https://doi.org/10.46810/tdfd.1440444



Statistical Analysis of Wind Speed Data with Different Distributions: Bitlis, Türkiye

Asuman YILMAZ ^{1*D}, Mahmut KARA^{1D}

1.2 Van Yüzüncü Yıl University, Faculty of Economics and Administrative Sciences, Department of Econometrics,

Van, Türkiye

Asuman YILMAZ ORCID No: 0000-0002-8653-6900 Mahmut KARA ORCID No: 0000-0001-7678-8824

**Corresponding author: asumanduva@yyu.edu.tr*

(Received: 21.02.2024, Accepted: 19.10.2024, Online Publication: 30.12.2024)

Keywords

Wind speed data, Weibull distribution, Log-normal distribution, Gamma distribution, Generalized Rayleigh, Distribution, Maximum likelihood estimation **Abstract:** Accurately modeling wind speed is important in estimating the wind energy potential of a specified region. Two- parameter Weibull distribution is the most widely used and accepted distribution in the energy literature. However, it does not model the all wind speed data encountered in nature. Therefore, in this study, different distributions are used for modeling wind energy, such as Gamma, lognormal, Generalized Rayleigh. The estimators of the unknown parameters of these distributions are found by using maximum likelihood estimators.

Rüzgar Hızı Verilerinin Farklı Dağılımlara Göre İstatistiksel Analizi: Bitlis, Türkiye

Anahtar Kelimeler

Rüzgar hızı verileri, Weibull dağılımı, Gamma dağılımı, Log-normal dağılımı, Genelleştirilmiş Rayleigh dağılımı, En çok olabilirlik tahmini Öz: Rüzgâr hızının doğru bir şekilde modellenmesi, belirli bir bölgenin rüzgâr enerjisi potansiyelinin tahmin edilmesi açısından önemlidir. İki parametreli Weibull dağılımı enerji literatüründe en yaygın kullanılan ve kabul edilen dağılımdır. Ancak doğada karşılaşılan tüm rüzgâr hızı verilerini modellemez. Bu nedenle bu çalışmada rüzgâr enerjisinin modellenmesinde Gamma, log-normal, Genelleştirilmiş Rayleigh gibi farklı dağılımlar kullanılmıştır. Bu dağılımların bilinmeyen parametrelerinin tahmin edicileri, maksimum olabilirlik tahmin edicileri kullanılarak bulunur.

1. INTRODUCTION

Energy consumption is rising substantially as a result of both population growth and advancements in technology. One of the most crucial elements in the growth of a country is its energy needs [1,2]. These days, fossil fuels are used extensively in practically every aspect of daily life, such as production, logistics, and heating. However, continued consumption of fossil fuels endangers both people and the world in the long run due to carbon emissions, air pollution, and climate change [3]. To solve these problems, it is recommended to use environmentally friendly alternative energy sources including geothermal, solar, and wind power. Therefore, renewable energy sources have attracted attention in recent decades, particularly in developed countries [4, 5].

A promising renewable energy source, wind power can be used for stand-alone, remote, and grid-connected applications in addition to direct energy delivery [6]. Over the past 20 years, wind power generation has grown amazingly rapidly, and it is now a mature, reliable, and efficient technology for producing electricity [3]. It has been used as an essential renewable energy source in electricity production in many developed countries.

To effectively and economically obtain wind energy, there are two key components. The first of these is where the wind system will be built. The second is to decide on the statistical distributions to be used to determine wind speed characteristics. Accurately determining the wind speed distribution has an important impact on wind power calculations [7]. Because small errors in the modeling of the wind speed data lead to significantly larger errors in the energy outcome computation [8]. The Weibull distribution is frequently used in literature to model wind speed data. However, it is not suitable for all wind regimes [9-10]. Therefore, different distributions including

Gamma, inverse Weibull, Inverted Kumaraswamy, lognormal, and inverse Gaussian are used for modelling the wind speed, see [8,11-15]. Considering all these points, Weibull, Gamma, log-normal, and Generalized Rayleigh distributions are used in modeling wind energy in this study. The estimators of the unknown parameters of these distributions are found by using maximum likelihood estimators (MLEs). The originality of this study comes from the fact that it considers different statistical distributions to model wind speed data.

The rest of the study is organized as follows. In section 2, the distributions used in the study are given along with the ML method for estimating parameters of the Weibull, Gamma, inverse Gauss, and lognormal distributions. In section 3, application results are presented and modeling performances of the given distributions are compared. In section 4, the study is finalized with some concluding remarks.

2. MATERIAL AND METHOD

This section provides parameter estimations and the probability distribution functions (pdf), the cumulative density function (cdf) for the Weibull, and gamma, lognormal, and generalized Rayleigh distributions.

2.1. Weibull Distribution

Among lifetime distributions, the Weibull distribution is one of the most often used. W. Weibull first proposed the Weibull distribution and used it to describe the distribution of a materials breaking strength. The reliability theory, numerous environmental science fields, and renewable energy have all benefited greatly from the distribution; see [16–18]. The probability density function (pdf) and the pdf cumulative density function (cdf) and of the Weibull distribution are:

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{\left(\frac{-x}{\beta}\right)^{\alpha}} < x < \infty; \ \alpha > 0, \beta > 0 \ (1)$$

and

$$F(x;\alpha,\beta) = 1 - e^{\left(\frac{x}{\beta}\right)^{-\alpha}} 0 < x < \infty; \ \alpha > 0, \beta > 0.$$
(2)

Here, α is the shape parameter and β is the scale parameter.

2.1.1 Parameter estimation of Weibull distribution

Let $X_1, X_2, ..., X_n$ be a random sample from Weibull distribution. The likelihood function is given by:

$$L(x;\alpha,\beta) = \frac{\alpha^n}{\beta^{n\alpha}} \prod_{i=1}^n x_i^{(\alpha-1)} e^{-\left(\frac{x_i}{\beta}\right)^{\alpha}}$$
(3)

Following that, the logarithm of $L(x; \alpha, \beta)$ is defined as follows to define the log-likelihood function:

$$lnL = nln\alpha - n\alpha ln\beta + (\alpha - 1)\sum_{i=1}^{n} lnx_i - \sum_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha}.$$
 (4)

By taking the partial derivative of (4) with respect to α and β , and equating them to zero, we obtain the following loglikelihood equations:

$$\frac{\partial lnL}{\alpha} = \frac{n}{\alpha} - nln\beta + \sum_{i=1}^{n} lnx_i - \sum_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha} ln\left(\frac{x_i}{\beta}\right) = 0$$
(5)

$$\frac{\partial lnL}{\beta} = \frac{-n\alpha}{\beta} + \frac{\alpha \sum_{i=1}^{n} x_i^{\alpha}}{\beta^{\alpha+1}} = 0.$$
 (6)

Then, from the solution of equation (6), the parameter β is obtained as

$$\beta = \left(\frac{\sum_{i=1}^{n} x_i^{\alpha}}{n}\right)^{\frac{1}{\alpha}}$$

By substituting the solution of β into equation (5), we have

$$\frac{\sum_{i=1}^{n} x_i^{\alpha} ln x_i}{\sum_{i=1}^{n} x_i^{\alpha}} - \frac{1}{\alpha} - \frac{\sum_{i=1}^{n} ln x_i}{n} = 0$$

The equation $g(\alpha)=0$ cannot be solved explicitly, but a numerical method should be used to solve this equation. The $\hat{\alpha}$ can be obtained by using the following well-known Newton-Raphson formula

$$\hat{a}_{m+1=\hat{a}_m} - \frac{g(\hat{a}_m)}{g'(\hat{a}_m)} \ m = 1,2,3, \dots$$

where $\hat{\alpha}_1$ is as initial value and $g'(\hat{\alpha})$ is the derivative of $g(\alpha)$. Then the ML estimator of β is given as follows:

$$\hat{\beta} = \left(\frac{\sum_{i=1}^{n} x_i^{\hat{\alpha}}}{n}\right)^{\frac{1}{\alpha}}.$$

2.2. Gamma Distribution

Because of its flexible shape, the positively skewed gamma distribution has numerous applications in a variety of industries, including hydrology, engineering, medicine, seismology, and reliability. The works of Aksoy [19], Shapiro and Chen [20], and Hristopulos et al. [21] provide a thorough summary of the literature's study on the application of gamma distribution.

The pdf and the cdf of the Gamma distribution are given by:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{\frac{-x}{\beta}}, x > 0; \alpha, \beta > 0,$$
(7)

and

$$F(x) = \int_0^x \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{\frac{-x}{\beta}} dx \ x > 0; \alpha, \beta > 0, \tag{8}$$
respectively.

2.2.1. Parameter estimation of Gamma distribution

Let $X_1, X_2, ..., X_n$ be a random sample of size *n* from Gamma (α, β) distribution. The likelihood function for X_i i = 1, 2, ..., n is written as follows:

$$L(x;\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x_{i}^{(\alpha-1)} e^{-\frac{x_{i}}{\beta}}$$
(9)

Then the log-likelihood function is defined as the logarithm of $L(x; \alpha, \beta)$ is given by:

$$lnL(x; \alpha, \beta) =$$
$$-nln\Gamma(\alpha) - n\alpha ln(\beta) + (\alpha - 1)\sum_{i=1}^{n} lnx_{i} - \frac{1}{\beta}\sum_{i=1}^{n} x_{i}.$$
 (10)

The likelihood equations are found by equating the first partial derivatives of $lnL(x; \alpha, \beta)$ with respect to the α and β parameters to zero. The α and β parameters are shown as follows:

$$\frac{\partial lnL}{\partial \alpha} = -n\Psi(\alpha) - nln\beta + \sum_{i=1}^{n} \ln x_i = 0 \quad (11)$$

and

$$\frac{\partial lnL}{\partial \beta} = \frac{-n\alpha}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i = 0.$$
(12)

Here $\Psi(\alpha) = \frac{\partial ln\Gamma(\alpha)}{\partial \alpha}$ is the digamma function. Then, by solving equation (11), the parameter β is found as:

$$\beta = \frac{1}{n\alpha} \sum_{i=1}^n x_i.$$

Substitution of β into equation (12), the resulting equation in α becomes

$$-n\Psi(\alpha) - nln\left(\frac{1}{n\alpha}\sum_{i=1}^n x_i\right) + \sum_{i=1}^n \ln x_i = 0.$$

This equation does not yield an explicit estimator for the α ; therefore, we resort to iterative methods. Hence, approximate solutions for the parameters are found using iterative numerical methods. In this study, we used the Newton Raphson method.

2.3. Lognormal Distribution

The lognormal distribution, a long-tailed, positively skewed distribution, is a suitable model for reliability and life span study. Furthermore, Johnson et al. [22] mention that it can be used as a model for several applications in engineering and medical.

The pdf of the log-normal distribution is given by:

$$f(x) = \frac{1}{x\tau\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{\ln x - \mu}{\tau}\right)^2}, x > 0, \mu \in \mathbb{R}, \tau > 0, \qquad (13)$$

where $\frac{1}{\tau}$ is the shape parameter and μ is the scale parameter.

2.3.1. Parameter estimation of lognormal distribution

The likelihood function based on the observed values of a random sample from the lognormal distribution with pdf (13) is given by

$$L(x;\mu,\tau) = \frac{1}{\sigma^{n}(2\pi)^{n/2}} \prod_{i=1}^{n} \frac{1}{x_{i}} e^{\frac{-1}{2\sigma^{2}} \sum_{i=1}^{n} (lnx_{i}-\mu)^{2}}$$
(14)

Then the log likelihood function is

$$lnL(x; \mu, \tau) = -nln\tau - \frac{n}{2}ln(2\pi) - \sum_{i=1}^{n} lnX_i - \frac{1}{2}\sum_{i=1}^{n} \left(\frac{(lnX_i - \mu)}{\tau}\right)^2$$
(15)

By taking the derivatives of *lnL* with respect to these parameters and equating them to zero, the following likelihood equations are obtained

$$\frac{\partial lnL}{\partial \mu} = \frac{\sum_{i=1}^{n} lnx_i}{\tau^2} - \frac{2n\mu}{2\tau^2} = 0 \quad (16)$$

and

$$\frac{\partial \ln L}{\tau^2} = \frac{-n}{2\tau^2} + \sum_{i=1}^n \frac{(\ln X_{i-\mu})^2}{2\tau^4} = 0.$$
(17)

Thus, the MLs are

$$\hat{\mu} = \frac{\sum_{i=1}^{n} lnx_i}{n} \text{ and } \hat{\tau}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(lnx_i - \frac{\sum_{i=1}^{n} lnx_i}{n} \right)^2,$$

respectively.

2.4. Generalized Rayleigh Distribution

For data modeling, Burr [23] provided twelve distinct kinds of cumulative distribution functions. Out of the twelve, Burr-Type X and Burr-Type XII were the two distribution functions that attracted the greatest attention. The generalized Rayleigh distribution was appropriately termed and the two parameter Burr Type X distribution was recently presented by Surles and Padgett [24]. Keep in mind that the exponentiated Weibull distribution, which was first put forth by Mudholkar and Srivastava [25], includes the two-parameter generalized Rayleigh distribution as one of its members. Both general lifetime modeling and the modeling of strength data can benefit greatly from the application of this distribution.

The pdf of the two-parameter generalized Rayleigh distribution is:

$$f(x; \alpha, \lambda) = 2\alpha\lambda^2 x e^{-(\lambda x)^2} (1 - e^{-(\lambda x)^2})^{\alpha - 1}, x > 0, \alpha, \lambda > 0.$$
(18)

2.4.1. Parameter estimation of generalized Rayleigh distribution

Let $X_1, X_2, ..., X_n$ be random sample of size n from generalized Rayleigh distribution with α and λ parameters, then the likelihood function can be written as: $L(x; \alpha, \lambda) = 2^n \alpha^n \lambda^{2n} \prod_{i=1}^n x_i e^{-(\lambda x_i)^2} (1 - e^{-(\lambda x_i)^2})^{\alpha - 1}$. (19)

Then, the log-likelihood function $lnL(x; \alpha, \lambda)$ as follows:

$$lnL(x; \alpha, \lambda) = nln2 + nln\alpha + 2nl + \sum_{i=1}^{n} lnx_i - \lambda^2 \sum_{i=1}^{n} x_i^2$$

$$+(\alpha-1)\sum_{i=1}^n ln(1-e^{-(\lambda x)^2}).$$

When we take the derivatives with respect to parameters normal equations become:

$$\frac{\partial lnL}{\partial \alpha} = \frac{n}{\alpha} + \sum_{l=1}^{n} ln \left(1 - e^{-(\lambda x)^2} \right) = 0$$
 (20)

and

$$\frac{\partial lnL}{\partial \lambda} = \frac{2n}{\lambda} - 2\lambda \sum_{i=1}^{n} x_i^2 + 2\lambda(\alpha - 1) \sum_{i=1}^{n} \frac{x_i^2 e^{-(\lambda x)^2}}{1 - e^{-(\lambda x)^2}} = 0 \quad (21)$$

Note that $\hat{\alpha}$ and $\hat{\lambda}$ are not in explicit form and hence, numerical methods, such as fixed-point solution. The ML estimators of generalized Rayleigh distribution can be seen from Esemen and Gürler [26].

3. DATA AND APPLICATION

In this section, the data set introduced and the Weibull, Gamma, lognormal, and generalized Rayleigh distributions are applied to the hourly wind speed data (m/s). The performances of the distributions are compared using several well-known criteria, such as the AIC,

 Table 1. ML estimators of the parameters for the given distributions

 R^2 and RMSE. For analyzing MatlabR2021 software is used.

3.1. The Data Set

In this section, hourly wind speed data (m/s), measured hourly at 10m from Bitlis, Turkey during January, February, March, and April 2017 is used. There are 2978 observations recorded. The data is taken with official permission from the Turkish State Meteorological Service. For the hourly wind speed data (m/s), estimates of the parameters of above mentioned distributions are obtained as reported in Table 1. To determine the distribution providing better determination (R^2) and Akaike information criteria (AIC) values for each distribution, as shown in Table 2. In addition to these statistical criteria, the cumulative density function of the given distributions is presented in Figure 1 for wind speed data.

Table 2 gives the values of the evaluation criteria for the generalized Rayleigh, Gamma, Weibull, and lognormal distributions. It is well-known that a better fit is indicated by lower values of the AIC and RMSE and higher values of the R^2 .

Table 1. Will estimators of the parameters for the given distributions										
Month	Weibull		Gamma		Logi	normal	G.Rayleigh			
	â	β	â	β	μ	î	â	Â		
January	1.207	2.796	1.477	1.767	0.895	0.585	1.567	0.706		
February	1.432	2.364	1.999	1.067	0.488	0.781	2.292	0.906		
March	1.853	3.064	2.827	0.961	0.813	0.679	3.175	0.725		
April	1.639	3.643	2.220	1.467	0.939	0.776	2.299	0.588		

Table 2. Modeling performances of the given distribution for wind speed data

Month	Weibull			Gamma			Lognormal			G. Rayleigh		
	R^2	RMSE	AIC	R^2	RMSE	AIC	R^2	RMSE	AIC	R^2	RMSE	AIC
January	0.9675	0.0513	2739.63	0.9969	0.0498	2706.23	0.9932	0.0242	2546.33	0.9527	0.1534	2765.87
February	0.9778	0.0401	2134.90	0.9989	0.0309	2109.78	0.9829	00363	2194.20	0.9193	0.2145	0.2180.32
March	0.9968	0.0160	2625.55	0.9998	0.0132	2635.83	0.9717	0.0471	2725.20	0.9085	0.2338	2778.88
April	0.9990	0.0094	2907.13	0.9995	0.0203	2905.42	0.9715	0.0477	2990.68	0.9049	0.2391	2986.68





Figure 1. Cumulative density functions of Weibull, Gamma, lognormal, and generalized Rayleigh distributions, along with observed data (X Data), for wind speed measurements in Bitlis, Turkey, during (a) January, (b) February, (c) March, and (d) April (2017)

It can be seen from Table 1 that the Gamma distribution provided the smallest values for the AIC, and RMSE and the largest values for the R^2 . In other words, the Gamma distribution performed better than the other three distributions in modeling the first months of 2017 of Bitlis province. Furthermore, in Figure 1 are consistent with Table 2.

4. CONCLUSION

In this study, the Weibull, Gamma, lognormal, and generalized Rayleigh are used in modelling the wind speed data of Bitlis in Turkey for the first four months of 2017. In parameter estimation, the ML estimation methods are considered. The results summarized in both Figure 1 and Table 2 show that the Gamma distribution described wind speed data in Bitlis during January, February, March, and April 2017 better than Weibull, lognormal, and generalized Rayleigh distributions.

Acknowledgement

We would like to thank Van Yüzüncü Yıl University Scientific Research and Project Coordination for supporting our project number "FBA-2018-6855".

REFERENCES

- Keyhani, A., Ghasemi-Varnamkhasti, M., Khanali, M., Abbaszadeh, R. An assessment of wind energy potential as a power generation source in the capital of Iran, Tehran. Energy,2010; 35(1), 188-201.
- [2] Akpinar, E. K., Akpinar, S. An assessment on seasonal analysis of wind energy characteristics and wind turbine characteristics. Energy conversion and management, 2005; 46(11-12), 1848-1867.
- [3] Fyrippis, I., Axaopoulos, P. J., Panayiotou, G. Wind energy potential assessment in Naxos Island, Greece. Applied Energy, 2010; 87(2), 577-586.
- [4] Köse, R. An evaluation of wind energy potential as a power generation source in Kütahya, Turkey. Energy conversion and management, 2004; 45(11-12), 1631-1641.
- [5] Kaplan, Y. A. Overview of wind energy in the world and assessment of current wind energy policies in Turkey. Renewable and Sustainable Energy Reviews, 2015; 43, 562-568.
- [6] Mabel, M. C., Fernandez, E. Growth and future trends of wind energy in India. Renewable and Sustainable Energy Reviews, 2008;12(6), 1745-1757.
- [7] Mohammadi, K., Alavi, O., Mostafaeipour, A., Goudarzi, N., & Jalilvand, M. Assessing different parameters estimation methods of Weibull distribution to compute wind power density. Energy Conversion and Management, 2016; 108, 322-335.
- [8] Akgül, F. G., Şenoğlu, B., Arslan, T.. An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. Energy Conversion and Management, 2016; 114, 234-240.
- [9] Akdağ, S. A., Dinler, A. A new method to estimate Weibull parameters for wind energy applications.

Energy conversion and management, 2009; 50(7), 1761-1766.

- [10] Kusiak, A., Zheng, H., Song, Z. On-line monitoring of power curves. Renewable Energy, 2009; 34(6), 1487-1493.
- [11] Brano, V. L., Orioli, A., Ciulla, G., Culotta, S. Quality of wind speed fitting distributions for the urban area of Palermo, Italy. Renewable Energy, 2011; 36(3), 1026-1039.
- [12] Carta, J. A., Ramírez, P., Velázquez, S. Influence of the level of fit of a density probability function to wind-speed data on the WECS mean power output estimation. Energy Conversion and Management, 2008; 49(10), 2647-2655.
- [13] Morgan, E. C., Lackner, M., Vogel, R. M., Baise, L. G. Probability distributions for offshore wind speeds. Energy Conversion and Management, 2011; 52(1), 15-26.
- [14] Sohoni, V., Gupta, S., Nema, R. A comparative analysis of wind speed probability distributions for wind power assessment of four sites. Turkish Journal of Electrical Engineering and Computer Sciences, 2016; 24(6), 4724-4735.
- [15] Bagci, K., Arslan, T., Celik, H. E. Inverted Kumarswamy distribution for modeling the wind speed data: Lake Van, Turkey. Renewable and Sustainable Energy Reviews, 2021;135, 110110
- [16] Raghunathan, K., Subramaniam, V., Srinivasamoorthy, V. R. Studies on the tensile characteristics of ring and rotor yarns using modified Weibull distribution, 2002.
- [17] Chiang, Y. J., Shih, C. D., Lin, C. C., Tseng, Y. Y. Examination of tyre rubber cure by Weibull distribution functions. International Journal of Materials and Product Technology, 2004; 20(1-3), 210-219.
- [18] Wood, M. A., Gunderson, B., Xia, A., Zhou, X., Padmanabhan, V., Ellenbogen, K. A. Temporal patterns of ventricular tachyarrhythmia recurrences follow a Weibull distribution. Journal of cardiovascular electrophysiology, 2005;16(2), 181-185.
- [19] Aksoy, H. Use of gamma distribution in hydrological analysis. Turkish Journal of Engineering and Environmental Sciences, 2000; 24(6), 419-428.
- [20] Shapiro, S. S., Chen, L. Composite tests for the gamma distribution. Journal of Quality Technology, 2001;33(1), 47-59.
- [21] Hristopulos, D. T., Petrakis, M. P., Kaniadakis, G. Weakest-link scaling and extreme events in finitesized systems. Entropy, 2015;17(3), 1103-1122.
- [22] Johnson, N. L., Kotz, S., Balakrishnan, N. Continuous univariate distributions, volume 2 (Vol. 289). John wiley & sons;1995.
- [23] Burr, I. W. Cumulative frequency functions. The Annals of mathematical statistics, 1942; 13(2), 215-232.
- [24] Surles, J. G., Padgett, W. J. Inference for reliability and stress-strength for a scaled Burr type X distribution. Lifetime data analysis, 2001; 7, 187-200.

- [25] Mudholkar, G. S., Srivastava, D. K. Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE transactions on reliability, 1993;42(2), 299-302.
- [26] Esemen, M., Gürler, S. Parameter estimation of generalized Rayleigh distribution based on ranked set sample. Journal of Statistical Computation and Simulation, 2018; 88(4), 615-628.