, M.; Lux, R.; Kottmeier, C.h. Assessment of extreme wind speeds from regional climate models—part 1: Estimation of return values and their evaluation. *J. Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 907–922. [CrossRef]
18. Soukissian, T.H.; Christos, T. The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction. *Nat. Hazards* **2015**, *78*, 1777–1809. [CrossRef]
19. Palutikof, J.P.; Brabson, B.B.; Lister, D.H.; Adcock, S.T. A review of methods to calculate extreme wind speeds. *Meteorological* **1999**, *6*, 119–132. [CrossRef]
20. De Fondeville, R.; Davison, A.C. Functional peaks-over-threshold analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2022**, *84*, 1392–1422. [CrossRef]
21. Balkema, A.A.; de Haan, L. Residual life time at great age. *Ann. Probab.* **1974**, *2*, 792–804. [CrossRef]
22. Pickands, J. Statistical inference using extreme order statistics. *Ann. Stat.* **1975**, *3*, 119–131.
23. Leadbetter, M.R. On a basis for 'Peaks over Threshold' modeling. *Stat. Probab. Lett.* **1991**, *12*, 357–362. [CrossRef]
24. Davison, A.C. *Modelling Excesses over High Thresholds, with an Application*; Statistical Extremes and Applications; Reidel: Dordrecht, The Netherlands, 1984; pp. 461–482.
25. Smith, R.L. *Threshold Methods for Sample Extremes*; Statistical Extremes and Applications; Reidel: Dordrecht, The Netherlands, 1984; pp. 621–638.
26. Davison, A.C.; Smith, R.L. Models for Exceedances over High Thresholds (with discussion). *J. R. Stat. Soc. Ser. B* **1990**, *52*, 393–442.
27. Chiodo, E.; Mazzanti, G.; Karimian, M.; Zoh, R. Comparison of two different estimation methods of wind speed extreme values. In Proceedings of the International Conference on Clean Electrical Power (ICCEP), Taormina, Italy, 16–18 June 2015.
28. Coles, S.G.; Elwyn, A.P. Bayesian methods in extreme value modelling: A review and new developments. *Int. Stat. Rev.* **1996**, *64*, 119–136. [CrossRef]

29. Li, G.; Jing, S. Applications of Bayesian methods in wind energy conversion systems. *Renew. Energy* **2012**, *43*, 1–8. [CrossRef]
30. Mazzanti, G.; Diban, B.; Chiodo, E.; De Falco, P.; Noia, L.P.D. Forecasting the Reliability of Components Subjected to Harmonics Generated by Power Electronic Converters. *Electronics* **2020**, *9*, 1266. [CrossRef]
31. De Angelis, F.; Cancellara, D.; Grassia, L.; D'Amore, A. The influence of loading rates on hardening effects in elasto/viscoplastic strain-hardening materials. *Mech. Time-Depend. Mater.* **2018**, *22*, 533–551. [CrossRef]
32. De Angelis, F.; Cancellara, D. Dynamic analysis and vulnerability reduction of asymmetric structures: Fixed base vs base isolated system. *Compos. Struct.* **2019**, *219*, 203–220. [CrossRef]
33. De Angelis, F.; Cancellara, D. Multifield variational principles and computational aspects in rate plasticity. *Comput. Struct.* **2017**, *180*, 27–39. [CrossRef]
34. Leadbetter, M.R.; Lindgren, G.; Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*; Springer Verlag: New York, NY, USA, 1983.
35. Galambos, J.; Lechner, J.; Simiu, E. *Extreme Value Theory and Applications*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994; pp. 139–158.
36. Gross, J.; Heckert, A.; Lechner, J.; Simiu, E. Novel Extreme Value Estimation Procedures: Application to Extreme Wind Data. In *Extreme Value Theory and Applications*; Springer: Boston, MA, USA, 1994.
37. Chiodo, E.; Noia, L.P.D. Stochastic Extreme Wind Speed Modeling and Bayes Estimation under the Inverse Rayleigh Distribution. *Appl. Sci.* **2020**, *10*, 5643. [CrossRef]
38. Fréchet, M. Sur. La loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math* **1927**, *6*, 93.
39. Sarkar, A.; Deep, S.; Datta, D.; Vijaywargiya, A.; Roy, R.; Phanikanth, V.S. Weibull and Generalized Extreme Value Distributions for Wind Speed Data Analysis of Some Locations in India. *KSCE J. Civ. Eng.* **2019**, *23*, 3476–3492. [CrossRef]
40. Shu, Z.R.; Li, Q.S.; Chan, P.W. Statistical Analysis of Wind Characteristics and Wind Energy Potential in Hong Kong. *Energy Convers. Manag.* **2015**, *101*, 644–657. [CrossRef]
41. Wais, P. Two and Three-Parameter Weibull Distribution in Available Wind Power Analysis. *Renew. Energy* **2017**, *103*, 15–29. [CrossRef]
42. Kiss, P.; Jánosi, I.M. Comprehensive Empirical Analysis of ERA-40 Surface Wind Speed Distribution over Europe. *Energy Convers. Manag.* **2008**, *49*, 2142–2151. [CrossRef]
43. Bilir, L.; İmir, M.; Devrim, Y.; Albostan, A. An Investigation on Wind Energy Potential and Small Scale Wind Turbine Performance at İncek Region—Ankara, Turkey. *Energy Convers. Manag.* **2015**, *103*, 910–923. [CrossRef]
44. Celik, A.N. A Simplified Model for Estimating Yearly Wind Fraction in Hybrid-Wind Energy Systems. *Renew. Energy* **2006**, *31*, 105–118. [CrossRef]
45. Alves, M.F.; Neves, C. *Extreme Value Distributions*; Springer: Berlin, Germany, 2011.
46. Pinheiro, E.C.; Ferrari, S.L.P. A comparative review of generalizations of the Gumbel extreme value distribution with an application to wind speed data. *J. Stat. Comput. Simul.* **2016**, *86*, 2241–2261. [CrossRef]
47. Torrielli, A.; Repetto, M.P.; Solari, G. Extreme Wind Speeds from Long-Term Synthetic Records. *J. Wind Eng. Ind. Aerod.* **2013**, *115*, 22–38. [CrossRef]
48. Kang, D.; Ko, K.; Huh, J. Determination of Extreme Wind Values Using the Gumbel Distribution. *Energy* **2015**, *86*, 51–58. [CrossRef]
49. Baran, S.; Szokol, P.; Szabó, M. Truncated generalized extreme value distribution-based ensemble model output statistics model for calibration of wind speed ensemble forecasts. *Environmetrics* **2021**, *32*, 2678. [CrossRef]
50. Holmes, J.D.; Moriarty, W.W. Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *J. Wind Eng. Ind. Aerodyn.* **1999**, *83*, 1–10. [CrossRef]
51. Brabson, B.B.; Palutikof, J.P. Tests of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds. *J. Appl. Meteor. Climatol.* **2000**, *39*, 1627–1640. [CrossRef]
52. Samal, R.K. Assessment of Wind Energy Potential Using Reanalysis Data: A Comparison with Mast Measurements. *J. Clean. Prod.* **2021**, *313*, 127933. [CrossRef]
53. Raynal, J.; Guevara, J. Maximum Likelihood Estimators for the Two Populations Gumbel Distribution. *Hydrol. Sci. Technol.* **1997**, *13*, 47–56.
54. Raynal, J.; Santillan, O. Maximum Likelihood Estimators of the Parameters of the Mixed GEV Distribution. In *Proceedings of Congreso Nacional de Hidráulica*; Asociación Mexicana de Hidráulica AMH: Querétaro, Mexico, 1986.
55. Escalante, C. Application of Bivariate Extreme Value Distribution to Flood Frequency Analysis: A Case Study of Northwestern Mexico. *Nat. Hazards* **2006**, *42*, 37–46. [CrossRef]
56. Escalante, C. Estimation of Extreme Wind Speeds by Using Mixed Distributions. *Ing. Investig. Y Tecnol.* **2013**, *14*, 153–162.
57. Rossi, F.; Fiorentino, M.; Versace, P. Two-Component Extreme Value Distribution for Flood Frequency Analysis. *Water Resour. Res.* **1984**, *20*, 847–856. [CrossRef]
58. Hu, Q.; Wang, Y.; Xie, Z.; Zhu, P.; Yu, D. On Estimating Uncertainty of Wind Energy with Mixture of Distributions. *Energy* **2016**, *112*, 935–962. [CrossRef]
59. Ouarda, T.B.M.J.; Charron, C. On the Mixture of Wind Speed Distribution in a Nordic Region. *Energy Convers. Manag.* **2018**, *174*, 33–44. [CrossRef]

60. Chiodo, E.; Fantauzzi, M.; Mazzanti, G. The Compound Inverse Rayleigh as an Extreme Wind Speed Distribution and Its Bayes Estimation. *Energies* **2022**, *15*, 861. [CrossRef]
61. Han, Q.; Hao, Z.; Hu, T.; Chu, F. Non-parametric models for joint probabilistic distributions of wind speed and direction data. *Renew. Energy* **2018**, *126*, 1032–1042. [CrossRef]
62. Zhou, J.; Erdem, E.; Li, G.; Shi, J. Comprehensive Evaluation of Wind Speed Distribution Models: A Case Study for North Dakota Sites. *Energy Convers. Manag.* **2010**, *51*, 1449–1458. [CrossRef]
63. McElroy, T.; Penny, R. Maximum entropy extreme-value seasonal adjustment. *Aust. N. Z. J. Stat.* **2019**, *61*, 152–174. [CrossRef]
64. Lechner, J.A.; Simiu, E.; Heckert, N.A. Assessment of 'peaks over threshold methods for estimating extreme value distribution tails. *Struct. Saf.* **1993**, *12*, 305–314. [CrossRef]
65. Beirlant, J.; Goegebeur, Y.; Segers, J.J.; Teugels, J. *Statistics of Extremes: Theory and Applications*; Wiley: Chichester, UK, 2004.
66. Powell, M.D. Wind measurement and archival under the Automated Surface Observing System (ASOS): User concerns and opportunity for improvement. *Bull. Am. Meteorol. Soc.* **1993**, *74*, 615–624. [CrossRef]
67. Simiu, E.; Heckert, N.A. Extreme wind distribution tails: A "peaks over threshold" approach. *J. Struct. Eng.* **1996**, *122*, 539–547. [CrossRef]
68. Cook, N.J. Towards better estimation of extreme winds. *J. Wind Eng. Ind. Aerodyn.* **1982**, *9*, 295–323. [CrossRef]
69. Ross, W.H. A peaks-over-threshold analysis of extreme wind speeds. *Can. J. Stat.* **1987**, *15*, 328–335. [CrossRef]
70. Coles, S.; Bawa, J.; Trenner, L.; Dorazio, P. *An Introduction to Statistical Modeling of Extreme Values*; Springer: Berlin/Heidelberg, Germany, 2001; p. 208.
71. Ceppi, P.; Della-Marta, P.M.; Appenzeller, C. Extreme value analysis of wind speed observations over Switzerland. *Arbeitsberichte der MeteoSchweiz* **2008**, *219*, 48.
72. Folgueras, P.; Solari, S.; Losada, M.Á. The selection of directional sectors for the analysis of extreme wind speed. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 221–236. [CrossRef]
73. Fawcett, L.; Walshaw, D. Bayesian inference for clustered extremes. *Extremes* **2008**, *11*, 217–233. [CrossRef]
74. Bernardo, J.M.; Smith, A.F.M. *Bayesian Theory*; Wiley: Hoboken, NJ, USA, 2000.
75. Cowles, M.K. *Applied Bayesian Statistics*; Springer: Berlin/Heidelberg, Germany, 2013.
76. Press, S.J. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*, 2nd ed.; Wiley: New York, NY, USA, 2002.
77. Diriba, T.A.; Debusho, L.K.; Botai, J.; Hassen, A. Bayesian modelling of extreme wind speed at Cape Town, South Africa. *Env. Ecol. Stat.* **2017**, *24*, 243–267. [CrossRef]
78. Sloughter, J.; McLean, T.G.; Raftery, A.E. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Am. Stat. Assoc.* **2010**, *105*, 25–35. [CrossRef]
79. Touqeer, A.; Ahmad, I.; Arshad, I.A.; Almanjahie, I.M. An efficient Bayesian modelling of extreme winds in the favour of energy generation in Pakistan. *Energy Rep.* **2023**, *9*, 2980–2992.
80. Berger, J.O. Bayesian analysis: A look at today and thoughts of tomorrow. *J. Am. Stat. Assoc.* **2000**, *95*, 1269–1276. [CrossRef]
81. Jackman, S. *Bayesian Analysis for the Social Sciences*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
82. Cooley, D.; Naveau, P.; Jomelli, V.; Rabatel, A.; Grancher, D. A Bayesian hierarchical extreme value model for lichenometry. *Env. Off. J. Int. Env. Soc.* **2006**, *17*, 555–574. [CrossRef]
83. Chiodo, E.; Falco, P.D.; Noia, L.P.D.; Mottola, F. Inverse Log-logistic distribution for Extreme Wind Speed modeling: Genesis, identification and Bayes estimation. *AIMS Energy* **2018**, *6*, 926–948. [CrossRef]
84. Chiodo, E.; Mazzanti, G. Theoretical and practical aids for the proper selection of reliability models for power system components. *Int. J. Reliab. Saf.* **2008**, *2*, 99–128. [CrossRef]
85. Guo, Z.; Chi, D.; Wu, J.; Zhang, W. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm. *Energy Convers. Manag.* **2014**, *84*, 140–151. [CrossRef]
86. Casella, G.; Berger, R.L. *Statistical Inference*, 2nd ed.; Duxbury Press: Pacific Grove, CA, USA, 2001.
87. Tanoe, V.; Henderson, S.; Shahirinia, A.; Bina, M.T. Bayesian and non-Bayesian regression analysis applied on wind speed data. *J. Renew. Sustain. Energy* **2021**, *13*, 053303. [CrossRef]
88. Battistelli, L.; Chiodo, E.; Lauria, D. Bayes assessment of photovoltaic inverter system reliability and availability. In Proceedings of the 2010 International Symposium on Power Electronics Electrical Drives Automation and Motion (SPEEDAM), Pisa, Italy, 14–16 June 2010; pp. 1–5.