Comparison of Estimators of Gumbel Distribution for Modelling Wind Speed Data

N. Vivekanandan and Dr. S.K. Roy

Abstract— Estimation of extreme wind speed potential at a region is of importance while designing tall structures such as cooling towers, stacks, transmission line towers, etc. Assessment of wind speed in a region can expediently be carried out by probabilistic modelling of historic wind speed data using an appropriate extreme value distribution. This paper illustrates the use of five parameter estimation methods of Gumbel distribution for modelling Hourly Maximum Wind Speed (HMWS) data recorded at Delhi and Visakhapatnam regions. Goodness-of-Fit (GoF) tests involving Anderson-Darling and Kolmogorov-Smirnov are used for checking the adequacy of fitting of the method to the recorded data. Root Mean Square Error (RMSE) is used for selection of a suitable method for determination of parameters of Gumbel distribution for modelling HMWS data. The results of GoF tests and RMSE show that order statistics approach is better suited for estimation of design wind speed for the regions under study.

Keywords— Anderson-Darling, Gumbel, Kolmogorov-Smirnov, Mean Square Error, Order Statistics, Wind Speed

I. INTRODUCTION

Wind forces, and their static and dynamic effects, need to be taken into account while designing buildings, structures and their components thereof. If a structure is tall and slender, the effect of wind on the structure can be critical. The distribution of wind speed is also important in determining the serviceability of buildings. The estimated wind speed at 1000-year (yr) return period is often used to arrive at the design load that a structure must withstand during its lifetime [1-2]. For arriving at such design values, a standard procedure is to analyse historic wind data over a period of time and arrive at statistical estimates.

A theoretical analysis of extreme hydrologic phenomena has led researchers to identify Gumbel distribution as a standard distribution for frequency analysis of recorded extreme meteorological data such as rainfall, temperature, wind speed, evaporation, etc; and hence used in the present study. Standard analytical procedures such as Method of Moments (MOM), Maximum Likelihood Method (MLM), Method of Least Squares (MLS), Order Statistics Approach (OSA) and Probability Weighted Moments (PWM) are commonly available for determination of parameters of Gumbel distribution [3]. Number of studies has been carried out by different researchers on analysing the characteristics of the parameter estimation methods of Gumbel distribution. Research reports indicated that MOM is a natural and relatively easy parameter estimation method. MLM is considered the most efficient method, since it provides the smallest sampling variance of the estimated parameters and hence of the estimated quantiles compared to other methods. But, the method has the disadvantage of frequently giving biased estimates and often failed to give the desired accuracy in estimating extremes from hydrological data. It may not produce good estimators in small samples, especially when the random variable is restricted to an interval that depends on the parameters [4-6]. PWM and MLS are much less complicated, and the computations are simpler. Parameter estimates from small samples using PWM and MLS are sometimes more accurate than the MLM estimates for Gumbel distribution. On the other hand, OSA estimators are unbiased and having minimum variance [7-10]. Since there is no general agreement in applying particular method for a region because of the characteristics of the estimators, an attempt is made to apply five parameter estimation methods of Gumbel distribution for modelling HMWS data recorded at Delhi and Visakhapatnam regions. GoF tests involving Anderson-Darling (AD) and Kolmogorov-Smirnov (KS) are employed for checking the adequacy of fitting of the method to the recorded data. Diagnostic analysis involving RMSE is used for selection of a suitable method of Gumbel distribution for modelling HMWS for the regions under study. The methodology adopted in estimation of design wind speed using all five methods of Gumbel distribution, GoF tests and diagnostic analysis are briefly described in the ensuing sections.

II. METHODOLOGY

A) Probability Distribution

The Probability Density Function [PDF; f(W)] and Cumulative Distribution Function [CDF; F(W)] of Gumbel distribution is given by:

\[
f(W) = \frac{e^{-(W-\alpha)/\beta} e^{-(W-\alpha)/\beta}}{\beta}, \quad W, \beta > 0 \quad (1)
\]

\[
F(W) = e^{-e^{-(W-\alpha)/\beta}} \quad (2)
\]
where $\alpha$ and $\beta$ are location and scale parameters of the distribution [11]. The parameters are computed by MOM, MLM, MLS, OSA and PWM; and further used to estimate extreme wind speed ($W_T$) for different return periods from

$$W_T = \alpha + Y_T \beta$$  \hspace{1cm} (3)

The Lower and Upper Confidence Limits (LCL and UCL) of the estimated wind speeds at 95% level are computed from

$$LCL = W_T - 1.96(SE)$$

and

$$UCL = W_T + 1.96(SE)$$

where $Y_T = -\ln(-\ln(1-(1/T)))$ and SE is the Standard Error of the estimated wind speed.

A1) Methods for Estimation of Parameters

Method of Moments

$$\alpha = \bar{W} - 0.5772157\beta$$

and

$$\beta = \left(\frac{\sqrt{\pi}}{\sqrt{6}}\right)\bar{W}$$

where $\bar{W}$ and $\bar{W}$ are the mean and standard deviation of the recorded HMWS.

Maximum Likelihood Method

$$\beta = \bar{W} - \frac{\sum_{i=1}^{N} W_i \exp(-W_i/\beta)}{\sum_{i=1}^{N} \exp(-W_i/\beta)}$$

and

$$\alpha = -\beta \log\left(\frac{\sum_{i=1}^{N} \exp(-W_i/\beta)}{N}\right)$$

Method of Least Squares

$$\beta = \frac{\sum_{i=1}^{N} W_i \ln(-\ln(P_i))}{\sum_{i=1}^{N} \ln(-\ln(P_i))}$$

and

$$\alpha = \bar{W} + \frac{\left(\sum_{i=1}^{N} \ln(-\ln(P_i))\right)\beta}{N}$$

where $P_i = (i - 0.44)/(N + 0.12)$ and $\ln(-\ln(P_i))$ defines the cumulative probability of non-exceedance for each $W_i$.

Order Statistics Approach

OSA is based on the assumption that the set of extreme values constitutes a statistically independent series of observations. The OSA estimators of Gumbel distribution are given by:

$$\alpha = r^* \alpha_M + r' \beta_M$$

and

$$\beta = r^* \beta_M + r' \beta_M$$

where $r^*$ and $r'$ are proportionality factors, which can be obtained from the selected values of $k$, $n$ and $n'$ using the relations as follows:

$$r^* = kn/N$$

and

$$r' = n'/N$$

Here $N$ is the sample size containing the basic data that are divided into $k$ sub groups of $n$ elements each leaving $n'$ remainders. $\alpha_M$ and $\beta_M$ are the distribution parameters of the groups, and $\alpha_M$ and $\beta_M$ are the parameters of the remainders, if any. These can be computed from the following equations:

$$\alpha_M = \left(\frac{1}{k}\right)\sum_{i=1}^{k} \alpha_{M_i}$$

and

$$\beta_M = \left(\frac{1}{k}\right)\sum_{i=1}^{k} \beta_{M_i}$$

where $S_i = k \sum_{i=1}^{k} W_i$, $i = 1,2,3,...,n$.

Probability Weighted Moments

$$\alpha = M_{100} - 0.5772157\beta$$

and

$$\beta = (M_{100} - 2M_{101})/\ln 2$$  \hspace{1cm} (11)

where $M_{100} = \bar{W}$ and $M_{101} = \sum_{i=1}^{N} W_i(N-1)/(N(N-1))$. Here ‘$i$’ is the rank assigned to each sample arranged in ascending order.

Computation of Standard Error

The values of SE of the estimated wind speed using MOM, MLM, MLS and PWM methods may be computed from the following equation:

$$SE = \frac{\beta}{\sqrt{N}} \left[1.1589 + 0.1919Y_T + 1.1Y_T^2\right]^{0.5}$$

The SE of the estimated wind speed by OSA can be obtained from

$$SE = \left[\frac{r^* W_n + r' W_n'}{2}\right]^{1/2}$$

where $r^* = (k/n)N$ and $r' = (n'/N)^2$. $W_n$ and $W_n'$ are defined by the general form as $W_n = (A_n \gamma_G^2 + B_n \gamma_G + C_n)^{1/2}$. The weights $A_n$, $B_n$, and $C_n$ used in determining OSA estimators, and values of $A_n$, $B_n$, and $C_n$ used in computing the SE for OSA, are given in AERB safety guide [12].

B) Goodness-of-Fit Tests

The AD and KS statistics are defined by:

$$AD = \sum_{i=1}^{N} \left[2(1-\frac{1}{N})\ln(1-Z(i)) + (2N+1-2i)\ln(1-Z(i))\right]$$

and

$$KS = \max_{i=1}^{N} \left|F_G(W_i) - F_D(W_i)\right|$$

where $Z(i) = F(W_i)$, for $i = 1,2,3,...,N$ and $W_1 < W_2 < ... < W_N$.

Also, $F_G(W_i) = (i-0.44)/(N+0.12)$ is the empirical CDF of $W_i$, and $F_D(W_i)$ is the computed CDF of $W_i$. If the computed values of GoF tests statistics given by the method are less than that of theoretical values at the desired significance level, then the method is suitable for determination of parameters of Gumbel distribution for modelling HMWS [13].

C) Diagnostic Test

Theoretical description of RMSE is given by:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} (W_i - W_i^*)^2\right]^{1/2}$$

where $W_i$ and $W_i^*$ are the recorded and estimated wind speed of $i$th observation given by Gumbel distribution. The method provides minimum RMSE is considered as the most suitable method for estimation of design wind speed [14].

III. RESULTS AND DISCUSSIONS

A) Estimation of Design Wind Speed

By adopting the methodology described above, a computer program was developed and used for modelling HMWS for Delhi and Visakhapatnam regions. The program computes the parameters of Gumbel distribution by five methods, GoF tests
and RMSE. HMWS data recorded at Delhi for the period 1991-2003 and Visakhapatnam for the period 1987-1997 was used to estimate the extreme wind speed for different return periods adopting five methods of Gumbel distribution [15].

Tables 1 and 2 give the extreme wind speed estimates for different return periods together with SE given by five methods of Gumbel distribution for Delhi and Visakhapatnam.

### Table 1: Extreme Wind Speed Estimates for Different Return Periods together with SE given by Five Methods of Gumbel Distribution for Delhi

<table>
<thead>
<tr>
<th>Return period (yr)</th>
<th>MOM</th>
<th>MLM</th>
<th>MLS</th>
<th>OSA</th>
<th>PWM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W_T$</td>
<td>$SE$</td>
<td>$W_T$</td>
<td>$SE$</td>
<td>$W_T$</td>
</tr>
<tr>
<td>2</td>
<td>55.7</td>
<td>3.1</td>
<td>55.6</td>
<td>3.1</td>
<td>55.8</td>
</tr>
<tr>
<td>5</td>
<td>66.6</td>
<td>5.3</td>
<td>66.6</td>
<td>4.8</td>
<td>67.8</td>
</tr>
<tr>
<td>10</td>
<td>73.8</td>
<td>7.1</td>
<td>73.8</td>
<td>6.2</td>
<td>75.7</td>
</tr>
<tr>
<td>20</td>
<td>80.7</td>
<td>9.0</td>
<td>80.8</td>
<td>7.6</td>
<td>83.3</td>
</tr>
<tr>
<td>50</td>
<td>89.7</td>
<td>11.5</td>
<td>89.8</td>
<td>9.4</td>
<td>93.1</td>
</tr>
<tr>
<td>100</td>
<td>96.4</td>
<td>13.4</td>
<td>96.5</td>
<td>10.8</td>
<td>100.5</td>
</tr>
<tr>
<td>200</td>
<td>103.1</td>
<td>15.3</td>
<td>103.3</td>
<td>12.2</td>
<td>107.9</td>
</tr>
<tr>
<td>500</td>
<td>112.0</td>
<td>17.9</td>
<td>112.1</td>
<td>14.1</td>
<td>117.6</td>
</tr>
<tr>
<td>1000</td>
<td>118.7</td>
<td>19.8</td>
<td>118.8</td>
<td>15.5</td>
<td>124.9</td>
</tr>
</tbody>
</table>

From Tables 1 and 2, it may be noted that the extreme wind speed estimates given by OSA are consistently higher when compared with the corresponding values of other four methods for both the regions.

### Table 2: Extreme Wind Speed Estimates for Different Return Periods together with SE given by Five Methods of Gumbel Distribution for Visakhapatnam

<table>
<thead>
<tr>
<th>Return period (yr)</th>
<th>MOM</th>
<th>MLM</th>
<th>MLS</th>
<th>OSA</th>
<th>PWM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W_T$</td>
<td>$SE$</td>
<td>$W_T$</td>
<td>$SE$</td>
<td>$W_T$</td>
</tr>
<tr>
<td>2</td>
<td>55.9</td>
<td>4.5</td>
<td>55.8</td>
<td>4.8</td>
<td>56.0</td>
</tr>
<tr>
<td>5</td>
<td>70.3</td>
<td>7.6</td>
<td>71.1</td>
<td>7.3</td>
<td>72.1</td>
</tr>
<tr>
<td>10</td>
<td>79.9</td>
<td>10.3</td>
<td>81.2</td>
<td>9.4</td>
<td>82.8</td>
</tr>
<tr>
<td>20</td>
<td>89.1</td>
<td>13.0</td>
<td>90.9</td>
<td>11.5</td>
<td>93.0</td>
</tr>
<tr>
<td>50</td>
<td>101.0</td>
<td>16.6</td>
<td>103.4</td>
<td>14.3</td>
<td>106.2</td>
</tr>
<tr>
<td>100</td>
<td>110.0</td>
<td>19.4</td>
<td>112.8</td>
<td>16.4</td>
<td>116.2</td>
</tr>
<tr>
<td>200</td>
<td>118.9</td>
<td>22.1</td>
<td>122.2</td>
<td>18.5</td>
<td>126.0</td>
</tr>
<tr>
<td>500</td>
<td>130.6</td>
<td>25.8</td>
<td>134.6</td>
<td>21.4</td>
<td>139.1</td>
</tr>
<tr>
<td>1000</td>
<td>139.4</td>
<td>28.6</td>
<td>143.9</td>
<td>23.6</td>
<td>148.9</td>
</tr>
</tbody>
</table>

### Table 3: Computed Values of GoF Tests using Five Methods of Gumbel Distribution for Delhi and Visakhapatnam

<table>
<thead>
<tr>
<th>Method</th>
<th>AD</th>
<th>KS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>0.295</td>
<td>0.173</td>
</tr>
<tr>
<td>MLM</td>
<td>0.293</td>
<td>0.173</td>
</tr>
<tr>
<td>MLS</td>
<td>0.271</td>
<td>0.160</td>
</tr>
<tr>
<td>OSA</td>
<td>0.285</td>
<td>0.169</td>
</tr>
<tr>
<td>PWM</td>
<td>0.274</td>
<td>0.168</td>
</tr>
</tbody>
</table>

From Table 3, it may be noted that the computed values of GoF tests statistics given by all five methods of Gumbel distribution are less than the theoretical values ($AD_{0.05}=0.757$; $KS_{0.05,11}=0.410$; $KS_{0.05,13}=0.377$) at five percent significance level, and at this level, all five methods are found to be suitable for determination of estimators of Gumbel distribution for the regions under study.

### C) Analysis Based on Diagnostic Test

The RMSE values were computed from Eq. (16) using the estimators of Gumbel distribution and given in Table 4.

### Table 4: RMSE values given by Five Methods of Gumbel Distribution for Delhi and Visakhapatnam

<table>
<thead>
<tr>
<th>Method</th>
<th>Delhi</th>
<th>Visakhapatnam</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>3.28</td>
<td>4.45</td>
</tr>
<tr>
<td>MLM</td>
<td>3.26</td>
<td>4.03</td>
</tr>
<tr>
<td>MLS</td>
<td>2.76</td>
<td>3.61</td>
</tr>
<tr>
<td>OSA</td>
<td>2.71</td>
<td>3.41</td>
</tr>
<tr>
<td>PWM</td>
<td>2.96</td>
<td>3.77</td>
</tr>
</tbody>
</table>

From Table 4, it may be noted that the difference between the RMSE values given by MLS, OSA and PWM are very meager. However, by considering the amount of variation in RMSE, OSA is identified as the best suitable method for determination of parameters of Gumbel distribution for estimation of design wind speed for both the regions. Figures 1 and 2 show the PDF plots of recorded and estimated wind speeds given by Gumbel (using OSA) for Delhi and Visakhapatnam.
Figure 1: Plot of recorded and Estimated Wind Speeds given by Gumbel distribution (using OSA) for Delhi

Figure 2: Plot of Recorded and Estimated Wind Speeds given by Gumbel Distribution (using OSA) for Visakhapatnam

Figures 3 and 4 show the plots of recorded and estimated wind speeds together with confidence limits at 95% level given by Gumbel distribution (using OSA) for Delhi and Visakhapatnam regions.

From Figures 3 and 4, it can be seen that the recorded annual HMWS data falls within the confidence limits of the estimated wind speed given by Gumbel distribution (using OSA) for both the regions. The Correlation Coefficient (CC) between the recorded and estimated wind speeds by all five methods of Gumbel distribution is computed as 0.977 for Delhi and 0.979 for Visakhapatnam.

IV. CONCLUSIONS

The paper presented a computer aided procedure for modelling HMWS data for estimation of extreme wind speeds for Delhi and Visakhapatnam regions adopting Gumbel distribution using MOM, MLM, MLS, OSA and PWM. GoF tests results confirmed the use of five methods for determination of parameters of Gumbel distribution for modelling HMWS for both the regions. The diagnostic analysis indicated that OSA is found to be an appropriate method for estimation of design wind speed based on the amount of variation on RMSE. The results showed that the RMSE on the estimated wind speed given by Gumbel distribution (using OSA) for Delhi and Visakhapatnam regions are 2.71 km/ hr and 3.41 km/ hr respectively. The results also showed that the CC values for Delhi and Visakhapatnam are 0.977 and 0.979 respectively. The study suggested that the Mean+SE (where Mean denotes the estimated extreme wind speed) values of about 147 km/ hr and 183 km/ hr related to 1000-yr return period may be adopted for design purposes of buildings and structures at Delhi and Visakhapatnam regions.

ACKNOWLEDGEMENTS

The authors are grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The authors are thankful to M/s NPCIL, Mumbai, and India Meteorological Department, Pune, for making available the wind speed data of Delhi and Visakhapatnam regions.

REFERENCES