

Assessing Adequacy of Probability Distributional Model for Estimation of Design Storm

Roman U.C.¹, Porey P.D.¹, Patel P.L.¹ and Vivekanandan N.²

¹Sardar Vallabhai National Institute of Technology, Surat 395007, INDIA

²Central Water and Power Research Station, Pune 411024, INDIA

Available online at: www.isca.in

Received 30th May 2012, revised 15th June 2012, accepted 18th June 2012

Abstract

Estimation of rainfall for a desired return period is one of the pre-requisites for any design purposes at a particular site, which can be achieved by probabilistic approach. In the present study, six probability distributions such as extreme value type-1 (EV1), normal, lognormal (LN2), gamma, pearson type-3, log pearson type-3 (LP3) are used to fit to annual 1-day maximum rainfall (ADMR) for Atner, Multai and Dharni sites in upper Tapi basin. Goodness-of-Fit tests such as Anderson-Darling, Chi-square and Kolmogorov-Smirnov are used to judge the applicability of the distributions for modelling recorded ADMR data. Diagnostic test, involving D-index, is used for selection of suitable distribution for estimation of rainfall for different return periods. The study shows the EV1 distribution is better suited, amongst six distributions studied, for estimation of design storm for Atner while LN2 for Multai and LP3 for Dharni.

Keywords: Anderson-Darling, Chi-square, D-index, Kolmogorov-Smirnov, probability, rainfall, return period.

Introduction

Estimation of design flood for a desired return period is a pre-requisite for planning and operation of various hydraulic structures such as dams, bridges, barrages and design of storm water drainage systems. These include different types of flood such as standard project flood, probable maximum flood and design basis flood. In case of large river basins, the hydrological and streamflow series of a significant duration are generally available. However, for ungauged basins, more data is not available other than rainfall. The rainfall data is also of shorter duration and may pertain to a neighbouring basin¹. Rainfall depth thus becomes an important input in derivation of flood discharge. Depending on the size and the proposed life of the structure, the estimated rainfall corresponding to a desired return period is used. Generally, 1-day maximum rainfall pertaining to 1000-year (yr) return period is used to arrive at a design parameter that a structure must withstand during its lifetime. For arriving at such design values, frequency analysis is identified as an effective and expedient tool for modelling annual 1-day maximum rainfall (ADMR) data^{2,3}. The procedure enables estimation of the probability of occurrence of a certain hydrological event of practical importance by fitting a probability distribution to one that is empirically obtained from recorded data.

A number of methods, based on probability distributions such as extreme value type-1 (EV1), normal (N2), lognormal (LN2), gamma (G2), pearson type-3 (P3) and log Pearson type-3 (LP3) are widely used for modelling ADMR data⁴⁻¹⁰. Hydrologists have recommended different distributions for fitting of ADMR data for estimation of rainfall for a desired return period. When

different distributions are used for modelling ADMR, a common problem that arises is how to determine which model fits best for a given set of data. This can be answered by formal statistical procedures involving goodness-of-fit tests; and the results are quantifiable and reliable than those from the empirical procedures.

Parameters of EV1 are determined by method of maximum likelihood and method of moments used for N2 and LN2. The parameters of G2, P3 and LP3 are determined from sample moments and then adjusted with moment bias¹¹⁻¹⁵. Qualitative assessment is made from the probability plot of the recorded and estimated rainfall. For quantitative assessment on rainfall data within the recorded range, Goodness-of-Fit (GoF) tests such as Anderson-Darling (A^2), Chi-square (χ^2) and Kolmogorov-Smirnov (KS) are applied. Diagnostic analysis, involving D-index, is used for selection of suitable distribution for modelling ADMR data. The objective of the paper is to assess the adequacy of a probability distributional model for estimation of rainfall for Atner, Multai and Dharni sites in upper Tapi basin. The methodology adopted in estimating the rainfall by six probability distributions, GoF and diagnostic tests are briefly described in the ensuing sections.

Methodology

Fitting probability distributions to the recorded ADMR data provides rainfall estimates for different return periods such as 2-yr, 5-yr, 10-yr, 20-yr, 50-yr, 100-yr, 200-yr, 500-yr and 1,000-yr. Table 1 gives the probability density function (PDF) with the corresponding rainfall estimator (R_T) of six distributions for modelling ADMR data.

Table-1
Probability density function and rainfall estimator of six probability distributions

| No. | Distribution | Probability density function | Rainfall estimator (R_T) |
|-----|--------------|--|--|
| 1 | EV1 | $f(r; \alpha, m) = \frac{e^{-(r-m)/\alpha} e^{-e^{-(r-m)/\alpha}}}{\alpha}$ | $R_T = m + Y_T \alpha$ |
| 2 | N2 | $f(r; \alpha, \lambda) = (1/\lambda \sqrt{2\pi}) \exp \left[-\frac{(r - \alpha)^2}{2\lambda^2} \right], \alpha, \lambda > 0$ | $R_T = \alpha + \lambda K_T$ |
| 3 | LN2 | $f(r; \alpha, \lambda) = (1/\lambda r \sqrt{2\pi}) \exp \left[-\frac{(\ln r - \alpha)^2}{2\lambda^2} \right]$ | $R_T = \exp(\alpha + \lambda K_T)$ |
| 4 | G2 | $f(r; \alpha, \lambda) = \frac{ \alpha e^{-\alpha r} (\alpha r)^{\lambda-1}}{\Gamma(\lambda)}, r > 0, \alpha, \lambda > 0, -\infty < r, \alpha < \infty, m > 0$ | $R_T = \left(\frac{1}{\alpha} \right) (K_T \sqrt{\lambda} + \lambda)$ |
| 5 | P3 | $f(r; \alpha, \lambda, m) = \frac{ \alpha }{\Gamma(\lambda)} e^{-\alpha(r-m)} [\alpha(r-m)]^{\lambda-1}, r > 0, -\infty < m < +\infty$ | $R_T = m + \left(\frac{\lambda + K_p \sqrt{\lambda}}{\alpha} \right)$ |
| 6 | LP3 | $f(r; \alpha, \lambda, m) = \frac{ \alpha }{\Gamma(\lambda)} \left(\frac{e^{\alpha m}}{r^{1+\alpha}} \right) [\alpha(\ln r - m)]^{\lambda-1}$ | $R_T = 10^{m + ((\lambda + K_p \sqrt{\lambda})/\alpha)}$ |

In the above PDF, α , λ and m are scale, shape and location parameters respectively; Y_T is a reduced variate for EV1 and $Y_T = -\ln[-\ln(1-(1/T))]$; K_T is the frequency factor corresponding to the coefficient of skewness (C_s) and $C_s = 2/\sqrt{\lambda}$ for G2, $C_s = 0.0$ for N2 and LN2; K_p is the frequency factor corresponding to C_s of the original and log-transformed series of the recorded ADMR for P3 and LP3 respectively. For N2 and LN2, the parameters are determined from mean and standard deviation of the original and log-transformed series of the recorded ADMR respectively¹⁶.

Goodness-of-Fit tests: GoF tests are either based on cumulative distribution function (CDF) or PDF. χ^2 -test is based on PDF, and A^2 and KS tests are based on CDF approach; and hence belong to the class of distance tests¹⁷. The theoretical descriptions of GoF tests are as follows:

$$\chi^2 \text{ statistics is defined by: } \chi^2 = \sum_{j=1}^{NC} \frac{(O_j(r) - E_j(r))^2}{E_j(r)} \quad (1)$$

where $O_j(r)$ is the recorded frequency value of ADMR for j^{th} class, $E_j(r)$ is the expected frequency value of ADMR for j^{th} class, NC is the number of frequency classes and p is the number of parameters of the distribution.

A^2 statistics is defined by:

$$A^2 = (-N) - (1/N) \sum_{i=1}^N \{ (2i-1) \log(Z_i) + (2N+1-2i) \log(1-Z_i) \} \quad (2)$$

where, $Z_i = F(r_i)$, for $i=1, 2, 3, \dots, N$, and $r_1 < r_2 < r_3 < \dots < r_N$; where $F(r_i)$ represents the CDF of r_i . The distribution of A^2 statistics does not depend on $F(r)$, but on the N sample values.

$$\text{KS statistics is defined by: } KS = \max_{i=1}^N (F_e(r_i) - F_D(r_i)) \quad (3)$$

where, $F_e(r_i) = (i - 0.35)/N$ is the empirical CDF of r_i , $F_D(r_i)$ is the computed CDF of r_i and N is the number of observations.

The rejection region of χ^2 , A^2 and KS statistics at the desired significance level ' η ' are $\chi_C^2 > \chi_{NC-p-1, 1-\eta}^2$, $A_C^2 > A_{N, 1-\eta}^2$ and $KS_C > K_{N, 1-\eta}$ respectively. If the computed values (A_C^2 , χ_C^2 and KS_C) of GoF test statistics of the distribution are less than that of critical values at the desired significance level ' η ' then the selected distribution is accepted to be adequate than any other distributions¹⁸.

Diagnostic test: D-index statistics is defined by

$$D\text{-index} = \frac{1}{R} \sum_{i=1}^6 |r_i - r_i^*| \quad (4)$$

where r_i and r_i^* are the i^{th} highest recorded and estimated ADMR values using six probability distributions, and \bar{R} is the average value of recorded ADMR. The distribution having the least value for D-index is considered as the best distribution for estimation of rainfall for a given return period¹⁹

Results and Discussion

Study area and data used: An attempt has been made to estimate the rainfall for different return periods for Atner, Multai and Dharni sites in upper Tapi basin. Figure 1 shows the map of the study area with the locations of rain gauge stations considered in the analysis²⁰. The ADMR recorded at the sites for the period 1943 to 2004 are used. The drainage areas of Atner, Multai and Dharni are 650 km², 932 km² and 2,860 km² respectively. Table 2 gives the statistical parameters of original and log-transformed series of ADMR for the sites under study.

Estimation of rainfall using probability distributions: By applying the methodology as described above, a computer program was used to fit the ADMR data recorded at Atner, Multai and Dharni using six probability distributions. The program computes the parameters of the distributions, rainfall estimates for different return periods, GoF and Diagnostic test values for the data under study. Tables 3-5 give the rainfall estimates from six distributions for different return periods ranging from 2-yr to 1,000-yr for the sites under study. The estimated rainfall for different return periods using six

distributions were further used to develop the rainfall frequency curves for Atner, Multai and Dharni; and delineated in figures 2-4.

From tables 3-5, it may be noted that estimated rainfall for return periods above 5-yr given by N2 distribution is comparatively less than the corresponding values given by other five distributions for Atner. Also, from tables 3-5, it may be noted that N2 distribution provides lower estimates for return periods above 20-yr for Multai and Dharni sites.

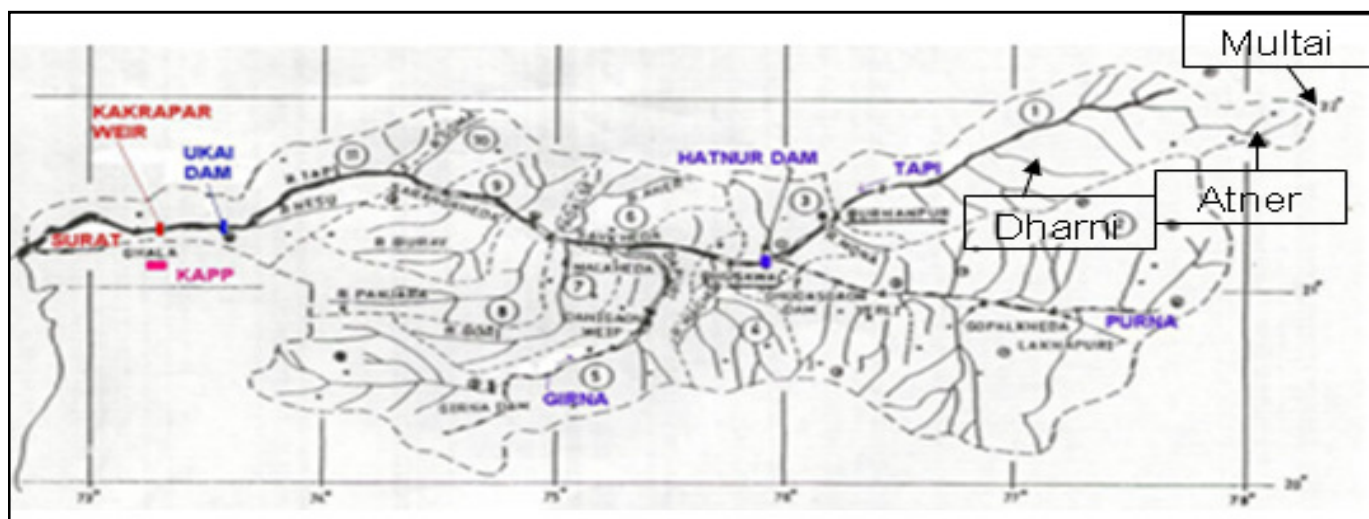


Figure-1
Location map of the study area

Table-2
Statistical parameters of annual 1-day maximum rainfall for Atner, Multai and Dharni

| Site | Statistical parameters of recorded annual 1-day maximum rainfall based on | | | | | | | |
|--------|---|-------|-------|--------|------------------------|-------|--------|-------|
| | Original series | | | | Log-transformed series | | | |
| | \bar{R} | S_R | C_s | C_K | \bar{R} | S_R | C_s | C_K |
| Atner | 88.1 | 33.2 | 0.670 | 0.541 | 4.405 | 0.399 | -0.564 | 0.941 |
| Multai | 102.9 | 55.1 | 3.010 | 14.589 | 4.530 | 0.448 | 0.201 | 1.604 |
| Dharni | 131.8 | 60.3 | 2.071 | 6.921 | 4.798 | 0.398 | 0.485 | 0.385 |

S_R : Standard deviation; C_K : Coefficient of kurtosis; C_s : Coefficient of skewness

Table-3
Rainfall estimates for different return periods by six distributions for Atner

| Return period (yr) | Estimated 1-day maximum rainfall (mm) using | | | | | |
|--------------------|---|-----|-----|-----|-----|-----|
| | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| 2 | 83 | 88 | 82 | 84 | 81 | 84 |
| 5 | 112 | 116 | 114 | 115 | 116 | 114 |
| 10 | 131 | 130 | 136 | 134 | 139 | 133 |
| 20 | 149 | 142 | 157 | 151 | 161 | 149 |
| 50 | 173 | 156 | 184 | 172 | 188 | 168 |
| 100 | 191 | 165 | 205 | 186 | 207 | 182 |
| 200 | 209 | 173 | 227 | 201 | 226 | 195 |
| 500 | 233 | 183 | 255 | 219 | 251 | 211 |
| 1000 | 250 | 190 | 278 | 232 | 269 | 223 |

Table-4
Rainfall estimates for different return periods by six distributions for Multai

| Return period (yr) | Estimated 1-day maximum rainfall (mm) using | | | | | |
|--------------------|---|-----|-----|-----|-----|-----|
| | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| 2 | 95 | 103 | 93 | 96 | 86 | 89 |
| 5 | 133 | 149 | 135 | 139 | 144 | 131 |
| 10 | 159 | 173 | 164 | 166 | 185 | 165 |
| 20 | 184 | 193 | 192 | 191 | 225 | 203 |
| 50 | 216 | 215 | 231 | 221 | 277 | 258 |
| 100 | 240 | 230 | 260 | 243 | 315 | 306 |
| 200 | 264 | 244 | 291 | 265 | 354 | 361 |
| 500 | 295 | 260 | 333 | 292 | 404 | 443 |
| 1000 | 319 | 272 | 365 | 312 | 443 | 514 |

Table-5
Rainfall estimates for different return periods by six distributions for Dharni

| Return period (yr) | Estimated 1-day maximum rainfall (mm) using | | | | | |
|--------------------|---|-----|-----|-----|-----|-----|
| | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| 2 | 122 | 132 | 121 | 124 | 114 | 117 |
| 5 | 166 | 182 | 169 | 174 | 175 | 165 |
| 10 | 196 | 208 | 201 | 204 | 218 | 203 |
| 20 | 224 | 230 | 232 | 232 | 261 | 243 |
| 50 | 261 | 255 | 273 | 266 | 316 | 301 |
| 100 | 289 | 271 | 304 | 290 | 357 | 350 |
| 200 | 316 | 286 | 335 | 313 | 398 | 405 |
| 500 | 352 | 304 | 377 | 342 | 452 | 486 |
| 1000 | 380 | 317 | 410 | 364 | 493 | 555 |

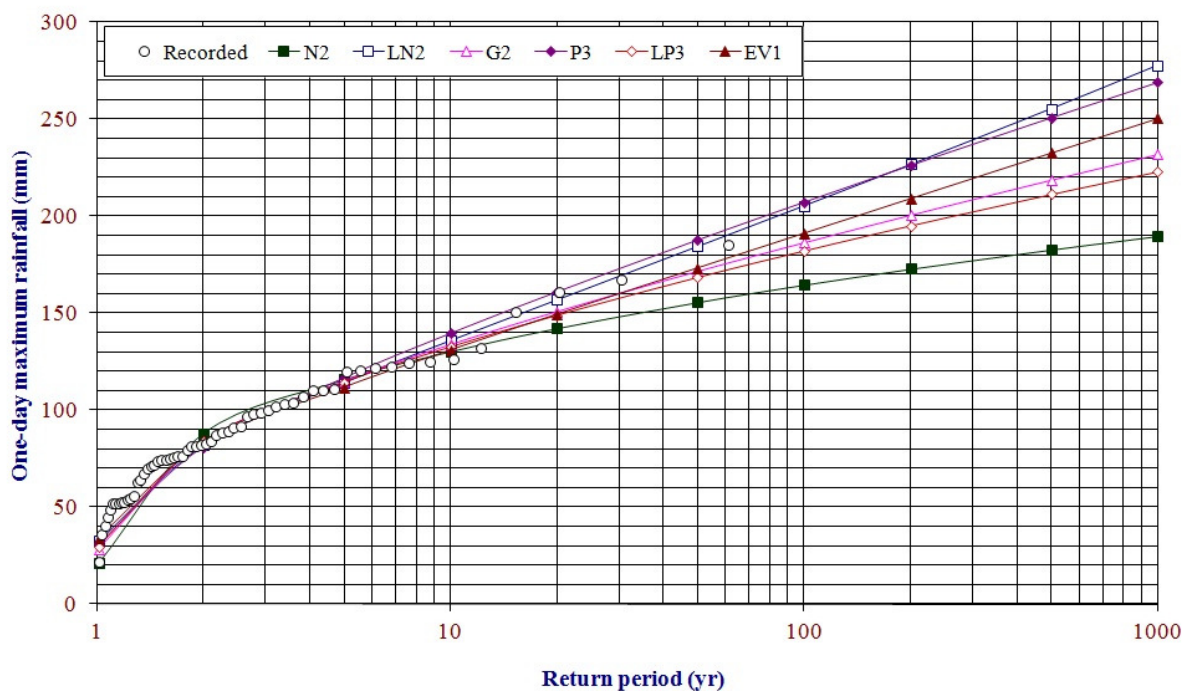


Figure-2
Probability plot of recorded and estimated 1-day maximum rainfall for different return periods using six distributions for Atner

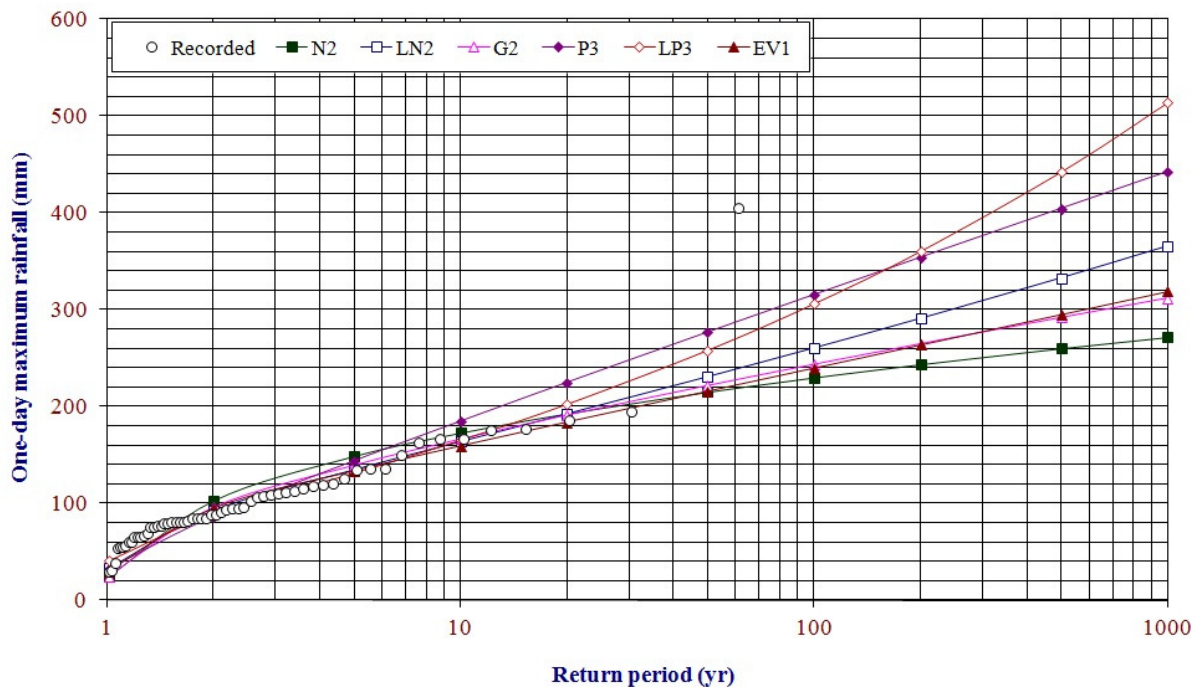


Figure-3
Probability plot of recorded and estimated 1-day maximum rainfall for different return periods using six distributions for Multai

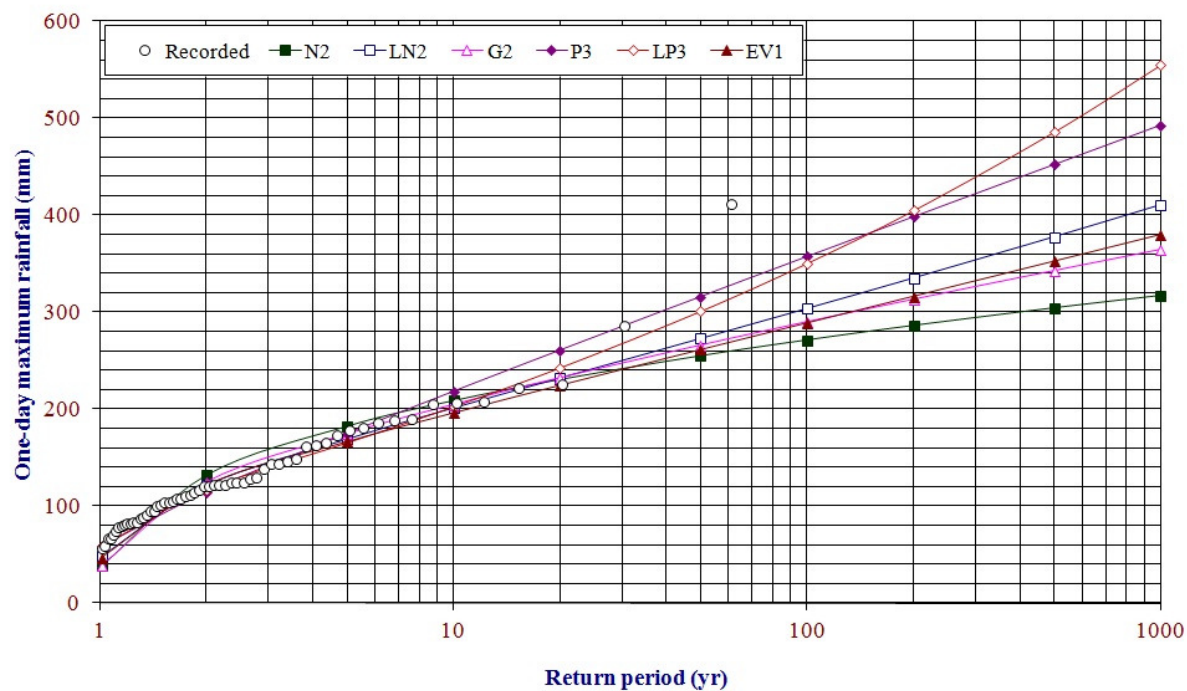


Figure-4
Probability plot of recorded and estimated 1-day maximum rainfall for different return periods using six distributions for Dharni

From figures 2-4, it can be seen that the fitted curves using six distributions show converging trend for all three sites. Also, from figures 2-4, it can be seen that the LN2 distribution gave higher estimates for return period above 100-yr for Atner while LP3 for Multai and Dharni when compared with other five distributions.

Analysis based on GoF tests: For assessment on fitting of statistical distributions to the observed ADMR data series, GoF test statistics for six distributions were computed using equation (1-3), and given in tables 6-8. In the present study, degrees of freedom for all six distributions were considered as eleven while computing χ^2 statistics for the data under study. By using the values given in tables 6-8, the adequacy on fitting of probability distributional model to the ADMR data recorded at the sites was analysed and are: i. The computed values of GoF tests involving χ^2 , A^2 and KS statistics, adopting six distributions, are less than the critical values at five percent level of significance, and hence at this level, all six distributions are accepted to fit the ADMR data recorded at Atner; ii. GoF test results don't support the use of P3 for modelling ADMR for Multai; iii. χ^2 and A^2 test

results don't support the use of N2 for estimation of rainfall for Multai and Dharni; iv. A^2 results indicated that G2 and LP3 distributions are not acceptable for modelling ADMR data recorded at Multai; and v. A^2 results showed that P3 is not suitable for fitting ADMR data recorded at Dharni.

From the analysis based on GoF test statistics, it is noticed that EV1, LN2, G2 and LP3 distributions are found to be uniformly acceptable for estimation of rainfall for Atner and Dharni while EV1 and LN2 for Multai.

Diagnostic analysis: Diagnostic analysis, using D-index, was adopted to identify the most suitable distribution for estimation of rainfall though GoF tests gave sufficient information on fitting of six distributions to the recorded ADMR data for the sites under study. By using equation (4), D-index values using six distributions for Atner, Multai and Dharni sites were computed, and given in table 9.

Table-6
Computed values of Chi-square (χ^2) statistics using six distributions

| Site | Critical value at 5% level | Computed values of χ^2 statistics using | | | | | |
|--------|----------------------------|--|------|-----|------|------|------|
| | | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| Atner | 19.675 | 13.2 | 14.0 | 8.4 | 9.2 | 9.2 | 13.6 |
| Multai | 19.675 | 12.0 | 30.4 | 8.0 | 12.4 | 34.8 | 13.2 |
| Dharni | 19.675 | 10.0 | 22.8 | 9.6 | 14.0 | 14.0 | 10.4 |

Table-7
Computed values of Ander-Darling (A^2) statistics using six distributions

| Site | Critical value at 5% level | Computed values of A^2 statistics using | | | | | |
|--------|----------------------------|---|-------|-------|-------|-------|-------|
| | | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| Atner | 0.757 | 0.353 | 0.538 | 0.350 | 0.229 | 0.659 | 0.216 |
| Multai | 0.757 | 0.502 | 4.469 | 0.513 | 0.888 | 2.558 | 0.799 |
| Dharni | 0.757 | 0.384 | 1.982 | 0.298 | 0.641 | 0.764 | 0.219 |

Table-8
Computed values of Kolmogorov-Smirnov (KS) statistics using six distributions

| Site | Critical value at 5% level | Computed values of KS statistics using | | | | | |
|--------|----------------------------|--|-------|-------|-------|-------|-------|
| | | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| Atner | 0.173 | 0.078 | 0.083 | 0.082 | 0.064 | 0.105 | 0.058 |
| Multai | 0.173 | 0.085 | 0.154 | 0.094 | 0.098 | 0.184 | 0.120 |
| Dharni | 0.173 | 0.094 | 0.167 | 0.093 | 0.118 | 0.093 | 0.060 |

Table-9
Computed values of D-index using six distributions

| Site | Indices of D-index using | | | | | |
|--------|--------------------------|-------|-------|-------|-------|-------|
| | EV1 | N2 | LN2 | G2 | P3 | LP3 |
| Atner | 0.282 | 0.943 | 0.372 | 0.533 | 0.504 | 0.609 |
| Multai | 2.009 | 2.138 | 1.922 | 1.945 | 2.748 | 1.888 |
| Dharni | 1.580 | 1.607 | 1.377 | 1.452 | 1.355 | 1.119 |

From table 9, it may be noted that the indices of D-index given by EV1, LN2 and LP3 are minimum when compared with the corresponding values of other distributions for Atner, Multai and Dharni respectively. From the results of GoF and diagnostic tests, it may be noticed that the EV1, LN2 and LP3 distributions are found to be suitable for estimation of rainfall for Atner, Multai and Dharni respectively. The results showed that the 1000-yr return period estimated rainfalls of 250 mm, 365 mm and 555 mm could be considered as a design parameter for planning and operation of hydraulic structures at Atner, Multai and Dharni sites respectively.

Conclusion

The paper presented a study on assessing adequacy in fitting of EV1, LN2, G2, P3 and LP3 distributions for estimation of design storm using GoF and diagnostic tests. The GoF test results uniformly supported the use of EV1, LN2, G2 and LP3 distributions for modelling ADMR data recorded at Atner and Dharni while EV1 and LN2 distributions for Multai. Diagnostic test results indicated that EV1, LN2 and LP3 are better suited for rainfall estimation for Atner, Multai and Dharni respectively. The study showed that the 1000-yr return period estimated rainfalls of 250 mm, 365 mm and 555 mm could be considered as the design parameter for planning and operation of hydraulic structures at Atner, Multai and Dharni respectively. The study also showed that the hydrographs derived from the estimated rainfall values could be served as input in the design of storm water drainage systems at the sites under study.

Acknowledgements

The authors are thankful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The authors are also thankful to the Director, Sardar Vallabhai National Institute of Technology, Surat, for making available the rainfall data of Tapi river basin.

References

1. National Institute of Hydrology (NIH), Technical Note on Hydrological Process in an Ungauged Catchment, 1-163 (2011)
2. Singh R.D., Mishra S.K. and Chowdhary H., Regional flow duration models for 1200 ungauged Himalayan watersheds for planning micro-hydro projects, *ASCE J. Hydrologic Engineering*, **6(4)**, 310-316 (2001)
3. Vaidya V.B., Karande B.I., Pandey Vyas, Lunagaria M.M. and Shekh A.M., Rainfall probability analysis for crop planning in Gujarat state, *J. Agrometeorology*, **10(1-2)**, 183-185 (2008)
4. Aksoy H., Use of Gamma Distribution in Hydrological Analysis, *Turkey J. Engineering Environmental Sciences*, **24(6)**, 419-428 (2000)
5. May W., Variability and extremes of daily rainfall during the Indian summer monsoon in the period 1901-1989, *Global and Planetary Change*, **44(1-2)**, 83-105 (2004)
6. Sharda V.N. and Das P.K., Modelling weekly rainfall data for crop planning in a sub-humid climate of India, *J. Agricultural Water Management*, **76(2)**, 120-138 (2005)
7. Chen J. and Adams B.J., Integration of Artificial Neural Networks with Conceptual Models in Rainfall-Runoff Modelling, *J. Hydrology*, **318(3)**, 232-249 (2006)
8. Carta J.A. and Ramirez P., Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions, *Renewable Energy*, **32(3)**, 518-531 (2007)
9. Rosatto H.G., Becerra A.T., Botta G. and Presutti M.E., Runoff estimation in small rural watersheds using DEMS in North West of Argentina, *J. Soil and Tillage Research*, **112(1)**, 8-17 (2011)
10. Sreekala P.P., Vijaya Bhaskara Rao S. and Rajeevan M., Northeast monsoon rainfall variability over south peninsular India and its teleconnections, *J. Theoretical and Applied Climatology*, **108(1-2)**, 73-83 (2012)
11. Hiremath Deepa B. and Shiyani R.L., Adapting Gujarat to Climatic Vulnerabilities: The Road Ahead, *Research J. Recent Sciences*, **1(5)**, 38-45 (2012)
12. Orlov Alexei M. and Ul'chenko Vasily A., Multi-annual changes of bottom temperatures in the Pacific off the North Kuril Islands and South Kamchatka (Northwestern Pacific, Russia) and demography of selected groundfish species, *Research J. Recent Sciences*, **1(2)**, 61-84 (2012)
13. Xu Y.P., Yu C., Zhang X., Zhang Q. and Xu X., Design rainfall depth estimation through two regional frequency analysis methods in Hanjiang River Basin, China, *J. Theoretical and Applied Climatology*, **107(3-4)**, 563-578 (2012)
14. Lee J.H. and Heo J.H., Evaluation of estimation methods for rainfall erosivity based on annual precipitation in Korea, *J. Hydrology*, **409(1-2)**, 30-48 (2011)
15. Saleh A.A., Developing an empirical formulae to estimate rainfall intensity in Riyadh region, *J. Engineering Sciences*, **23(2)**, 81-88 (2011)
16. Suhaila J. and Jemain A.A., Fitting Daily Rainfall Amount in Peninsular Malaysia using Several Types of Exponential Distributions, *J. Applied Sciences Research*, **3(10)**, 1027-1036 (2007)
17. D'Agostino B.R. and Stephens A.M., Goodness of Fit Statistic, Marcel Dekkar Inc., 270 Madison Avenue, New York 10016, USA (1986)
18. Zhang J., Powerful Goodness-of-Fit Tests Based on the Likelihood Ratio, *J. Royal Statistical Society*, **64(2)**, 281-294 (2002)
19. United States Water Resources Council (USWRC), Guidelines for Determining Flood Flow Frequency, Bulletin No. **17B**, 15-19 (1991)
20. Report on 'Comprehensive planning for the scheme flood protective measures in river Tapi in upper Tapi basin', Narmada Water Resources and Water Supply Department, Government of Gujarat, 1-45 (2000)