





Article

On the Hybrid Normal Distribution and Its Application in Fiber Data on the Strength of Glass

Damodaran Santhamani Shibu ¹, Soman Latha Nitin ¹, Christophe Chesneau ^{2,*}, Muhammed Rasheed Irshad ³, Sobhanam Padmini Shibin ⁴ and Radhakumari Maya ¹

¹ Department of Statistics, University College, Thiruvananthapuram 695 034, India

² LMNO, UFR de Sciences, Université de Caen Basse-Normandie, F-14032 Caen, France

³ Department of Statistics, Cochin University of Science and Technology, Cochin 682 022, India

⁴ National Center for Aquatic Animal Hospital, Lakeside Campus, Cochin University of Science and Technology, Cochin 682 022, India

* Correspondence: christophe.chesneau@gmail.com

Abstract: The hybrid normal (HN) distribution is a new generalization of the normal distribution that we introduce and study in this article. Its mathematical foundation is based on the logarithmically transformed version of the famous hybrid log-normal (HLN) distribution, which is an unexplored direction in the literature. We emphasize the applicability of the HN distribution with the aim of fitting versatile data, such as, in this paper, fiber data on the strength of glass. In particular, the unknown parameters are estimated using both Bayesian and maximum likelihood estimation approaches, with Bayesian estimation carried out using the MCMC approach. A thorough simulation study is performed to determine the flexibility of the estimates' performance. The glass fiber data are then analyzed, with an assessment of several existing distributions from the literature used to demonstrate how the HN distribution is relevant in this regard.

Keywords: hybrid log-normal distribution; MCMC estimation; simulation; strength of glass fiber data

MSC: 60E05; 62E10; 62F10



Citation: Shibu, D.S.; Nitin, S.L.; Chesneau, C.; Irshad, M.R.; Shibin, S.P.; Maya, R. On the Hybrid Normal Distribution and Its Application in Fiber Data on the Strength of Glass. *Dynamics* **2022**, *2*, 399–413. <https://doi.org/10.3390/dynamics2040023>

Academic Editor: Christos Volos

Received: 10 October 2022

Accepted: 16 November 2022

Published: 18 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The majority of real-life datasets resulting from complex processes are infrequently characterized and predicted adequately by standard fundamental distributions in practice. The choice of a flexible distribution for data analysis is crucial, as the perfection of results from statistical analysis greatly relies on the model that is adopted. In order to obtain statistical outputs of higher quality and greater accuracy, it is essential to identify more connected distributions. Several continuous distributions are readily accessible in the statistical literature for modelling real data. Among them, the skewed distributions have more strategic significance, and are frequently used in many areas of applied research. For this reason, numerous families of skewed distributions have been proposed. Azzalini (1985) (see [1]) introduced a method for introducing skewness to the normal (N) distribution. It should be noted that while the majority of skewed distributions have the positive real values set as the support, the selected based distribution has the full real line as the support. Subsequently, many authors have continued to develop this concept; for example, see [2,3].

In the same respect, we cite Coles (2001) [4], Castillo et al., (2005) [5], and Ferrari and Pinheiro (2012) [6], all of whom emphasize the Gumbel distribution (defined on the real line) for analyzing real datasets. As a fair competitor, [7] developed the log-Dagum (LD) distribution, which is derived by transforming the Dagum distribution logarithmically. The LD distribution has the feature of having leptokurtic shapes.

In order to create vast classes of standard distributions and pertinent statistical procedures that can be used as models for a variety of real-world problems, much effort has

been put forth throughout the years. Real-world phenomena need to be introduced to new statistical distributions that are defined on the entire real line and have bimodal behaviour for the probability density function (pdf). As a consequence, in this paper, we offer a novel distribution called the hybrid N (HN) distribution, which can be described as a generalized version of the well-known N distribution. It can additionally be represented as the logarithmically transformed version of the hybrid log-N (HLN) distribution, as defined by the cumulative distribution function (cdf) provided in Equation (1). It has the interesting feature of having a bimodal pdf. The leading motivations behind the introduction of this new distribution are:

- (i) to develop a new generalization for the well-known N distribution, which continues to pique the interest of statisticians across the board;
- (ii) as mentioned, to investigate the characteristics of the HLN distribution's logarithmically transformed version, which has surprisingly not received much attention in the statistical literature.

The practical viewpoint is considered as well; we aim to demonstrate how the HN distribution can be concretely used as a model for dealing with the analysis of important data, and show that it outperforms comparable well-known models. We examine an experimental dataset of glass fiber strength from the National Physical Laboratory in England as an application in this research. Because the HN distribution is an extended version of the well-known N distribution, which has numerous applications in physics, including various dynamical systems (see [8,9]), it is quite conceivable to employ the HN distribution for further physical applications.

The remainder of the present article is divided into several sections, and is structured as follows. We introduce the HN distribution and discuss its special cases and moments in Section 2. The various functions and moments related to reliability measures are discussed in Section 3. In Section 4, the unknown parameters of the new model are estimated using efficient procedures, namely, the maximum likelihood (ML) and Bayesian estimation procedures. A simulation study is conducted in Section 5 to analyze the performance and flexibility of the ML estimates (MLEs). Additionally, Section 6 presents a parametric bootstrap simulation technique utilizing these estimates. Section 7 illustrates the application of the HN model based on a real-life dataset. The final concluding remarks are presented in Section 8.

2. The Hybrid Normal Distribution

For fitting data on occupational radiation exposure, Kumazawa and Numakunai [10] introduced a new distribution in 1981 called the HLN distribution. An absolutely continuous random variable (acrv) Y is said to follow an HLN distribution if its cdf is indicated as

$$H_Y(y) = \int_0^y \frac{1}{\sqrt{2\pi}} \left(\frac{\alpha}{u} + \sigma \right) e^{-\frac{1}{2}(\alpha \log u + \sigma u + \mu)^2} du, \quad (1)$$

and the pdf is provided by

$$h_Y(y) = \frac{1}{\sqrt{2\pi}} \left(\frac{\alpha}{y} + \sigma \right) e^{-\frac{1}{2}(\alpha \log y + \sigma y + \mu)^2}, \quad (2)$$

where $y > 0$, $\alpha \geq 0$, $\sigma \geq 0$, $\alpha + \sigma > 0$, and $\mu \in \mathbb{R}$. For $\sigma = 0$, the pdf in Equation (2) reduces to the well-known two-parameter log-N distribution denoted as $\text{LN}(-\mu\alpha^{-1}, \alpha^{-2})$. Although the HLN distribution might seem simple, there do not seem to be many references to it in the statistical literature. A scalar function was defined in [11], showing a hybrid pdf made up of p N and q log-N variants. The hybrid multivariate stochastic differential equation for stand basal area and volume models in a forest stand is presented in [12].

Here, the proposed three-parameter distribution is referred to as the hybrid N distribution, denoted as $\text{HN}(\alpha, \mu, \sigma)$ or simply HN.

Definition 1. If an acrv Y follows the HLN distribution as provided by its cdf and pdf in Equations (1) and (2), respectively, then the acrv $X = \log Y$ is said to follow the HN distribution. The cdf of X is specified by

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} (\alpha + \sigma e^u) e^{-\frac{1}{2}(\alpha u + \sigma e^u + \mu)^2} du,$$

which can be written as

$$F(x) = \Phi(\alpha x + \sigma e^x + \mu), \quad (3)$$

where $\Phi(x)$ is the cdf of the standard N (SN) distribution, conventionally denoted as $N(0, 1)$. Then, the pdf of the acrv X is provided by

$$f(x) = \frac{1}{\sqrt{2\pi}} (\alpha + \sigma e^x) e^{-\frac{1}{2}(\alpha x + \sigma e^x + \mu)^2}, \quad x \in \mathbb{R}, \quad (4)$$

where $\alpha \geq 0$, $\sigma \geq 0$, $\alpha + \sigma > 0$, and $\mu \in \mathbb{R}$.

Two extremely alluring characteristics instantly follow from Definition 1, which are as follows:

Property A: When $\sigma = 0$, the pdf in Equation (4) reduces to the pdf of the N distribution with parameters $-\mu\alpha^{-1}$ and α^{-2} , i.e., $N(-\mu\alpha^{-1}, \alpha^{-2})$, where μ is not the typical location parameter of the N distribution.

Property B: If the acrv X follows the HN distribution, then $Y = e^X$ has the HLN distribution.

Plots of the corresponding cdf and pdf of the HN distribution are shown in Figures 1 and 2. It is possible to see from Figure 2 that the pdf is bimodal in nature, with a possible spike, which is an intriguing feature of this distribution.

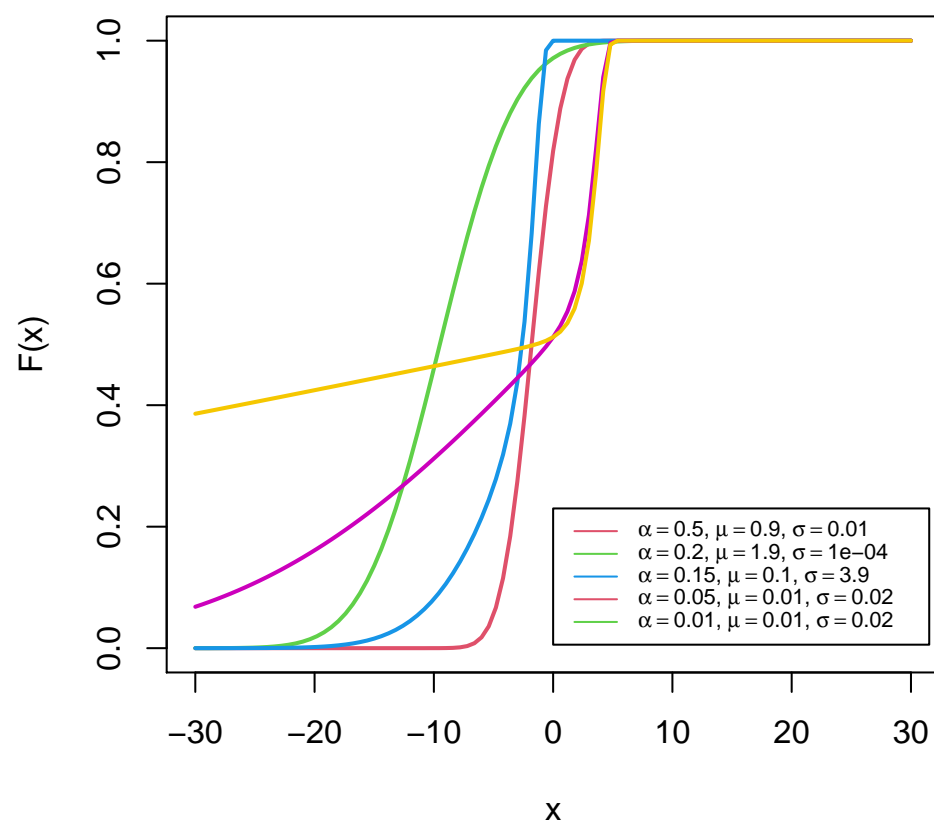


Figure 1. Cdf plots of the HN distribution.

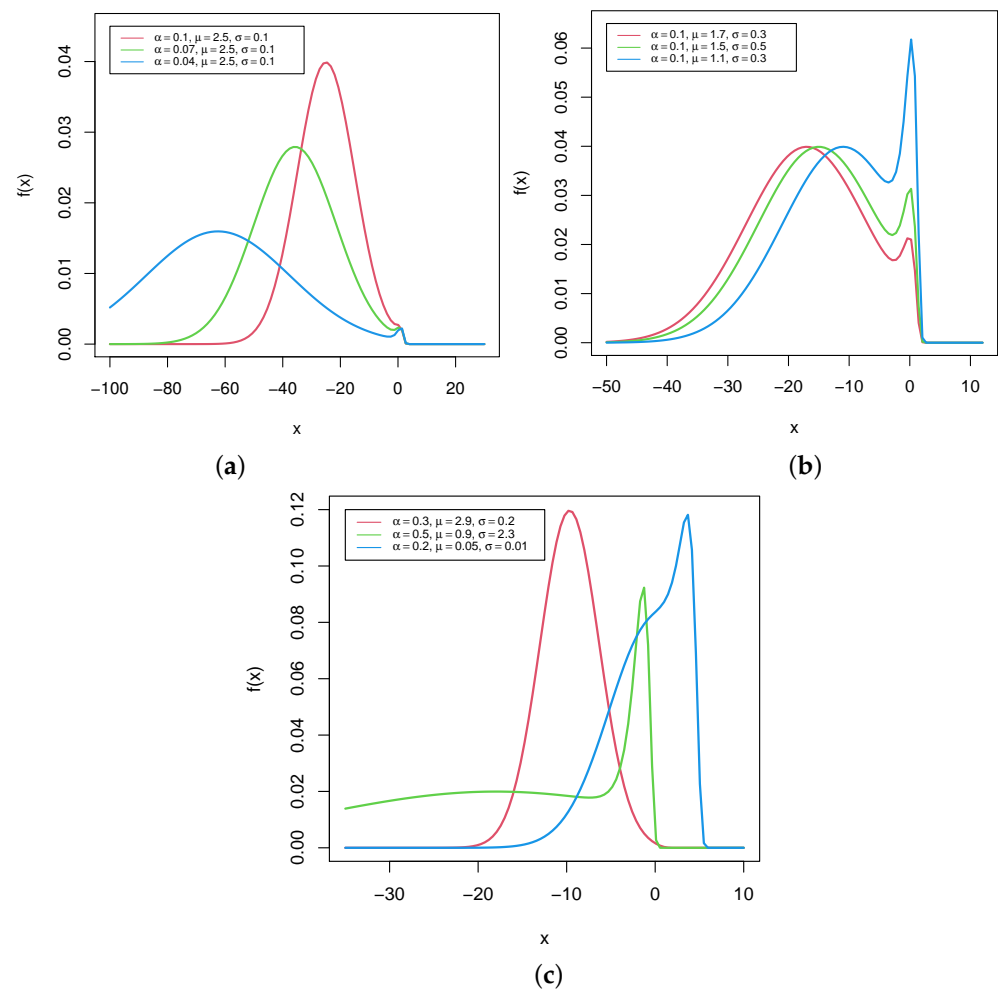


Figure 2. Pdf plots of the HN distribution: (a) bimodal with a dominant bell shape, (b) bimodal with a bell-shape and a spike shape, and (c) original skewness shapes.

Now, the moment generating function of the HN distribution is indicated as

$$M(t) = \frac{I^*(t)}{\sqrt{2\pi}}, \quad (5)$$

where

$$I^*(t) = \int_{-\infty}^{\infty} (\alpha + \sigma e^x) e^{tx - \frac{1}{2}(\alpha x + \sigma e^x + \mu)^2} dx, \quad (6)$$

which is well-defined for all $t \in \mathbb{R}$.

Theorem 1. If X has the $HN(\alpha, \mu, \sigma)$ distribution, then $Y = |X|$ follows the half-HN distribution (HHN) with parameters α, μ , and σ , with the cdf and pdf provided by

$$G(y) = \Phi(\alpha y + \sigma e^y + \mu) - \Phi(-\alpha y + \sigma e^{-y} + \mu), \quad (7)$$

and

$$g(y) = (\alpha + \sigma e^y) \phi(\alpha y + \sigma e^y + \mu) + (\alpha + \sigma e^{-y}) \phi(-\alpha y + \sigma e^{-y} + \mu), \quad (8)$$

respectively, where $y > 0$ and $\phi(x)$ is the pdf of the SN distribution.

Proof. For any $y > 0$, the cdf of Y is specified by

$$\begin{aligned} G(y) &= P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F(y) - F(-y) \\ &= \Phi(\alpha y + \sigma e^y + \mu) - \Phi(-\alpha y + \sigma e^{-y} + \mu). \end{aligned}$$

The pdf of the HHN distribution is provided by

$$\begin{aligned} g(y) &= \frac{d}{dy} G(y) = \frac{d}{dy} \Phi(\alpha y + \sigma e^y + \mu) - \frac{d}{dy} \Phi(-\alpha y + \sigma e^{-y} + \mu) \\ &= (\alpha + \sigma e^y) \phi(\alpha y + \sigma e^y + \mu) + (\alpha + \sigma e^{-y}) \phi(-\alpha y + \sigma e^{-y} + \mu), \end{aligned} \quad (9)$$

which ends the proof. \square

It is significant to mention that the HHN distribution is not addressed in the current statistical literature. In a subsequent publication, we intend to investigate the various mathematical properties of this distribution.

3. Reliability Measures

Identification of a system's important components and estimation of the effects of component failures are the fundamental goals of systems reliability analysis. Therefore, it is essential to derive the functions of the HN distribution's reliability measures.

Assume that X is an acrv with pdf $f(x)$ and cdf $F(x)$. Then, the general formulas for the survival, hazard rate, cumulative hazard rate and reversed hazard rate functions are $S(x) = 1 - F(x)$, $h(x) = f(x)/S(x)$, $R(x) = -\log[S(x)]$, and $r(x) = f(x)/F(x)$, respectively.

3.1. Survival, Hazard Rate, Cumulative Hazard Rate, and Reversed Hazard Rate Functions

The survival function of the HN distribution is indicated as

$$S(x) = 1 - \Phi(\alpha x + \sigma e^x + \mu), \quad (10)$$

where $x \in \mathbb{R}$ (as hereinafter).

Thus, the hazard rate function of the HN distribution is provided as

$$h(x) = \frac{(\alpha + \sigma e^x) e^{-\frac{1}{2}(\alpha x + \sigma e^x + \mu)^2}}{\sqrt{2\pi} [1 - \Phi(\alpha x + \sigma e^x + \mu)]}. \quad (11)$$

The plots in Figure 3 portray the increasing nature of the hazard rate function.

The cumulative hazard rate function of the HN distribution is specified by

$$R(x) = -\log[1 - \Phi(\alpha x + \sigma e^x + \mu)], \quad (12)$$

while the reversed hazard rate function of the HN distribution is expressed as

$$r(x) = \frac{(\alpha + \sigma e^x) e^{-\frac{1}{2}(\alpha x + \sigma e^x + \mu)^2}}{\sqrt{2\pi} \Phi(\alpha x + \sigma e^x + \mu)}. \quad (13)$$

3.2. Conditional Moments

For studying the reliability of a system, it is of great interest to know the conditional moments that are important tools for prediction purposes. If X is an acrv that follows the HN distribution, then its n th conditional moment is provided by

$$E(X^n | X > t) = \frac{1}{\sqrt{2\pi} S(t)} I_1(n, t), \quad (14)$$

where

$$I_1(n, t) = \int_t^\infty x^n (\alpha + \sigma e^x) e^{-\frac{1}{2}(\alpha x + \sigma e^x + \mu)^2} dx. \quad (15)$$

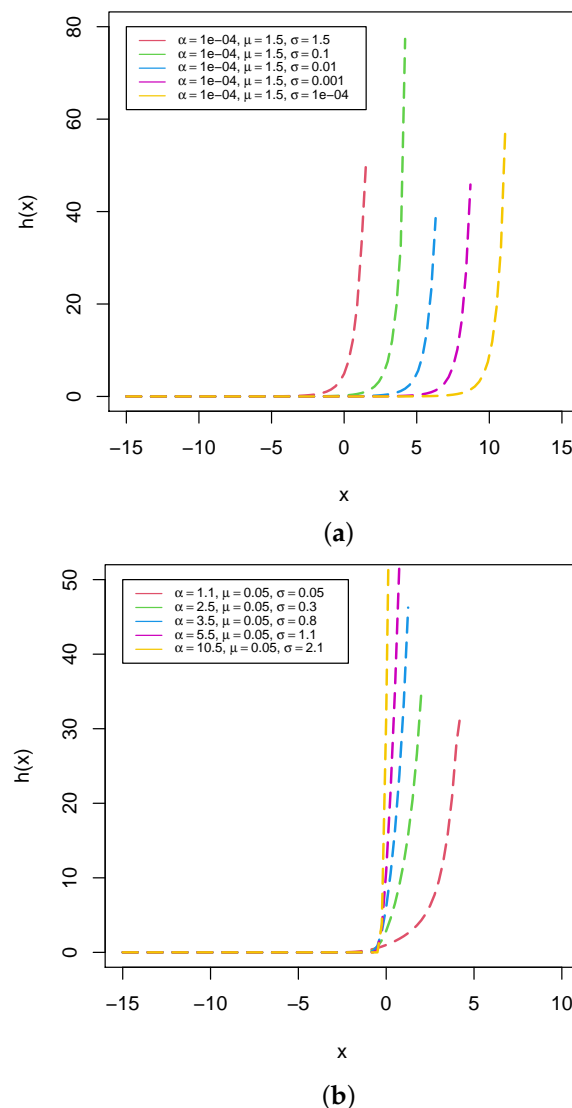


Figure 3. Hazard rate function plots of the HN distribution: (a) moving σ and (b) combined action of α and σ .

A direct application is discussed below. In reliability engineering, biomedical research, and survival analysis, the vitality function is crucial. A component must have low relative vitality to age quickly, whereas high vitality implies comparatively gradual aging over the course of the given period. The vitality function of an acrv X that follows the HN distribution is expressed by the conditional moments provided in Equation (14) for $n = 1$, that is,

$$V(t) = E(X|X > t) = \frac{I_1(1, t)}{\sqrt{2\pi} S(t)}, \quad (16)$$

where $I_1(n, t)$ is presented in Equation (15). For more details on the vitality function, see [13].

3.3. Geometric Vitality Function

The geometric mean of the residual lifetime serves as the foundation for the geometric vitality function. If X is an acrv which represents the lifetime of a component, then the geometric mean of the component that lives up to time t is represented by $\log G(t) =$

$E(\log X|X > t)$, provided that $E(\log X)$ exists. Hence, the geometric vitality function for a non-negative acrv X with pdf $f(t)$ is defined as

$$\log G(t) = \frac{1}{S(t)} \int_t^\infty (\log x) f(x) dx, \quad (17)$$

where $S(t) = P(X > t)$ represents the survival function. For more information on the geometric vitality function, see [14]. Now, the geometric vitality function of an acrv X that follows the HN distribution is provided by

$$\log G(t) = \frac{I_2(t)}{\sqrt{2\pi} S(t)}, \quad (18)$$

where

$$I_2(t) = \int_t^\infty (\log x) (\alpha + \sigma e^x) e^{-\frac{1}{2}(\alpha x + \sigma e^x + \mu)^2} dx. \quad (19)$$

3.4. Moments of Residual Life

The r th order moment of the residual life of an acrv X that follows the HN distribution is obtained as follows:

$$\begin{aligned} \mu_r(t) &= E[(X - t)^r | X > t] \\ &= \frac{1}{S(t)} \int_t^\infty \sum_{i=0}^r \binom{r}{i} (-t)^{r-i} x^i f(x) dx \\ &= \frac{1}{\sqrt{2\pi} S(t)} \sum_{i=0}^r \binom{r}{i} (-t)^{r-i} I_1(i, t), \end{aligned} \quad (20)$$

where $I_1(n, t)$ is provided in Equation (15).

As a special case, the mean residual life (MRL) function is expressed as

$$\mu_1(t) = V(t) - t = \frac{I_1(1, t)}{\sqrt{2\pi} S(t)} - t. \quad (21)$$

Similarly, the second moment of the residual life time is provided by

$$\begin{aligned} \mu_2(t) &= t^2 - 2t V(t) + \frac{I_1(2, t)}{\sqrt{2\pi} S(t)} \\ &= t^2 - 2t \frac{I_1(1, t)}{\sqrt{2\pi} S(t)} + \frac{I_1(2, t)}{\sqrt{2\pi} S(t)}. \end{aligned} \quad (22)$$

Using $\mu_1(t)$ and $\mu_2(t)$, it is possible to calculate the variance of the residual life function.

4. Estimation of the Parameters

In this section, we go over how to estimate the HN distribution's parameters using the ML and Bayesian approaches, the two most widely used techniques.

4.1. Maximum Likelihood Estimation

Let x_1, x_2, \dots, x_n represent the observed values of a random sample formed of acrvs, say, X_1, X_2, \dots, X_n , taken from the HN distribution. The log-likelihood function for the parameter vector $\Theta = (\alpha, \mu, \sigma)^T$ is provided by

$$\mathcal{L}_n = -n \log \sqrt{2\pi} + \sum_{i=1}^n \log(\alpha + \sigma e^{x_i}) - \frac{1}{2} \sum_{i=1}^n (\alpha x_i + \sigma e^{x_i} + \mu)^2. \quad (23)$$

By maximising Equation (23) with regard to the parameters, the MLEs can be obtained more quickly in practice. Hereafter, we denote by $\hat{\mu}$, $\hat{\alpha}$, and $\hat{\sigma}$ the MLEs of μ , α , and σ , respectively. From a computational viewpoint, the score vector function is

$$\mathbf{U} = \left(\frac{\partial \mathcal{L}_n}{\partial \alpha}, \frac{\partial \mathcal{L}_n}{\partial \mu}, \frac{\partial \mathcal{L}_n}{\partial \sigma} \right)^T, \quad (24)$$

where

$$\frac{\partial \mathcal{L}_n}{\partial \mu} = - \sum_{i=1}^n (\alpha x_i + \sigma e^{x_i} + \mu), \quad (25)$$

$$\frac{\partial \mathcal{L}_n}{\partial \alpha} = \sum_{i=1}^n \frac{1}{\alpha + \sigma e^{x_i}} - \sum_{i=1}^n x_i (\alpha x_i + \sigma e^{x_i} + \mu) \quad (26)$$

and

$$\frac{\partial \mathcal{L}_n}{\partial \sigma} = \sum_{i=1}^n \frac{e^{x_i}}{\alpha + \sigma e^{x_i}} - \sum_{i=1}^n e^{x_i} (\alpha x_i + \sigma e^{x_i} + \mu). \quad (27)$$

By computationally solving the nonlinear system of equations $\mathbf{U} = (0, 0, 0)^T$, it is possible to derive the MLEs. In this study, we obtain the MLEs numerically using the `fitdistrplus` package in R software (see [15]). The following link provides more information about this package: (<https://CRAN.R-project.org/package=fitdistrplus>).

We now depict the asymptotic confidence intervals (CIs) for the parameters. On taking the second partial derivatives of Equations (25)–(27) at the MLEs, the observed Hessian matrix of the HN distribution can be obtained, and is specified by

$$\hat{H} = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_n}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \partial \sigma} \\ \frac{\partial^2 \mathcal{L}_n}{\partial \mu \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma^2} \end{pmatrix} \bigg|_{(\alpha, \mu, \sigma) = (\hat{\alpha}, \hat{\mu}, \hat{\sigma})}.$$

Based on the above, the observed Fisher's information matrix is $\hat{J} = -\hat{H}$, from which we can derive the estimated dispersion matrix as

$$\hat{\Xi} = \hat{J}^{-1} = \begin{pmatrix} \hat{\Xi}_{11} & \hat{\Xi}_{12} & \hat{\Xi}_{13} \\ \hat{\Xi}_{21} & \hat{\Xi}_{22} & \hat{\Xi}_{23} \\ \hat{\Xi}_{31} & \hat{\Xi}_{32} & \hat{\Xi}_{33} \end{pmatrix},$$

and $\hat{\Xi}_{ij} = \hat{\Xi}_{ji}$ for $i \neq j = 1, 2, 3$. The asymptotic N distribution of the random versions of the MLEs is guaranteed by the current general theory, with technical regular assumptions. More precisely, the random version of $\Theta = (\hat{\alpha}, \hat{\mu}, \hat{\sigma})^T$ follows the multivariate N distribution $N_3(\Theta, \hat{\Xi})$. Thus, we obtain $100 \times (1 - \delta)\%$ asymptotic CIs of the parameters using the following formulas:

$$I_{\alpha} = \left[\hat{\alpha} \mp v_{\delta/2} \sqrt{\hat{\Xi}_{11}} \right], \quad I_{\mu} = \left[\hat{\mu} \mp v_{\delta/2} \sqrt{\hat{\Xi}_{22}} \right], \quad I_{\sigma} = \left[\hat{\sigma} \mp v_{\delta/2} \sqrt{\hat{\Xi}_{33}} \right],$$

where v_{δ} is the upper δ th percentile of the SN distribution.

4.2. MCMC Estimation Using Bayesian Approach

In this subsection, we implement the MCMC estimation using the Bayesian estimation procedure on the HN distribution parameters. Each parameter demands a prior pdf to accomplish this. For this, we use two different variants of the almost improper uniform priors, as when using the almost improper uniform priors the comparison with MLEs is not confounded by the choice of proper prior distributions. Thus, we choose Uniform $(-1000, 1000)$ ($U(-1000, 1000)$) for the parameter μ , and $U(0, 1000)$ for the other two parameters. With this configuration, the joint posterior pdf is described by

$$\psi(\mu, \alpha, \sigma | x_1, x_2, \dots, x_n) \propto L_n \psi(\mu) \psi(\alpha) \psi(\sigma), \quad (28)$$

where $L_n = e^{\mathcal{L}_n}$ is the likelihood function of the HN distribution, where $\psi(\mu)$, $\psi(\alpha)$ and $\psi(\sigma)$ are the almost improper prior pdfs of the corresponding parameters. From Equation (28), there is obviously no analytical way to determine the MCMC Bayesian estimates. To perform numerical computations for MCMC Bayesian estimation, the LaplacesDemon package of the R software, which provides a comprehensive environment for Bayesian inference, is used. For more information and examples about this package, the following link is recommended: <https://cran.r-project.org/package=LaplacesDemon> (accessed on 14 September 2022).

5. Simulation Study

In this section, we execute simulation tests to evaluate the long-term performance of the MLEs of the HN distribution parameters. Several finite sample sizes are considered. More precisely, we generate samples of sizes $n = 25, 50, 100, 250$, and 500 from the HN distribution for two sets of parameter values.

5.1. Simulation Study for the MLEs

Here, the iteration is conducted 500 times, and the mean values of the biases, root mean squared errors (RMSEs), 95% (asymptotic) coverage probabilities (CPs), and ALs of the 95% (asymptotic) CIs corresponding to each of the parameter estimates for every replication are calculated with respect to the corresponding sample sizes. The results corresponding to each of the parameter sets are reported in Tables 1 and 2.

It can be seen that the RMSEs and ALs corresponding to each estimate decrease as the sample size increases. Additionally, for each parameter the CPs are relatively close to expected value of 0.95. This attests to the MLEs of the HN distribution perform consistently.

Table 1. The simulation results of the MLEs for the set $(\alpha = 0.2, \mu = 0.9, \sigma = 0.6)$.

Parameters	n	MLE	Bias	RMSE	CP	AL
α	25	0.2087	0.0087	0.0394	0.966	0.1548
	50	0.2045	0.0045	0.0273	0.948	0.1063
	100	0.2022	0.0022	0.0178	0.964	0.0742
	250	0.2015	0.0015	0.0120	0.942	0.0466
	500	0.2007	0.0007	0.0086	0.944	0.0328
μ	25	0.9237	0.0237	0.3106	0.958	1.2256
	50	0.9173	0.0173	0.2240	0.936	0.8528
	100	0.9051	0.0051	0.1532	0.942	0.5976
	250	0.9039	0.0039	0.0942	0.964	0.3757
	500	0.8998	-0.00018	0.0650	0.966	0.2651
σ	25	0.7869	0.1869	0.5733	0.954	1.5415
	50	0.6754	0.0754	0.2842	0.956	0.9509
	100	0.6376	0.0376	0.1695	0.950	0.6353
	250	0.6133	0.0133	0.1017	0.944	0.3889
	500	0.6067	0.0067	0.0683	0.950	0.2717

Table 2. The simulation results of the MLEs for the set ($\alpha = 1.1, \mu = 0.2, \sigma = 0.05$).

Parameters	n	MLE	Bias	RMSE	CP	AL
α	25	1.0490	−0.0510	0.2432	0.968	1.1842
	50	1.0689	−0.0311	0.1681	0.964	0.7897
	100	1.0803	−0.0197	0.1133	0.988	0.5403
	250	1.0894	−0.0106	0.0753	0.974	0.3350
	500	1.0935	−0.0065	0.0566	0.968	0.2327
μ	25	0.0443	−0.1557	0.3837	0.966	1.5773
	50	0.1190	−0.0810	0.2566	0.966	1.0372
	100	0.1509	−0.0491	0.1727	0.962	0.7024
	250	0.1746	−0.0254	0.1050	0.968	0.4336
	500	0.1864	−0.0136	0.0749	0.964	0.2997
σ	25	0.1916	0.1416	0.2948	0.978	1.0386
	50	0.1250	0.0750	0.1707	0.978	0.6460
	100	0.0913	0.0413	0.1025	0.984	0.4234
	250	0.0720	0.0220	0.0639	0.972	0.2570
	500	0.0596	0.0096	0.0439	0.968	0.1743

5.2. Simulation Study for MCMC Bayesian Estimates

Here, we consider the prior distributions for the HN distribution parameters provided in Subsection 4.2. Iteration is conducted 10,001 times. For each parameter set of the respective sample sizes, the posterior summary results, such as mean, standard deviation (SD), Monte Carlo error (MCE), ALs of the 95% (asymptotic) CIs, and median, are presented in Tables 3 and 4.

Table 3. Posterior summary results for ($\alpha = 0.2, \mu = 0.9, \sigma = 0.6$).

Parameters	n	Mean	SD	MCE	AL	Median
α	25	0.2901	0.0511	0.0084	0.1891	0.2866
	50	0.2940	0.0429	0.0072	0.1607	0.2883
	100	0.2109	0.0215	0.0034	0.0837	0.2115
	250	0.2108	0.0106	0.0032	0.0357	0.2133
	500	0.2090	0.0052	0.0008	0.0144	0.2038
μ	25	1.0248	0.2868	0.0440	1.2101	1.0270
	50	1.0688	0.2519	0.0389	0.9015	1.0579
	100	0.8076	0.1592	0.0251	0.5298	0.7742
	250	0.9186	0.1071	0.0201	0.3615	0.9101
	500	0.9651	0.0384	0.0046	0.0984	0.9855
σ	25	0.4747	0.2909	0.0592	1.1074	0.4390
	50	0.4946	0.2443	0.0528	0.8478	0.5202
	100	0.7444	0.2018	0.0429	0.8019	0.7700
	250	0.5579	0.0741	0.0235	0.2961	0.5930
	500	0.6533	0.0597	0.0156	0.1831	0.5656

Table 4. Posterior summary results for ($\alpha = 1.1, \mu = 0.2, \sigma = 0.05$).

Parameters	n	Mean	SD	MCE	AL	Median
α	25	1.2324	0.2087	0.0255	0.8182	1.2660
	50	1.1025	0.1481	0.0177	0.5810	1.0935
	100	1.0528	0.0917	0.0159	0.3420	1.0744
	250	1.1009	0.0861	0.0132	0.2751	1.1148
	500	1.0993	0.0590	0.0058	0.1472	1.1162
μ	25	0.2278	0.2569	0.0494	1.1499	0.2437
	50	0.1767	0.1690	0.0305	0.8105	0.2195
	100	0.2530	0.1482	0.0231	0.6799	0.2523
	250	0.2139	0.0910	0.0141	0.3427	0.2457
	500	0.2151	0.0838	0.0085	0.2558	0.1958
σ	25	0.0476	0.1330	0.0277	0.3482	0.03409
	50	0.0316	0.0659	0.0129	0.1901	0.0272
	100	0.0469	0.0541	0.0121	0.1761	0.0237
	250	0.0741	0.0509	0.0118	0.0909	0.0676
	500	0.0300	0.0331	0.0108	0.0693	0.0120

It can be seen that the SD, MCE, and AL decline as the sample size rises, which indicates that the MCMC Bayesian estimates of the HN distribution parameters perform consistently.

6. Bootstrap CIs

In this section, we approximate the distribution of the MLEs of the HN model parameters using the parametric bootstrap method. Then, we can estimate CIs for each parameter for the fitted HN distribution using the bootstrap distribution. Using the dataset comprised of x_1, x_2, \dots, x_n , let the MLE of $\theta \in \{\alpha, \mu, \sigma\}$ be $\hat{\theta}$. In our setting, the bootstrap stands for a method for estimating the distribution of $\hat{\theta}$ by drawing a random sample $(\theta_1^*, \theta_2^*, \dots, \theta_B^*)$ for θ based on B random samples that are drawn with replacement from x_1, x_2, \dots, x_n . The sample of bootstrapping, i.e., $(\theta_1^*, \theta_2^*, \dots, \theta_B^*)$, can be used to construct bootstrap CIs for each HN distribution parameter.

Thus, using the following formulas, we calculate the $100 \times (1 - \delta)\%$ bootstrap CIs for the parameters as

$$I_{\alpha,boot} = [\hat{\alpha} \mp z_{\delta/2} \hat{se}_{\alpha,boot}], I_{\mu,boot} = [\hat{\mu} \mp z_{\delta/2} \hat{se}_{\mu,boot}], I_{\sigma,boot} = [\hat{\sigma} \mp z_{\delta/2} \hat{se}_{\sigma,boot}],$$

where z_{δ} represents the δ th percentile of the bootstrap sample; for $\theta \in \{\alpha, \mu, \sigma\}$,

$$\hat{se}_{\theta,boot} = \sqrt{\frac{1}{B} \sum_{b=1}^B \left(\theta_b^* - \frac{1}{B} \sum_{b=1}^B \theta_b^* \right)^2}.$$

7. Applications and Empirical Study

The purpose of this section is to illustrate the empirical significance of the HN distribution. We consider a real dataset from the area of subject physics. The data consist of a sample from an experimental dataset of the strength of glass fibers with a length 15 cm from the National Physical Laboratory in England. This dataset, usually called the glass fiber (GF) dataset, can be found in [16]. The summary statistics of the data are provided in Table 5. We use the R software for numerical evaluations.

Table 5. Descriptive statistics of the strength of the GF dataset.

Statistic	n	Min	First Quartile	Median	Mean	Third Quartile	Max
Values	46	0.37	0.9575	1.16	1.13	1.3375	1.61

We study the empirical hazard rate function of the dataset using the concept of total time on test (TTT) plot. The TTT plot is a graph that is primarily used to distinguish between various ageing types indicated by hazard rate shapes. For more details, see [17]. Figure 4 indicates that the dataset has an increasing hazard rate shape for its empirical hazard rate function. Therefore, the HN distribution represents a credible choice for this dataset.

We demonstrate the potentiality of the HN distribution by comparing it with the Laplace, logistic, Gumbel, and N distributions, all of which are defined on the entire real line. Using the statistical tools of the negative log-likelihood ($-\log L$), Kolmogorov–Smirnov statistics (KS), Cramér–von Mises (W^*), Anderson–Darling (A^*), the values of the Akaike information criterion (AIC), and Bayesian information criterion (BIC), we compare the competing models with the suggested model. The corresponding MLEs and goodness-of-fit (GOF) statistics are presented in Table 6.

From this table, it can be seen that the GOF statistics values of the HN distribution are lower than those of the examined distributions. In light of this, we can draw the conclusion that the suggested HN model provides a better fit than those from the compared distributions.

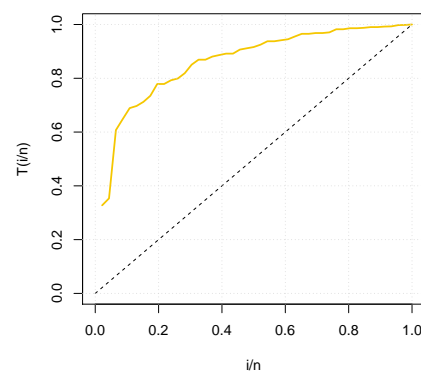


Figure 4. The TTT plot of the strength of the GF dataset.

Table 6. MLEs and GOF statistics results of the strength of the GF dataset.

Estimates	Laplace	Logistic	Gumbel	N	Hybrid N
$\hat{\alpha}$	-	-	-	-	0.0142
$\hat{\mu}$	1.1609	1.1504	0.9865	1.1300	-4.0688
$\hat{\sigma}$	0.2095	0.1498	0.3111	0.2684	1.2659
$-\log L$	5.9996	4.4613	13.4978	4.7700	2.2719
KS	0.0864	0.0734	0.1497	0.0928	0.0619
W^*	0.0648	0.0411	0.2947	0.0742	0.0187
A^*	0.5121	0.3790	1.9881	0.5267	0.1397
AIC	15.9993	12.9225	30.9955	13.5400	10.5439
BIC	19.6566	16.5798	34.6528	17.1973	16.0298

The plots of the empirical cdfs, empirical pdfs, and the theoretical probabilities against empirical ones (P–P plot) for the real dataset (for the HN model only for the P–P plot) are displayed in Figure 5.

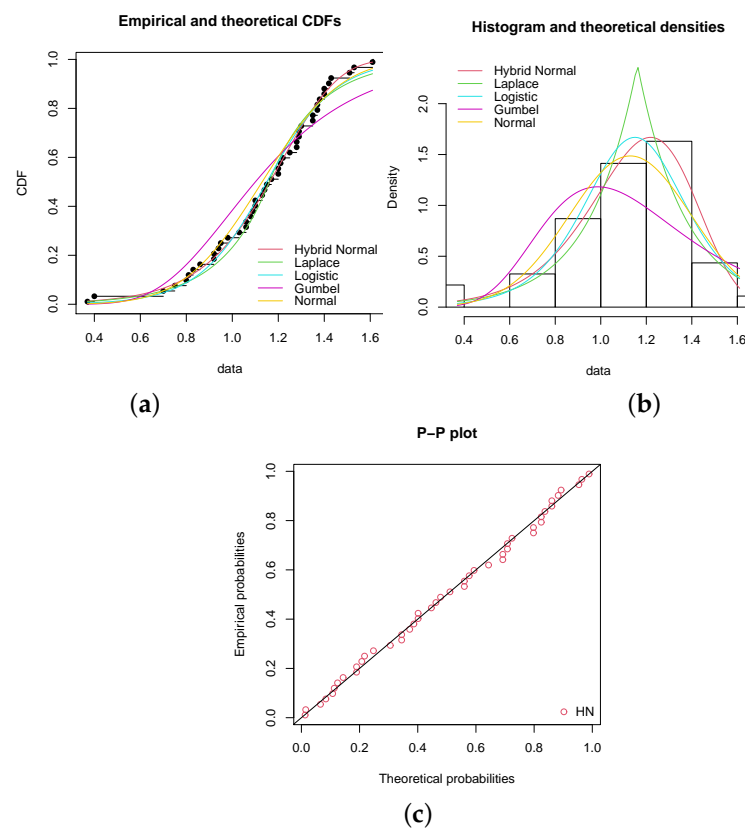


Figure 5. Empirical plots on the strength of the GF dataset: (a) empirical versus theoretical cdfs, (b) histogram and fitted pdfs, and (c) P–P plot.

This graphic displays attractive curves for the fitted and empirical functions. In light of this, we can draw the conclusion that the HN distribution is the most appropriate for this dataset among the alternative distributions.

Now, the empirical Hessian matrix is obtained as follows:

$$\hat{H} = \begin{pmatrix} 65.5553 & 51.9800 & 185.5325 \\ 51.9800 & 46.0000 & 147.2665 \\ 185.5325 & 147.2665 & 528.4451 \end{pmatrix},$$

and the associated estimated dispersion matrix is

$$\hat{\Xi} = \begin{pmatrix} 2.4932 & -0.1388 & -0.8367 \\ -0.1388 & 0.2093 & -0.0096 \\ -0.8367 & -0.0096 & 0.2983 \end{pmatrix}.$$

In order to complete the above analysis, Table 7 provides the 95% (asymptotic) CIs for the HN distribution parameters.

Table 7. The 95% (asymptotic) CIs of the HN distribution parameters based on the strength of the GF dataset.

Parameter	Lower	Upper
α	−3.0806	3.1090
μ	−4.9655	−3.1720
σ	0.1954	2.3364

On the basis of the GF dataset outlined above, we now concentrate on estimating the HN distribution parameters using the MCMC Bayesian technique. The investigation was carried out using the MH algorithm of the MCMC method with respect to 1000 iterations in the perspective of Bayesian estimation. All three MCMC Bayesian estimates of the HN distribution parameters are included in Table 8 for comparison with the MLEs. Numerical computations for MCMC Bayesian estimation were performed using R software.

Table 8. ML and MCMC Bayesian estimates of the HN distribution parameters on the strength of the GF dataset.

Parameter	ML	MCMC Bayesian
α	0.0142	0.0135
μ	−4.0688	−4.1753
σ	1.2659	1.2977

We now focus on the 95% bootstrap CIs for the HN distribution parameters using the computed MLEs. To this end, we generate 1001 samples of the same size as the GF dataset using the HN distribution, with the true values of parameters chosen as the corresponding MLEs. For each obtained sample, we calculate the MLEs $\hat{\alpha}_b^*$, $\hat{\mu}_b^*$, and $\hat{\sigma}_b^*$ for $b \in \{1, 2, \dots, 1001\}$. Thus, for the parameters α , μ , and σ , Table 9 shows the median and 95% bootstrap CIs.

Table 9. The median and 95% bootstrap CIs for the HN distribution parameters on the strength of the GF dataset.

	Parameter	Median	Bootstrap CI
Strength of the GF dataset	α	0.2340	(0.0001, 4.3993)
	μ	−4.2647	(−5.6048, −3.4066)
	σ	1.1516	(0.0001, 1.5551)

In order to identify any potential structural correlation between the parameters, it is noteworthy to examine the joint distribution of the bootstrapped values in the matrix of scatter plots. Figure 6 demonstrates the matrix scatterplots of the bootstrapped HN parameter values, which reflect the joint uncertainty distribution with respect to the fitted parameters.

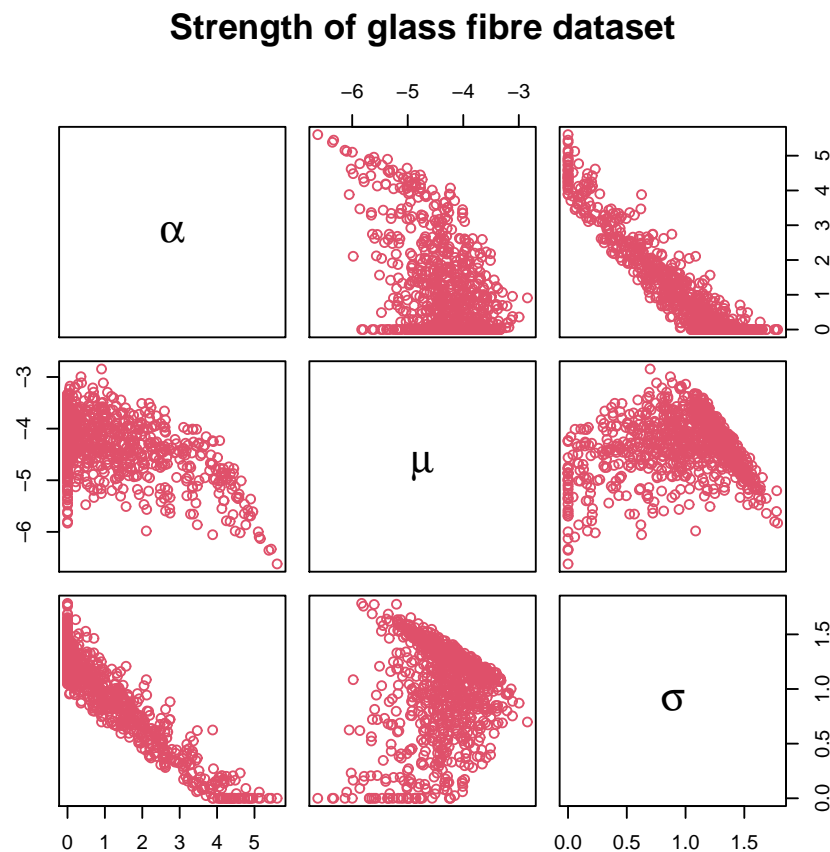


Figure 6. Matrix scatterplots of the bootstrapped HN parameter values due to strength of the GF dataset.

8. Concluding Remarks

In this study, we suggest a new distribution that generalizes the well-known normal distribution and is defined on the entire real line. We investigate the mathematical and statistical features of the new model, which we refer to as the hybrid normal (HN) distribution. With regard to the HN distribution, we present explicit expressions for several reliability metrics. The hazard rate function of the HN distribution possesses increasing shaped graphical representation. In terms of inference, the observed information matrix is presented together with the estimation of the model parameters using the maximum likelihood and Bayesian estimation methods, with Bayesian estimation performed using the MCMC approach. In addition, we use the parametric bootstrap method to obtain the model parameter confidence intervals. Goodness-of-fit tests are applied to a real dataset concerning the strength of glass fibers with a length of 15 cm from the National Physical Laboratory in England in order to demonstrate the new model's real-world applicability. In comparison to the other analyzed models, the new model consistently offers a superior fit. We expect that the proposed model can be used more frequently to represent actual datasets in a variety of fields, including physics, engineering, economics, hydrology, survival analysis, and others.

Author Contributions: Conceptualization, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M.; methodology, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M.; validation, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M.; formal analysis, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M.; investigation, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M.; writing—original draft preparation, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M.; writing—review and editing, D.S.S., S.L.N., C.C., M.R.I., S.P.S. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article.

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

References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2. Wahed, A.; Ali, M. The skew-logistic distribution. *J. Statist. Res.* **2001**, *35*, 71–80.
3. Genton, M. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; CRC Press: Boca Raton, FL, USA, 2004; pp. 1–396.
4. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 208. [\[CrossRef\]](#)
5. Castillo, E.; Hadi, A.; Balakrishnan, N.; Sarabia, J. *Extreme Value and Related Models with Applications in Engineering and Science*; Wiley: Hoboken, NJ, USA, 2005.
6. Ferrari, S.; Pinheiro, E.C. Small-sample likelihood inference in extreme-value regression models. *J. Stat. Comput. Simul.* **2012**, *84*, 582–595. [\[CrossRef\]](#)
7. Domma, F.; Perri, P.F. Some developments on the log-Dagum distribution. *Stat. Methods Appl.* **2009**, *18*, 205–220. [\[CrossRef\]](#)
8. Landauer, R. The physical nature of information. *Phys. Lett. A* **1996**, *217*, 188–193.1 [\[CrossRef\]](#)
9. Wang, X.; Wang, B.; Zhang, Y.; Suo, Y.; Jia, P.; Huang, F. Dispersion of mechanical properties of high-strength glass fibre composites in hygrothermal environment. *Polymers* **2022**, *14*, 3514. [\[CrossRef\]](#)
10. Kumazawa, S.; Numakunai, T. A New Theoretical analysis of occupational dose distributions Indicating the Effect of Dose Limits. *Health Phys.* **1981**, *41*, 465–475. [\[CrossRef\]](#)
11. Fletcher, S.J.; Zupanski, M. A hybrid multivariate normal and lognormal distribution for data assimilation. *Atmos. Sci. Lett.* **2006**, *7*, 43–46. [\[CrossRef\]](#)
12. Rupšys, P.; Narmontas, M.; Petrauskas, E. A multivariate hybrid stochastic differential equation model for whole-stand dynamics. *Mathematics* **2020**, *8*, 2230. [\[CrossRef\]](#)
13. Kupka, J.; Loo, S. The hazard and vitality measures of ageing. *J. Appl. Probab.* **1989**, *26*, 532–542. [\[CrossRef\]](#)
14. Nair, K.; Rajesh, G. Geometric vitality function and its applications to reliability. *IAPQR Trans.* **2000**, *25*, 1–8.
15. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: <https://www.R-project.org/> (accessed on 14 September 2022).
16. Smith, R.L.; Naylor, J.C. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *J. R. Stat. Soc. Ser. (Appl. Stat.)* **1987**, *36*, 358–369. [\[CrossRef\]](#)
17. Aarset, M.V. How to identify a bathtub hazard rate. *IEEE Trans. Reliab.* **1987**, *R-36*, 106–108. [\[CrossRef\]](#)