

Comparison of Four Distributions for Frequency Analysis of Wind Speed

Sabereh Darbandi (Corresponding author)

Department of Water Engineering, Tabriz University, Tabriz, Iran

Tel: 98-914-310-3249 E-mail: sadarbandi.tabrizu@yahoo.com

Mohammad Taghi Aalami

Department of Civil Engineering, Tabriz University, Tabriz, Iran

E-mail: mtaalami@tabrizu.ac.ir

Hakimeh Asadi

Department of Water Engineering, Tabriz University, Tabriz, Iran

E-mail: tu.i2006@yahoo.com

Received: October 11, 2011

Accepted: November 29, 2011

Published: March 1, 2012

doi:10.5539/enrr.v2n1p96

URL: <http://dx.doi.org/10.5539/enrr.v2n1p96>

Abstract

The increase in negative effects of fossil fuels on the environment has forced many countries to use renewable energy sources, especially wind energy. Wind speed is the most important parameter of the wind energy. Probability distributions are useful for estimating wind speed because it is a random phenomenon. This study analyzes wind speed frequencies using wind data from Tabriz synoptic station in Iran. Four different distributions are fitted to the maximum annual wind from station, and parameters of the distributions are estimated using the method of maximum likelihood and the method of moments. Calculations are performed with *Mathematica*, a computer algebra system developed by Wolfram Research. The advantage of using this software is that the symbolic, numerical, and graphical computations can be combined and that all quantities can be accurately calculated; in particular, there is no need to resort to any approximate methods for the calculation of quantiles. There is a ready-to-use command for calculating quantiles from distributions that are built in *Mathematica*, while for other distributions they can be easily and accurately calculated by inverting the cumulative distribution functions or by solving nonlinear equations where the inversion is not possible. The best distribution is selected based on the root mean square error (RMSE), the coefficient of determination (R^2), and the probability plot correlation coefficient (PPCC). The results indicate that the best performance can be obtained by the Gamma distribution.

Keywords: Mathematica, Probability distribution, Tabriz synoptic station, Wind energy, Wind speed frequency

1. Introduction

Nowadays more and more countries in the world have had to seek help to renewable resources, such as wind, solar and geothermal, not only to meet the increasing energy demand, but also for environmental reasons. Wind energy production is an attractive and feasible method employing renewable energy source without emission of pollutants.

Wind energy can be considered a green power technology as it has only minor impacts on the environment. Currently, wind energy is one of the fastest developing renewable energy source technologies across the globe. Countries all around the world are doing a thorough research on the specific energy that comes out by the uneven heating by the sun. Wind analysis gives remarkable information to researches involved in renewable energy studies.

Knowledge of the statistical properties of wind speed is essential for predicting the energy output of a wind energy conversion system. Because of the high variability in space and time of wind energy, it is important to verify that the analyzing method used in measuring wind data will yield the estimated energy collected that is close to the actual energy collected. The wind speed distribution, one of the wind characteristics, is of great importance not only for structural and environmental design and analysis, but also for the assessment of the wind energy potential and the performance of wind energy conversion system as well. For this reason, an accurate determination of probability distribution of wind speed values is very important in evaluating wind speed energy potential of a region. Wind energy potential can be determined by wind measurements of a certain investigation region depending on years. And also extreme wind speed frequency estimation is usually important in many fields of environmental studies such as climatology, hydrology, developing wind energy facilities, agricultural management, and structure designing (Lopez, 1998; Gomes *et al.*, 2003).

Many investigators have tried to fit different frequency distributions to wind data. Al Buhairi and Mahyoub (2006) found that, in recent years, many efforts have been made to construct an adequate model for the wind speed frequency distribution. In the literature, the Weibull distribution is commonly used in the practical studies related to the wind energy modeling (Steven, 1979; Toure, 2005; Zhou *et al.*, 2006). Auwera *et al.* (1980) used Weibull three-parameter model for estimating mean wind power densities. Weisser (2003) analyzes wind energy analysis of Grenada using the Weibull density function. Lun and Lan (2000) studied Weibull parameters using long-term wind observations. Seguro *et al.* (2000) estimated the parameters of Weibull wind speed distributions for wind energy analysis. Celik (2003) used Weibull distribution to estimate wind energy output of large- and small-scale turbines. Rehman *et al.* (1994) used Weibull parameters for wind speed distribution in Saudi Arabia.

Recently, Pandey and Sutherland (2003) fitted generalized Pareto distribution to peak-over-threshold extreme wind speed through bootstrapping. Holmes and Moriarty (1999) also suggested generalized Pareto distribution to fit the extreme wind speed in Australia. Recently, Zaharim *et al.* (2009) fitted gamma, lognormal and Weibull distributions to wind speed data in the east coast of Malaysia. The numerical and graphical results obtained from the specific statistics showed that the Weibull and gamma distributions, whose parameters are estimated using the maximum likelihood principle, provide the best fits for the year 2005 and 2006, respectively.

This study aims to find the best probability distribution to the annual maximum wind speed of Tabriz, Iran and is the first study in Iran that investigates the best probability distribution of maximum wind speed. The study uses the *Mathematica* code developed for (Ghorbani *et al.*, 2010); see <http://users.utu.fi/ruskeepa/>.

2. Methodology

2.1 Probability density functions

Four different probability distributions are considered in this study, some of which are very widely used in climatology frequency analysis. These distributions and their probability density functions are presented in Table 1. Only the Gamma, lognormal, truncated extreme value and truncated logistic distributions are used.

Since wind speed is always non-negative, it is more realistic to truncate the density functions so that they yield a domain that consists of only non-negative values. The truncation is done by simply dividing the original density function by a suitable constant, to make the integral of the truncated density function equal to one; the constant is given by $P(X \geq 0) = 1 - F(0)$, where $F(x)$ is the cumulative distribution function of the original distribution. The CDF of the truncated distribution is then calculated by integration the truncated density from 0 to x . [Note: In the truncated Pearson type III density function, the term $\Gamma(\alpha, -\varepsilon/\beta)$ is the value of the incomplete gamma function $\Gamma(\alpha, z) = \int_z^\infty t^{\alpha-1} e^{-t} dt$].

If $\varepsilon = 0$, then this distribution reduces to the gamma distribution. The extreme value is the limiting distribution for the largest values in large samples drawn from a variety of distributions, including Normal, Exponential, and Weibull distributions.

Table 1. Probability distributions and their density functions

2.2 Method of parameters estimating

2.2.1 Maximum-likelihood estimation (MLE)

In statistics, maximum-likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters.

The method of maximum likelihood corresponds to many well-known estimation methods in statistics. In general, for a fixed set of data and underlying statistical model, the method of maximum likelihood selects

values of the model parameters that produce a distribution that gives the observed data the greatest probability (i.e., parameters that maximize the likelihood function). Maximum-likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems. However, in some complicated problems, difficulties do occur: in such problems, maximum-likelihood estimators are unsuitable or do not exist.

Suppose there is a sample x_1, x_2, \dots, x_n of n iid observations, coming from a distribution with an unknown pdf $f_0(\cdot)$. It is however surmised that the function f_0 belongs to a certain family of distributions $\{f(\cdot|\theta), \theta \in \Theta\}$, called the parametric model, so that $f_0 = f(\cdot|\theta_0)$. The value θ_0 is unknown and is referred to as the "true value" of the parameter. It is desirable to find some estimator $\hat{\theta}$ which would be as close to the true value θ_0 as possible. Both the observed variables x_i and the parameter θ can be vectors.

To use the method of maximum likelihood, one first specifies the joint density function for all observations. For an iid sample this joint density function will be:

$$f(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta) \cdot f(x_2 | \theta) \dots f(x_n | \theta) \tag{1}$$

Now we look at this function from a different perspective by considering the observed values x_1, x_2, \dots, x_n to be fixed "parameters" of this function, whereas θ will be the function's variable and allowed to vary freely. From this point of view this distribution function will be called the likelihood:

$$\ell(\theta|x_1, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \tag{2}$$

2.2.2 Method of moments (MOM)

The method of moments (MOM) is a technique for constructing estimators of the parameters that is based on matching the sample moments with the corresponding distribution moments. First, let:

$$\mu_i(a) = E(X_i | a) \tag{3}$$

denote the i 'th moment of X about 0. Note that we are emphasizing the dependence of these moments on the vector of parameters a . Note also that $\mu_1(a)$ is just the mean of X , which we usually denote by μ . Next, let

$$M_i(X) = (X_{1i} + X_{2i} + \dots + X_{ni}) / n \tag{4}$$

denote the i 'th sample moment. Note that we are emphasizing the dependence of the sample moments on the sample X . Note also that $M_1(X)$ is just the ordinary sample mean, which we usually just denote by M_n .

To construct estimators W_1, W_2, \dots, W_k for our unknown parameters a_1, a_2, \dots, a_k , respectively, we attempt to solve the set of simultaneous equations:

$$\mu_1(W_1, W_2, \dots, W_k) = M_1(X_1, X_2, \dots, X_n) \tag{5}$$

$$\mu_2(W_1, W_2, \dots, W_k) = M_2(X_1, X_2, \dots, X_n) \tag{6}$$

...

$$\mu_k(W_1, W_2, \dots, W_k) = M_k(X_1, X_2, \dots, X_n) \tag{7}$$

for W_1, W_2, \dots, W_k in terms of X_1, X_2, \dots, X_n . Note that we have k equations in k unknowns, so there is hope that the equations can be solved.

2.3 Comparison of estimated probability density functions

Many methods are available for estimating the parameters of the above distributions, such as least-squares, maximum likelihood, moments, weighted moments, linear moments, and entropy. Extensive details of these methods are already available in the literature (Singh, 1996) and, therefore, are not reported here. In this study, only two of these methods are employed: maximum likelihood estimation and the method of moments.

There is no specific reason for preferring these two methods against the others, except that they are simple and also sufficient for the purpose of this study. They are neither treated as superior to the other methods nor any effort is made compare with them.

The probability density functions thus fitted are compared using quantiles. Assuming that there are n number of observations, Cunnane's plotting positions are first calculated as: $p_i = (i-0.4)/(n+0.2)$ for $i=1, \dots, n$, where i is the order of the i th observation arranged in ascending order and p_i is the probability of non-exceedance of the i th observation estimated by the Cunnane's plotting position formula. For each of the density functions, the p_i -quantiles, given by X_{pi} , $i=1, \dots, n$, are calculated. These quantiles are then compared with the observed values, denoted as X_i , the i th ordered value. Three statistical indicators are used to compare the computed and the

observed quantiles (O'Donnell, 1985): (1) the root mean square error (RMSE), which is the square root of $\sum_{i=1}^n (X_{pi} - X_i)^2 / n$;

(2) the coefficient of determination (R^2), which is the coefficient square of the correlation between the computed and the observed quantiles; and (3) the probability plot correlation coefficient (PPCC). The test uses the correlation r between the ordered X_i . The PPCC is a test statistic to measure linearity of the probability plot.

Critical values have been obtained for the normal and Gumbel distributions by Vogel (1986). The test has also been developed for the Weibull and the uniform distribution by Vogel & Kroll (1989), and for the extreme value distributions by Chowdhury *et al.* (1991). The critical values obtained for the normal distribution can also be used for the Log-normal distribution.

The correlation coefficient test statistic, r , is calculated between the ordered observed values and the inversed values of the cumulative distribution function of the fitted distribution. If the observed values conform to the applied distribution, the r statistic should be greater than the critical value for the selected significance level. The equations that give unbiased estimates of the inverse values, have been obtained by different researchers. The r test statistic is defined as

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(m_i - \bar{m})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (m_i - \bar{m})^2}} \quad (8)$$

in which x_i ($x_1 \leq \dots \leq x_i \leq \dots \leq x_n$) is the ordered i th value and \bar{X} is the mean of observed values. Also,

$$m_i = F_x^{-1}(p_i) \quad (9)$$

in which $F(\cdot)$ is the cumulative probability value and p_i is the value estimated by unbiased equations for different probability distributions.

2.4 Study area and data analysis

In this study, wind speed frequency analysis is performed for Tabriz synoptic station in Iran. The station is located in Iran north-west $38^\circ 4' N$, $46^\circ 17' E$ with elevation of 1364 meters above sea level. Figure 1 shows the geographical location of the study area. The data considered for the wind speed frequency analysis are the annual maximum wind values.

Figure 2 shows the variations of wind values for the Tabriz synoptic station (in the above order) and Table 2 presents some statistical parameters of wind speed. Wind speed frequency analysis is carried out by using four distribution functions and with the two parameter estimation methods (i.e. MLE and MOM).

Figure 1. The geographical location of the study area

Figure 2. Maximum annual wind speed at Tabriz station

Table 2. Some statistical parameters of wind speed (km/hr) data used

3. Results and Discussions

3.1 Distribution parameters

Table 3 shows the estimated parameters of the distributions when the estimation is done with the maximum likelihood method and the method of moments.

Table 3. Estimated values of parameters

Three goodness-of-fit methods including RMSE, R^2 , and PPCC are considered to select the best distribution. For the synoptic station, the results for selected distributions are presented in Table 4 for the maximum likelihood estimation and for the method of moments, respectively.

Table 4. Performance evaluation for selected distributions

With two methods, the Gamma distribution is better than the other distributions according to all the three criteria (RMSE, R^2 , and PPCC) and by both methods (MLE and MOM). Based on this result, it may be inferred that the Gamma is generally suitable for synoptic stations. Thus, the Gamma distribution may be suggested as an appropriate distribution for synoptic Tabriz station, and possibly for other Iranian conditions.

3.2 Comparison of distributions

With the maximum Likelihood estimation method, the best distribution is as follow:

Gamma (PPCC, R^2 , RMSE)

Figure 3 presents the best estimated density function (chosen based on RMSE, R^2 , and PPCC) for the Tabriz station obtained with these two parameter estimation methods, the best curve(s) is presented.

Figure 3. Histograms and the estimated probability density functions

Based on these results, it may be inferred that the Gamma distribution is generally suitable for Tabriz synoptic station, when the maximum likelihood method used for parameter estimation. Also with the method of moments, Gamma distribution seem more suitable. Taking these collectively, the Gamma distribution may be suggested as an appropriate for the Tabriz synoptic station, and possibly for other Iranian stations, although caution needs to be exercised in making such a generalization.

Rajabi and Modarres (2008) Estimated maximum wind speed is an essential task in many fields of environmental and engineering risk analysis. This study used prevalent westerly annual maximum wind speeds for the period of 1983-1998 for East Isfahan station in Isfahan Province, Iran. The frequency analysis of AM data wind speeds obtained by averaging the wind data over some chosen averaging periods showed that extreme value Type I distribution is the best distribution for 15, 30, 60 and 120 min wind durations.

Ghorbani *et al.* (2010) studied analyzes flood frequencies using discharge data from 6 gaging stations in the Aji River basin in Iran. The results also indicate that, among the 18 different distribution, the inverse gamma distribution is the most appropriate for the Aji River basin, followed by the inverse Gaussian distribution.

Morgan *et al.* (2011) studied on offshore wind farms. They show that the widely-accepted Weibull distribution provides a poor fit to the distribution of wind speeds when compared with more complicated models. They compare distributions in terms of three different metrics: probability plot R^2 , estimates of average turbine power output, and estimates of extreme wind speed. While the Weibull model generally gives larger R^2 than any other 2-parameter distribution, the bimodal Weibull, Kappa, and Wakeby models all show R^2 values significantly closer to 1 than the other distributions considered (including the Weibull), with the bimodal Weibull giving the best fits. The Kappa and Wakeby distributions fit the upper tail (higher wind speeds) of a sample better than the bimodal Weibull, but may drastically over-estimate the frequency of lower wind speeds. Because the average turbine power is controlled by high wind speeds, the Kappa and Wakeby estimate average turbine power output very well, with the Kappa giving the least bias and mean square error out of all the distributions. The 2-parameter Lognormal distribution performs best for estimating extreme wind speeds, but still gives estimates with significant error.

Table 5 show distribution parameters and wind speed exceeding a given value (i.e. quantiles) with the given probability for MLE and MOM, respectively.

Table 5. Distribution parameters and wind speed exceeding a given value with a given probability.

3.3 Conclusion

In this study, wind speed frequency analysis was performed for Iranian conditions. Maximum annual wind speed values observed at Tabriz synoptic station were studied. Four different probability distributions were fitted, and the method of maximum likelihood and the method of moments were used for parameter estimation. This study is also the first one where the software *Mathematica* was used for performing any type of wind speed frequency analysis. The results also indicate that, among the four different distributions, the Gamma distribution is the most appropriate for the Tabriz station.

The present study has important implications for wind speed frequency analysis for Iran in particular, and for regional climatology in general. Further, the use of *Mathematica* provides a new dimension to the wind speed frequency analysis. With the many challenges faced in using the existing methods (often due to difficulties in calculations) for the selection of the most appropriate probability distribution for a given region, the symbolic, numerical, and graphical capabilities of *Mathematica* together with its flexibility can go a long way. Future work will focus on advancing the use of *Mathematica* towards developing a more generalized and flexible framework for wind speed frequency analysis, details of which will be reported elsewhere.

References

- Auwers, V., Meyer, L. F., & Malet, L. M. (1980). The use of Weibull three-parameter model for estimating mean wind power densities. *Journal of Apply Meteorology*, 19, 819-825. [http://dx.doi.org/10.1175/1520-0450\(1980\)019<0819:TUOTWT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1980)019<0819:TUOTWT>2.0.CO;2)
- Buhairi, Al, & Mahyoub, H. (2006). A statistical analysis of wind speed data and an assessment of wind energy potential in taiz-yemen. *Ass. Univ. Bull. Environ. Res.*, 9, 21-33.

- Celik, A. N. (2003). Energy output estimation for small-scale wind power generators using Weibull-representative wind data. *Journal of Wind Engineering and Industrial Aerodynamics*, 91, 693-707. [http://dx.doi.org/10.1016/S0167-6105\(02\)00471-3](http://dx.doi.org/10.1016/S0167-6105(02)00471-3)
- Chowdhury, J. U., Stedinger, J. R., & Lu, L. H. (1991). Goodness-of-fit tests for regional generalized extreme value flood distributions. *Journal of Water Resources Research*, 27, 1765-1776. <http://dx.doi.org/10.1029/91WR00077>
- Ghorbani, M. A., Ruskeepää, H., Singh, V. P., & Sivakumar, B. (2010). Flood frequency analysis using *Mathematica*. *Turkish J. Eng. Env. Sci.*, 34, 171-188.
- Gomes, L., Arrue, J. L., Lopez, M. V., Sterk, G., Richard, D., Gracia, R., Sabre, M., Gaudichet, A., & Frangi, J. P. (2003). Wind erosion in a semiarid agricultural area of Spain: the WELSONS project catena. *Catena*, 52, 235-256. [http://dx.doi.org/10.1016/S0341-8162\(03\)00016-X](http://dx.doi.org/10.1016/S0341-8162(03)00016-X)
- Holmes, J. D., & Moriarty, W. W. (1999). Application of the generalised Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83, 1-10. [http://dx.doi.org/10.1016/S0167-6105\(99\)00056-2](http://dx.doi.org/10.1016/S0167-6105(99)00056-2)
- Lopez, M. V. (1998). Wind erosion in agricultural soils: an example of limited supply of particles available for erosion. *Catena*, 33, 17-28. [http://dx.doi.org/10.1016/S0341-8162\(98\)00064-2](http://dx.doi.org/10.1016/S0341-8162(98)00064-2)
- Lun, I. Y. F., & Lan, J. C. (2000). A study of Weibull parameters using long-term wind observations. *Renewable Energy*, 20, 145-153. [http://dx.doi.org/10.1016/S0960-1481\(99\)00103-2](http://dx.doi.org/10.1016/S0960-1481(99)00103-2)
- Morgan, E.C., Lackner, M., Vogel, R. M., & Baise, L.G. (2011). Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52, 15-26. <http://dx.doi.org/10.1016/j.enconman.2010.06.015>
- O'Donnell, T. (1985). A direct three-parameter Muskingum procedure incorporating lateral inflow. *Journal of Hydrology Science*, 30, 494-495.
- Pandey, M. D., & Sutherland, H. J. (2003). Probabilistic analysis of list data for the estimation of extreme design loads for wind turbine components. A collection of the 2003 ASME Wind.
- Rajabi M. R., & Modarres, R. (2008). Extreme value frequency analysis of wind data from Isfahan, Iran. *Journal of Wind Engineering and Industrial Aerodynamics*, 96, 78-82. <http://dx.doi.org/10.1016/j.jweia.2007.03.005>
- Rehman, S., Halawan, T. O., & Husain, T. (1994). Weibull parameters for wind speed distribution in Saudi Arabia. *Solar Energy*, 53, 473-479. [http://dx.doi.org/10.1016/0038-092X\(94\)90126-M](http://dx.doi.org/10.1016/0038-092X(94)90126-M)
- Seguro, J. V., & Lambert, T. W. (2000). Estimation of the parameters of Weibull wind speed distributions for wind energy analysis. *Journal of Wind Engineering and Industrial Aerodynamics*, 85, 75-84. [http://dx.doi.org/10.1016/S0167-6105\(99\)00122-1](http://dx.doi.org/10.1016/S0167-6105(99)00122-1)
- Singh V. P. (1996). *Entropy-based parameter estimation in hydrology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Stevens, D. E. (1979). Vorticity, momentum and divergence budgets of synoptic-scale wave disturbances in the tropical eastern Atlantic. *Mon. Wea. Rev.*, 107, 535-550. [http://dx.doi.org/10.1175/1520-0493\(1979\)107<0535:VMADBO>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1979)107<0535:VMADBO>2.0.CO;2)
- Toure, S. (2005). Investigations on the eigen-coordinates method for the 2-parameter Weibull distribution of wind speed. *Renewable Energy*, 30, 511-521. <http://dx.doi.org/10.1016/j.renene.2004.07.007>
- Vogel, R. M. (1986). The probability plot correlation coefficient test for the normal, lognormal and Gumbel distributional hypotheses. *Journal of Water Resources Research*, 22, 587-590. <http://dx.doi.org/10.1029/WR022i004p00587>
- Vogel, R. M., & Kroll, C. N. (1989). Low-flow frequency analysis using probability plot correlation coefficients. *J. Wat. Resour. Planning and Manag.*, 115, 338-357. [http://dx.doi.org/10.1061/\(ASCE\)0733-9496\(1989\)115:3\(338\)](http://dx.doi.org/10.1061/(ASCE)0733-9496(1989)115:3(338))
- Weisser, D. (2003). A wind energy analysis of Grenada: an estimation using the Weibull density function. *Renewable Energy*, 28, 1803-1812. [http://dx.doi.org/10.1016/S0960-1481\(03\)00016-8](http://dx.doi.org/10.1016/S0960-1481(03)00016-8)
- Zaharim, A., Razali, A. M., Abidin, R. Z., & Sopian K. (2009). Fitting of statistical distributions to wind speed data in Malaysia. *European Journal of Scientific Research*, 26, 6-12.

Zhou, J., Xu, F., Wang, T., Cao, A., & Yin, C. (2006). Cenozoic deformation history of the Qaidam Basin, NW China: Results from cross-section restoration and implications for Qinghai-Tibet Plateau tectonics. *Earth and Planetary Science Letters*, 243, 195-210. <http://dx.doi.org/10.1016/j.epsl.2005.11.033>

Table 1. Probability distributions and their density functions

| Distribution | PDF | Assumption | Domain |
|-----------------------------------|---|---|----------------------------|
| <i>Exponential</i> | $\lambda e^{-\lambda(x-\gamma)} \lambda e^{-\lambda(x-\lambda)}$ | $\gamma > 0, \lambda > 0$ | $x > \gamma$ |
| <i>Frechet</i> | $\frac{c\alpha^c}{x^{c+1}} e^{-\left(\frac{\alpha}{x}\right)^c}$ | $c > 0, \alpha > 0$ | $x > 0$ |
| <i>Gamma</i> | $\frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}}$ | $\alpha > 0, \beta > 0$ | $x > 0$ |
| <i>Generalized Pareto</i> | $\frac{1}{\sigma} \left(1 + k \frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}-1}$ | $k > 0, \sigma > 0, 0 < \mu < \frac{\sigma}{k}$ | $x > \mu$ |
| <i>Inverse Gamma</i> | $\frac{\beta^\alpha}{\Gamma(\alpha)x^{\alpha+1}} e^{-\frac{\beta}{x}}$ | $\alpha > 0, \beta > 0$ | $x > 0$ |
| <i>Inverse Gaussian</i> | $\sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda}{2\mu^2} \frac{(x-\mu)^2}{x}}$ | $\mu > 0, \lambda > 0$ | $x > 0$ |
| <i>Kumaraswamy</i> | $\frac{pq}{b} \left(\frac{x}{b}\right)^{p-1} \left(1 - \left(\frac{x}{b}\right)^p\right)^{q-1}$ | $p > 0, q > 0, b > 0$ | $0 < x < b$ |
| <i>Log-normal</i> | $\frac{1}{\sqrt{2\pi}\alpha x} e^{-\frac{1}{2} \left(\frac{\log(x)-\mu}{\sigma}\right)^2}$ | $\sigma > 0$ | $x > 0$ |
| <i>Log-Pearson Type III</i> | $\frac{1}{\Gamma(\alpha)\beta} \left(\frac{x-\varepsilon}{\beta}\right)^{\alpha-1} e^{-\frac{x-\varepsilon}{\beta}}$ | $\alpha > 0, \beta > 0, \varepsilon < 0$ | $x > 0, x > e^\varepsilon$ |
| <i>Maxwell</i> | $\frac{\sqrt{2}x^2}{\sqrt{\pi}\sigma^3} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ | $\sigma > 0$ | $x > 0$ |
| <i>Rayleigh</i> | $\frac{x}{\sigma^2} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ | $\sigma > 0$ | $x > 0$ |
| <i>Truncated Cauchy</i> | $\frac{1}{b\pi} \left(1 + \left(\frac{x-a}{b}\right)^2\right)^{-1} \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{a}{b}\right)\right)^{-1}$ | $b > 0$ | $x > 0$ |
| <i>Truncated Extreme Value</i> | $\frac{1}{\beta} \left(1 - e^{-\frac{x}{\beta}}\right)^{-1} e^{-\frac{x-\alpha}{\beta} - e^{-\frac{x-\alpha}{\beta}}}$ | $\alpha > 0, \beta > 0$ | $x > 0$ |
| <i>Truncated Gumbel</i> | $\frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}} e^{-\frac{x-\alpha}{\beta} - e^{-\frac{x-\alpha}{\beta}}}$ | $\alpha > 0, \beta > 0$ | $x > 0$ |
| <i>Truncated Logistic</i> | $\frac{1 + e^{-\frac{x-\mu}{\beta}}}{\beta} e^{-\frac{x-\mu}{\beta}} \left(1 + e^{-\frac{x-\mu}{\beta}}\right)^{-2}$ | $\mu > 0, \beta > 0$ | $x > 0$ |
| <i>Truncated Normal</i> | $\frac{\sqrt{2}}{\sigma\sqrt{\pi}} \left(1 + \operatorname{erf}\left(\frac{\mu}{\sigma\sqrt{2}}\right)\right)^{-1} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ | $\sigma > 0$ | $x > 0$ |
| <i>Truncated Pearson Type III</i> | $\frac{1}{\beta\Gamma\left(\alpha, \frac{\varepsilon}{\beta}\right)} \left(\frac{x-\varepsilon}{\beta}\right)^{\alpha-1} e^{-\frac{x-\varepsilon}{\beta}}$ | $\alpha > 0, \beta > 0, \varepsilon < 0$ | $x > 0$ |
| <i>Weibull</i> | $\frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-\left(\frac{x}{\beta}\right)^\alpha}$ | $\alpha > 0, \beta > 0$ | $x > 0$ |

Table 2. Some statistical parameters of wind speed (km/hr) data used

| Statistical parameter | Value |
|--|---------|
| Number | 55 |
| Average (km/hr) | 74.619 |
| Max (km/hr) | 107.416 |
| Min (km/hr) | 50.004 |
| Skewness (km ² /hr ²) | 0.3184 |
| Standard Deviation (km/hr) | 13.143 |

Table 3. Estimated values of parameters

| Method | Distribution | Distribution | | | |
|--------|--------------|------------------|-------------------|-------------------------|--------------------|
| | | Gamma | Log-normal | Truncated extreme value | Truncated logistic |
| MLE | | $\alpha=32.9867$ | $\mu=3.68089$ | $\alpha=36.8822$ | $\mu=40.0452$ |
| | | $\beta=1.22143$ | $\delta=0.174956$ | $\beta=6.19048$ | $\beta=4.10404$ |
| MOM | | $\alpha=32.8331$ | $\mu=3.68112$ | $\alpha=37.1263$ | $\mu=40.2895$ |
| | | $\beta=1.22714$ | $\delta=0.173212$ | $\beta=5.48247$ | $\beta=3.87901$ |

Table 4. Performance evaluation for selected distributions

| Method | Distribution | PPCC | R ² | RMSE |
|--------|-------------------------|-----------------|-----------------|-----------------|
| MLE | Gamma | 0.992093 | 0.984248 | 0.892496 |
| | Log-normal | 0.991606 | 0.983282 | 0.911586 |
| | Truncated logistic | 0.983074 | 0.966435 | 1.33081 |
| | Truncated extreme value | 0.981376 | 0.963099 | 1.53706 |
| MOM | Gamma | 0.992095 | 0.984253 | 0.89016 |
| | Log-normal | 0.991638 | 0.983345 | 0.917314 |
| | Truncated logistic | 0.983054 | 0.966396 | 1.29925 |
| | Truncated extreme value | 0.981376 | 0.963099 | 1.35977 |

Table 5. Values that the wind speed exceeds with a given probability

| Method | Criteria | Best Distribution | Wind speed quantiles | | | | | | | |
|--------|----------------|-------------------|----------------------|---------|---------|---------|---------|---------|---------|---------|
| | | | P=0.5 | P=0.6 | P=0.7 | P=0.8 | P=0.9 | P=0.95 | P=0.99 | P=0.999 |
| MLE | PPCC | Gamma | 39.8845 | 41.6775 | 43.6543 | 46.0454 | 49.502 | 52.4814 | 58.3805 | 65.4833 |
| | RMSE | | | | | | | | | |
| | R ² | | | | | | | | | |
| MOM | PPCC | Gamma | 39.8826 | 41.6798 | 43.6613 | 46.0584 | 49.5239 | 52.5113 | 58.4268 | 65.5504 |
| | RMSE | | | | | | | | | |
| | R ² | | | | | | | | | |

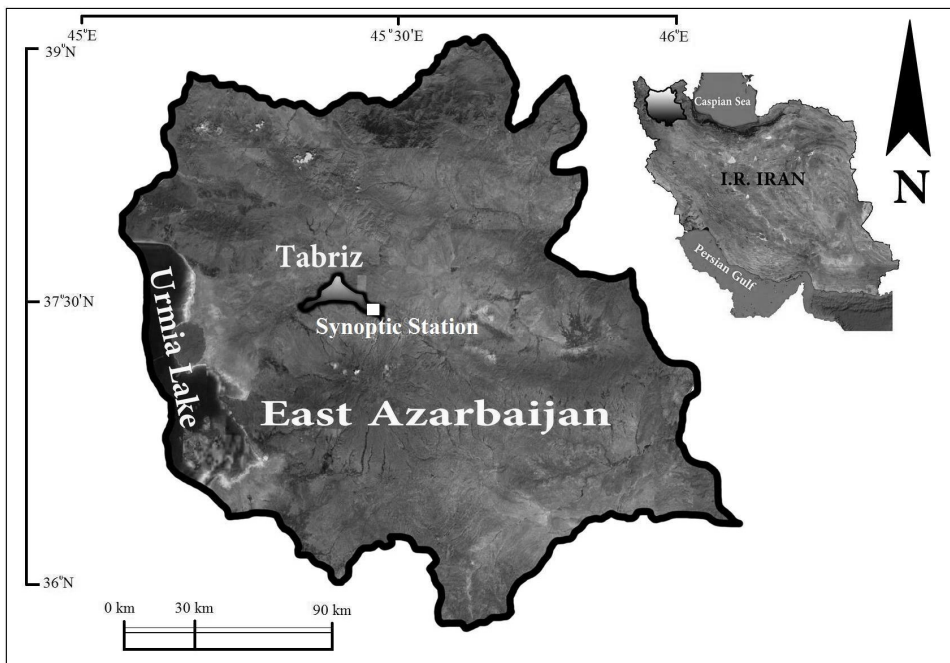


Figure 1. The geographical location of the study area

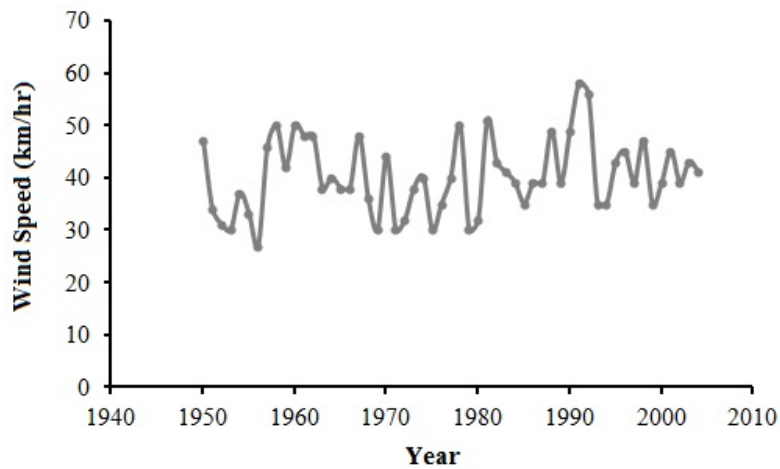


Figure 2. Maximum annual wind speed at Tabriz station

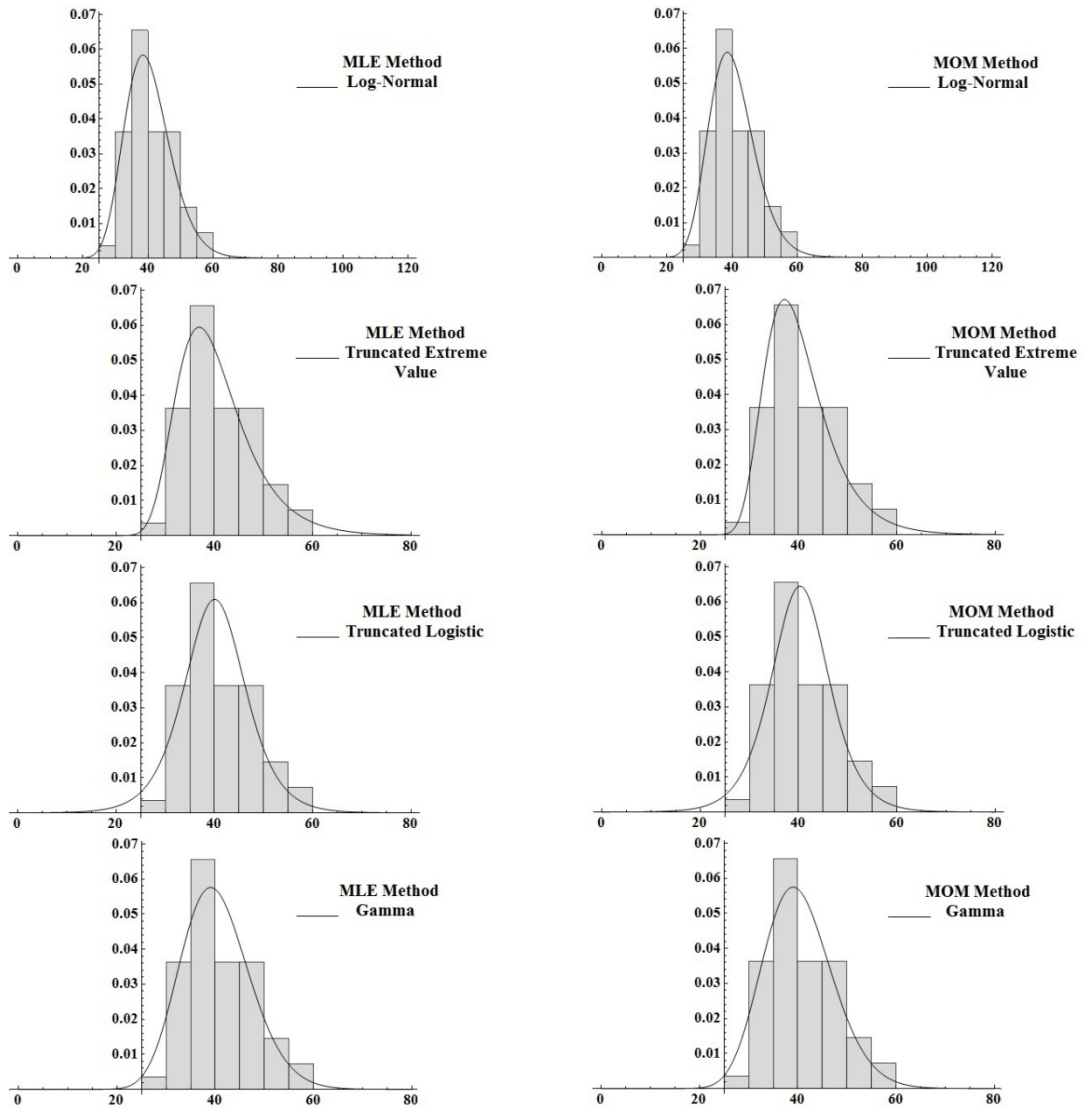


Figure 3. Histograms and the estimated probability density functions