The Annual Maximum Wind Speed at Pisa Airport in Italy

Paul Chiou, Weiwen Miao, and T. C. Ho

Abstract—Structural engineers use the extreme or fastest values of wind speed with return periods such as 25 years for structures having no human occupants or where there is a negligible risk to human life, 50 years for most permanent structures, and 100 years for structures with an unusually high degree of hazard of life and property in case of failure. The data of 41 annual maximum 10-minute average wind speeds at Pisa Airport in Italy from 1951 to 1991 were analyzed and modeled. Three extreme value models for the data were considered and compared. Subsequently, the required design value with a given return period of exceedance was obtained.

Index Terms—Anderson-Darling test, design value, Frechet distribution, general extreme value distribution, goodness-of-fit, Gumbel distribution, Kolmogorov-Smirnov test, return period.

I. INTRODUCTION

High wind speeds pose a threat to the integrity of structures such as air traffic control towers or wind turbines. An accurate estimation of the sustainability of extreme wind speeds is an important design factor in achieving an optimal balance between safety and cost of “over-design”. This kind of design problem arises in many engineering areas such as ocean engineering (wave height), hydraulics engineering (floods), structural engineering (earthquakes), and also in meteorology (tornadoes and rainfalls), fatigue strength (workloads), etc. [1]-[4]. All these applications have in common that the main interest is not the knowledge of the average behavior of the analyzed phenomena but the extreme behavior of them [5]. Typically, buildings are designed to resist a strong wind with a very long return period, such as 50 years or more. The design wind speed is determined from historical records using extreme value theory to predict future extreme wind speeds. When the probability law describing extreme wind speeds applies to homogeneous micro-meteorological conditions, one must consider initially the average time before using a specified probability law to represent wind data. If different sampling frequencies were used to collect the data, the whole sample must be adjusted to a unique averaging time such as a period of 10 minutes [6]. The statistical theory developed to deal with these problems and this type of data is known as extreme value theory. For example, the extreme wind speed estimates are used to determine critical design loads which the turbine must withstand during its lifetime. According to the International Standard IEC 61400-1 for wind turbine generator systems, the extreme wind speed $V_{\text{ref}}$ is a basic parameter for wind turbine classification and therefore strongly related to design of wind turbines. The parameter $V_{\text{ref}}$ termed as required design value is defined as the extreme 10-min average wind speed with a recurrence period of 50 years. In general $V_{\text{ref}}$ has to be determined statistically on the basis of on-site measurements [5].

The objectives of this study are to (i) model the annual maximum wind speed at Pisa Airport with the Gumbel, Frechet and general extreme value distributions, (ii) use the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests to compare the observed and theoretical cumulative frequencies as predicted by the three distributions, (iii) obtain required design values for given return periods of exceedance.

II. MATERIALS AND METHODS

A. Location and Measurements

Pisa is a city in Tuscany, Central Italy, on the right bank of the mouth of River Arno at Tyrrenhenian Sea (Fig. 1). It is the capital city of the Province of Pisa. Pisa Airport (43°41′02″N, 10°23′33″E) residing at an elevation of 6 feet (2 m) above mean sea level is an international airport located 1 km from the Pisa city centre. It is one of the two main airports in Tuscany. The airport was first developed for the military in the 1930 and 1940s. The airport had 4,067,012 passengers in 2010. The 10-minute average wind speeds at the airport were collected using anemometers from the year of 1951 to 1991. The data of 41 annual maximum 10-minute average wind speeds are summarized in Table I [6].

![Fig. 1. Location of Pisa in Italy.](image-url)
B. Model Formulations

The core of the extreme value theory is the study of the statistical behavior of $X = \max\{X_1, \ldots, X_n\}$, where $\{X_1, \ldots, X_n\}$ is a sequence of independent random variables having a common distribution function $F$. In applications, variables $X_i$ usually represent values of a process measured on a regular time-scale, as for example the 10 minutes average (or maximum) wind speeds. Then $X$ is the maximum of the observed process over $n$ time units. In well-behaved climates (with a stationary distribution of extreme winds), the annual maximum wind speed is often represented by the Gumbel distribution [7], [8]. An alternative to the Gumbel distribution is the Frechet distribution [9] or the general (generalized) extreme value (GEV) distribution [10]-[16].

1) \textit{Gumbel Distribution}

There are three types of extreme value distributions. The Type I distribution was extensively developed and applied to the flood flows by Gumbel; therefore, it is often referred to as the Gumbel distribution [7], [8]. The cumulative distribution function (cdf) of the Gumbel random variable $X$ is given as

$$F(x) = \exp[-e^{-(x-\beta)/\alpha}], \quad -\infty < x < \infty,$$  

where $\beta$ is the location parameter and $\alpha$ is the scale parameter. The location parameter $\beta$ is the mode of the distribution. The scale parameter $\alpha$ is a measure of dispersion and it only depends on the variance of $X$. The parameter $\beta$ is a measure of location that depends on both the variance and the mean. The moment-generating function is found to be

$$M_X(t) = \exp(\beta t) \Gamma(1-\alpha t), \quad t < (1/\alpha).$$  

Hence, the mean and the variance of $X$ from (2) are given respectively as

$$\mu = E[X] = \beta + n \alpha,$$

and

$$\sigma^2 = \text{Var}[X] = \frac{\pi^2 \alpha^2}{6},$$

where $n$ denotes the Euler constant, approximately equal to 0.5772. The skewness coefficient is 1.1396, and the kurtosis coefficient is 5.40. If the first two moments of $X$ are known, the values of $\alpha$ and $\beta$ can be determined by the method of moments from the mean $\mu$ and standard deviation $\sigma$ of the distribution $X$. From the above two equations, one obtains

$$\alpha = \frac{\sqrt{6}}{\pi} \sigma,$$  

and

$$\beta = \mu - n \alpha = \mu - \frac{n \sqrt{6}}{\pi} \sigma.$$

The two equations can be used to estimate $\alpha$ and $\beta$ if a finite sample of the values taken by $X$ is available, such as the annual maximum wind speeds for a period of $n$ years. To compute the estimated values of $\alpha$ and $\beta$, one must estimate the mean $\mu$ and standard deviation $\sigma$ of the population based on the sample.

2) \textit{Frechet Distribution}

The Frechet distribution is a particular form of Type II extreme value distribution. It was first developed and applied to the flood flows by Frechet [9]. The cumulative distribution function of the Frechet distribution $X$ is given as

$$F(x) = \exp[-(x/\eta)^{-\alpha}], \quad x > 0,$$  

where $\eta > 0$ denotes a scale parameter and $\alpha > 0$ is a shape parameter. The moments of order $r$ are given by

$$E[X^r] = \eta^r \Gamma(1-r/\alpha), \quad r < \alpha.$$  

Consequently, from (6) we have

$$\mu = E[X] = \eta \Gamma(1-1/\alpha), \quad \theta > 1,$$

and

$$\sigma^2 = \text{Var}[X] = \eta^2 \left[\Gamma(1-2/\alpha) - \Gamma^2(1-1/\alpha)\right], \quad \theta > 2.$$  

Since the shape parameter $\theta$ only depends on the square of the coefficient of variation $V$ given by

$$V^2 = \frac{\sigma^2}{\mu^2} = \frac{\Gamma(1-2/\theta)}{\Gamma^2(1-1/\theta)} - 1, \quad \theta > 2,$$

and

$$\eta = \frac{\mu}{\Gamma(1-1/\theta)}, \quad \theta > 1,$$

if the first two moments of $X$ exist and are known, the values of the parameters $\eta$ and $\theta$ can be determined from the mean $\mu$ and coefficient of variation $V$ of the distribution $X$. After $V$ is estimated as the ratio of the sample standard deviation to the sample mean, the value of $\theta$ in (7) can be solved via numerical iterations. Then, substituting the sample mean and the estimate of $\theta$ into (8), one can estimate the value of scale parameter $\eta$.

3) \textit{General Extreme Value Distribution}

The general extreme value (GEV) distribution was introduced by Jenkinson [17], [18] to identify the frequency distribution of the largest values of meteorological data when the limiting form of the extreme value distribution is

<table>
<thead>
<tr>
<th>Year</th>
<th>m/s</th>
<th>Year</th>
<th>m/s</th>
<th>Year</th>
<th>m/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>15.43</td>
<td>1965</td>
<td>18.00</td>
<td>1979</td>
<td>16.46</td>
</tr>
<tr>
<td>1955</td>
<td>22.03</td>
<td>1969</td>
<td>16.46</td>
<td>1983</td>
<td>15.95</td>
</tr>
<tr>
<td>1956</td>
<td>18.52</td>
<td>1970</td>
<td>13.43</td>
<td>1984</td>
<td>13.89</td>
</tr>
<tr>
<td>1958</td>
<td>18.00</td>
<td>1972</td>
<td>13.37</td>
<td>1986</td>
<td>19.05</td>
</tr>
<tr>
<td>1961</td>
<td>18.00</td>
<td>1975</td>
<td>11.32</td>
<td>1989</td>
<td>16.98</td>
</tr>
<tr>
<td>1964</td>
<td>13.89</td>
<td>1978</td>
<td>14.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
unknown. The cdf of the GEV distribution is given by
\[ F(x) = \exp\left\{-[1 + k(x - \varepsilon) / \alpha]^{-1/k}\right\}, \]
where \(\varepsilon\) denotes a location parameter, \(\alpha\) is a scale parameter, and \(k\) is the shape parameter. Note that for \(k > 0\), the GEV represents a Type II extreme value distribution and it is defined only for \(x > (\varepsilon - \alpha/k)\); in the opposite case, for \(k < 0\), this model becomes the Type III distribution and it is defined only for \(x < (\varepsilon - \alpha/k)\). If \(k = 0\), the GEV distribution corresponds to the Gumbel distribution in (1). The moments of order \(r\) exist only if \(k < 1/r\). The mean and variance of the GEV distribution are given by
\[ \mu = E[X] = \varepsilon + (\alpha / k)[\Gamma(1 - k) - 1], \]
as \(k < 1\), and
\[ \sigma^2 = \text{Var}[X] = (\alpha / k)^2[\Gamma(1 - 2k) - \Gamma^2(1 - k)], \]
for \(k < 0.5\), respectively. Therefore, the mean is not defined for \(k \geq 1\), and the variance is not defined for \(k \geq 0.5\). The coefficient of skewness is given by
\[ \gamma = \frac{\Gamma(1 - 3k) - 3\Gamma(1 - k)\Gamma(1 - 2k) + 2\Gamma^2(1 - k)}{[\Gamma(1 - 2k) - \Gamma^2(1 - k)]^{3/2}}, \]
for \(0 < k < \frac{1}{3}\), and it is given by
\[ \gamma = \frac{-\Gamma(1 - 3k) + 3\Gamma(1 - k)\Gamma(1 - 2k) - 2\Gamma^2(1 - k)}{[\Gamma(1 - 2k) - \Gamma^2(1 - k)]^{3/2}}, \]
for \(k < 0\). Note that the shape parameter depends only on the coefficient of skewness if the third moment exists.

If the first three moments of \(X\) exist and are known, then the values of the three parameters \(\varepsilon, \alpha, \text{ and } k\) can be determined from the mean, the variance, and the skewness coefficient of the data. Since the shape parameter \(k\) only depends on the coefficient of skewness for \(k < 1/3\), one can first solve for \(k\) by substituting the sample skewness coefficient into (10) or (11). From the expression of variance of \(X\), the scale parameter is found as
\[ \alpha = \sqrt{k^2\sigma^2 / [\Gamma(1 - 2k) - \Gamma^2(1 - k)]}, \]
To estimate the parameter \(\alpha\), the estimate for \(k\) and sample standard deviation are substituted for \(k\) and \(\sigma\) into (12). Finally, from the expression of \(E[X]\), the location parameter is given by
\[ \varepsilon = \mu - (\alpha / k)[\Gamma(1 - k) - 1], \]
and it can be estimated by substituting the sample mean and the estimates of \(\alpha\) and \(k\) for \(\mu\) and \(\alpha\) respectively.

C. Goodness of Fit

In assessing whether a given distribution is suited to a set of data, one can consider Kolmogorov-Smirnov and Anderson-Darling tests.

1) Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) goodness-of-fit test [19] is a nonparametric test that relates to the cdf rather than the probability density function (pdf) of a continuous variable. For a continuous variable \(X\), let \(x_{(1)}, x_{(2)}, \ldots, x_{(n)}\) represent the order statistics of a sample of size \(n\), that is, the values are arranged in ascending order. The empirical or sample distribution function \(F_n(x)\) is defined as
\[ F_n(x) = k / n, \quad x_{(k)} \leq x < x_{(k+1)}, \]
and
\[ F_n(x) = 0, \quad x < x_{(1)}; \quad F_n(x) = 1, \quad x \geq x_{(n)}. \]
Let \(F_0(x)\) denote a completely specified theoretical continuous cdf. The null hypothesis \(H_0\) is that the true cdf of \(X\) is the same as \(F_0(x)\). The test criterion is the maximum absolute difference between \(F_n(x)\) and \(F_0(x)\), formally defined as
\[ D_n = \sup_x |F_n(x) - F_0(x)|. \]
If the value of the test statistic \(D_n\) is large, then the null hypothesis is rejected, and the critical values \(D_{\alpha}\) for large sample, say, \(n > 35\), are 1.3581\(n^{1/2}\) and 1.6276\(n^{1/2}\) for levels of significance \(\alpha = 0.05\) and 0.01 respectively.

2) Anderson-Darling Test

The Anderson-Darling (A-D) test [20] is devised to give heavier weighting to the tails of a distribution where unexpectedly high or low values called outliers are located. This is made possible if one divides the difference between the empirical and hypothetical cdf’s, \(F_n(x)\) and \(F_0(x)\), by
\[ \sqrt{F_n(x)[1 - F_0(x)]}. \]
Anderson and Darling [20] showed the test statistic becomes
\[ A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \ln\left\{ 1 - F_0(x_{(i)}) \right\} - \frac{\ln\left\{ 1 - F_0(x_{(n)}) \right\}}{1 - F_0(x_{(n)})}, \]
where \(x_{(1)}, x_{(2)}, \ldots, x_{(n)}\) are the rearranged observations in increasing order. For large values of \(A^2\), the null hypothesis is rejected, and the critical values for \(A^2\) are 2.492 and 3.857 for \(\alpha = 0.05\) and 0.01 respectively when sample size \(n > 10\).

D. Design Values and Return Periods

A return level with a return period of \(T = 1/p\) years is a high threshold \(x(T)\) (e.g., annual peak for the 10-minute average wind speeds) such that the probability, \(P(X > x(T))\), is \(p\). For example, if \(p = 0.02\), then the return period is \(T = 50\) years, and the aforementioned parameter of interest \(V(\text{ref})\) is the return level \(x(50)\) with return period of 50 years for the 10 minutes average wind speeds. Using (1), (5), and (9), the return level with return period of \(T\) years for the Gumbel, Frechet, and GEV models are respectively given as
\[ x(T) = \beta - \alpha \ln\left\{ \ln[1 / (T - 1)] \right\}, \]
\[ x(T) = \eta \ln[\ln[1 / (T - 1)]]^{1/\theta}, \]
and
\[ x(T) = \varepsilon + (\alpha / k) \left\{ \ln[\ln[1 / (T - 1)]] \right\}^{\frac{-1}{k}}, \]
Substituting the estimated values into the above equations for unknown parameters, one can predict the return level for return period of \(T\) years. The predicted return level is the estimated design value until next exceedance. The two common interpretations of a return level with a return period
of $T$ years are: (i) Waiting time: Average waiting time until next occurrence of event (exceedance) is $T$ years; (ii) Number of events: Average number of events occurring within a $T$-year time period is one.

III. RESULTS AND DISCUSSION

The mean, standard deviation, coefficient of variation, and coefficient of skewness for the annual maximum wind speed estimated from the 41-year record at Pisa Airport in Italy, are 16.37 m/s, 2.83 m/s, 0.1729, and 0.3515, respectively.

Gumbel Model: From (3), the value

$$\hat{\alpha} = \frac{\sqrt{6}}{\pi} \hat{\sigma} = 0.780 \times 2.83 = 2.21 \text{ m/s}$$

is the estimated scale parameter of the Gumbel distribution by the method of moments. From (4), the location parameter is estimated as

$$\hat{\beta} = \hat{\mu} - \eta \hat{\alpha} = 16.37 - 0.5772 \times 2.21 = 15.09 \text{ m/s}.$$  

To test the null hypothesis $H_0$ that the annual maximum wind speed has a Gumbel distribution as specified earlier, we use these estimates for distribution $F_n(x)$ assuming the behavior of wind speed at Pisa Airport remains stationary. The value of $\alpha=0.05$ as the level of significance is adopted. The critical value is 0.212, and therefore the null hypothesis is not rejected. On the other hand, the calculated value of $D_n=0.107$ is less than the critical value 0.212, and therefore the null hypothesis is not rejected. Substituting the estimates 2.21 and 15.09 for $\alpha$ and $\beta$ into (16) gives the estimated 50-year required design value $V(\text{ref})$ as

$$\hat{x}(50) = 15.09 - 2.21 \times \ln[\ln(50 / (50 - 1))] = 23.71 \text{ m/s},$$

and the estimated 100-year design value as

$$\hat{x}(100) = 15.09 - 2.21 \times \ln[\ln(100 / (100 - 1))] = 25.26 \text{ m/s}.$$  

Frechet Model: Using the method of moments, the estimate of shape parameter of the Frechet distribution can be obtained. From (7), we have

$$\frac{\Gamma(1-2/\hat{\theta})}{\Gamma'(1-1/\hat{\theta})} = 1 + 0.1729^2,$$

and the equation is solved by numerical iterations to obtain

$$\hat{\theta} = 8.25.$$  

Then, from (8) we obtain the estimate of $\eta$ as

$$\hat{\eta} = \frac{\hat{\mu}}{\Gamma'(1-1/\hat{\theta})} = \frac{16.37}{\Gamma'(1-8.25)} = 15.07.$$  

To test the null hypothesis $H_0$ that the annual maximum wind speed has a Frechet distribution as specified earlier with $\eta=15.07$ and $\theta=8.25$, we compute the observed value of $D_n$ and obtain $d_n=0.136$ for K-S test. The value of $d_n=0.136$ is less than the critical value 0.212, and therefore the null hypothesis is not rejected. On the other hand, the calculated value of $A^2$ for A-D test is 1.617 which is less than the critical value of 2.492 and the null hypothesis is also not rejected. Substituting the estimates 15.07 and 8.25 for $\eta$ and $\theta$ into (17) gives the estimated 50-year required design value $V(\text{ref})$ as

$$\hat{x}(50) = 15.07 \times \left[\ln(50 / (50 - 1))\right]^{-(1/8.25)} = 24.18 \text{ m/s},$$

and the estimated 100-year design value as

$$\hat{x}(100) = 15.07 \times \left[\ln(100 / (100 - 1))\right]^{-(1/8.25)} = 26.32 \text{ m/s}.$$  

GEV Model: To estimate the shape parameter of the GEV distribution by the method of moments, the sample skewness of 0.3515 is substituted into (10) or (11), which is then solved for $k$ by numerical iterations. Thus, the estimate of $k$ is obtained as

$$\hat{k} = -0.1726.$$  

Then, from (12) and (13) respectively we obtain the estimates of $\alpha$ and $\epsilon$ as

$$\hat{\alpha} = \sqrt{\left(\hat{k}^2 \hat{\sigma}^2\right)} / \Gamma(1-2\hat{k}) - \Gamma(1-\hat{k}) = 2.636,$$

and

$$\hat{\epsilon} = \hat{\mu} - (\hat{\alpha} / \hat{k}) \Gamma(1-\hat{k}) - 1 = 15.236.$$  

To test the null hypothesis $H_0$ that the annual maximum wind speed has a GEV distribution as specified earlier with $\epsilon=15.236, \alpha=2.636$, and $k=-0.1726$, we compute the observed value of $D_n$ and obtain $d_n=0.093$ for K-S test. The value of $d_n=0.093$ is less than the critical value 0.212, and therefore the null hypothesis is not rejected. On the other hand, the calculated value of $A^2$ for A-D test is 0.302 which is less than the critical value of 2.492 and the null hypothesis is also not rejected. Substituting the estimates 15.236, 2.636, and -0.1726 respectively for $\epsilon, \alpha,$ and $k$ into (18) gives the estimated 50-year required design value $V(\text{ref})$ as

$$\hat{x}(50) = 15.236 + (2.636 / -0.1726) \times \left[\ln(50 / (50 - 1))\right]^{0.1726} - 1 = 22.72 \text{ m/s},$$

and the estimated 100-year design value as

$$\hat{x}(100) = 15.236 + (2.636 / -0.1726) \times \left[\ln(100 / (100 - 1))\right]^{0.1726} - 1 = 23.60 \text{ m/s}.$$  

Fig. 2 shows the histogram of the 41 data values from the year of 1951 to 1991, and the fitted three distributions. 

![Fig. 2. Histogram of wind speed and fitted models.](image-url)
IV. CONCLUSION

We consider three distributions to model the annual maximum wind speed at Pisa Airport in Italy. The observed values of K-S and A-D goodness-of-fit test for the GEV distribution are respectively the lowest among the Gumbel, Frechet, and GEV distributions. The visual comparison of the three fitted distributions to the data in Fig. 2 also confirms that the GEV distribution seems to be a better model for the annual maximum wind speed at Pisa Airport than the Gumbel and Frechet models. Consequently, the required design values of $V(\text{ref}) = 22.72 \, \text{m/s}$ and $23.60 \, \text{m/s}$ should be adopted to predict the return level with return period of 50 years and 100 years respectively.

REFERENCES