Monthly and Annual Maximum Rainfall Prediction using Best Fitted Probability Distributions in Junagadh Region (Gujarat- India)

Manoj Gundalia (mjgundalia@email.com)
CGPIT, Uka Tarsadia University

Research Article

Keywords: Rainfall, Frequency Analysis, Probability Distributions, Gumbel Distribution, Goodness of fit Chi-Square, Nash-Sutcliffe Efficiency.

Posted Date: July 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1770799/v1

Abstract

Rainfall is a meagre and crucial hydrological parameter in arid and semi-arid region. Junagadh (Gujarat-India) reels under monsoon rainfall uncertainties and thereby the agriculture and other water resources management activities suffer. Therefore, urgent attention is needed to address water resources conservation and crop damage issues due to deficits or excess rainfall. The amount of runoff produced and rainfall received determine the development of water resources in any region. Appropriate probability distributions need to be selected and fitted to the historical rainfall time series for better frequency analysis and forecasting of the rainfall. The daily rainfall data was collected for a period of 38 years i.e., from 1984 to 2021. In this study an attempt was made to find the most appropriate probability distributions for the better prediction of maximum rainfall by fitting the eight different hypothetical probability distributions to the monthly and annual maximum rainfall for one to five consecutive days. Chi-Square and Nash-Sutcliffe Efficiency were employed to determine goodness of fit. The results indicated that the Gumbel distribution appears to be the best fit to predict monthly and annual maximum rainfall of Junagadh region.

1 Introduction

India receives 75 to 80 percent precipitation out of total 4000 km^3 annual precipitation during rainy season under the influence of south-west monsoon (Kumar et al., 2005). In the arid and semi-arid regions, the yields of farmers are often badly affected due to extreme climate uncertainty. Due to its erratic nature and characteristic spatiotemporal variation, rainfall becomes the predominant key risk factor that has a direct or indirect effect on agriculture. Hence, the design and management of hydraulic structures, irrigation water supply, soil conservation planning, flood control systems, and optimal crop planning are based on rainfall depths that can be expected with a certain probability rather than the long-term average of rainfall data. There are various unpredictable sources of uncertainty concerning the physical processes that occur during hydrological events (Hosking and Wallis, 1997). The stochastic model (hydrological frequency analysis) can, however, be used as a tool to estimate how frequently a specified event will occur on average in a region based on the available data (Bhakar et al., 2006). In this method the magnitudes of events for design return periods are determined beyond the recorded range. Due to the high spatial and temporal variability of rainfall, irrigation water supply, flood control systems, and hydraulic structures are designed and managed based on rainfall depths that can be expected for a given probability.

Rainfall analysis using probability distribution models has been studied by a number of researchers. Kumar (2000) and Singh (2001) found that the Log Normal (LN) distribution is the best-fit probability distribution for annual maximum daily rainfall in India. Amin et al. (2016) found that the Log-Pearson Type-III (LP-III) distribution was the best-fit distribution to estimate annual maximum rainfall in the northern regions of Pakistan. Eslamian et al. (2007) suggested that the Generalized Extreme Value (GEV) and LP-III distributions provided the best fit to estimate maximum monthly rainfall as an extreme event in Iran. Lee (2005) and Ogunlela (2001) found that the LP-III distribution best fitted the rainfall distributions of Taiwan's Chia-Nan plain and Nigeria's Chia-Nan plain, respectively. For one to five consecutive days of maximum rainfall in Accra, Ghana, the LN-II distribution was shown to be the best-fit probability distribution (Kwaku et al., 2007). Olofintoye et al., (2009) identified that 50% of stations follow LP-III distributions and 40% follow Pearson Type-III distributions for peak daily rainfall in Nigeria. Sen et al. (1999) observed that the Gamma probability distribution provided the best fit to monthly maxima rainfall in arid regions of Libya. The US Water Resources Council (USWRC) recommended the LP-III distribution in 1967, and it was considered to be the best way of flood frequency analysis in the United States (Arora & Singh, 1989). Zalinaet al. (2002) concluded from their study on the annual maximum rainfall series in Malaysia that the GEV distribution is the best fit for analysing the annual maximum rainfall series. Hanson and Vogel (2008) reported that Pearson Type-III distribution fitted the best to the daily rainfall in the United States. According to Bhakar et al. (2008) the Gumbel distribution was the best fit for monthly maximum rainfall in India. Sharma and Singh (2010) evaluated that the LN and Gamma distributions were the best fit probability distributions for the annual and seasonal time scales, while the GEV distribution was found to be the best fit probability distribution for the weekly time scale in Pantnagar (India). The current study seeks the best-fit models for determining the frequency of extreme rainfall events as well as the maximum monthly and annual rainfall present over return periods of 3, 4, 5, 10, 15, 20, 25, 30 and 35 years in the Junagadh (Gujarat-India) region. These can be used to develop plans and policies for better management of water resources and agricultural issues. The findings would be useful for agriculturists, hydrologists, designers of hydraulic structures, irrigation engineers, environmental managers, and planners of water resources to develop better plans and policies.

2 Study Area And Data Sets

2.1 Study area

Junagadh (Gujarat-India) region lies between 20˚ 26ˊ to 21˚ 24ˊ North latitudes and 69˚ 24ˊ to71˚ 03’ East longitudes. Figure 1 shows the location of the study region. Junagadh District population in 2022 is 2,891,917. Junagadh district has a total geographical area of 8831 km^2, making it the state's seventh largest district by area. Population density of the district is 327 persons per km^2. The region is semi-arid region and it has mean annual rainfall of 1186 mm, mean maximum temperature 38.56˚C and mean minimum temperature 14.82˚C. The maximum mean of monthly rainfall was received as 93.79 mm in the month of July while minimum mean of monthly rainfall was received as 53.43 mm in the month of August.

Over-exploitation and mismanagement of ground water, limited aquifer water storage capacity, insufficient natural water conservation, and periodic unpredictable and inadequate rainfall patterns are all critical challenges for this region. Because water availability is a major issue in this area, rainfall has a significant impact on water resource management, cropping patterns and crop water requirements, irrigation scheduling, and environmental evaluation.

2.2 Data sets

The daily rainfall data for 38 years (1984–2021) were obtained from the website (http://www.jau.in/index.php/annual-weather-reports–weather-data) of Junagadh Agro-meteorological Cell which is located at between latitude of 21°31’ N and longitude of 70°33’ E with an altitude of 61 m. The weather parameters are recorded and published annually at the Agro-meteorological observatory; Department of Agronomy, Junagadh Agricultural University, Junagadh. The daily rainfall data for years 1984-2021were considered and analyzed for extreme rainfall events. The monthly and annual series of extreme rainfall datasets are derived from the daily rainfall data and used in frequency analysis. The relation between the magnitudes of is obtained by arranging the sample data in descending order of magnitude. Then each data is assigned with rank m = 1, for the first entry and so on till m = N, for the last event. The probability P of an event to or exceeded is given by the Weibull formula P = m / (N + 1).

3 Methodology

3.1 Selection of Probability Distributions

It is difficult to select probability distribution models in order to find the best-fit probability distribution for any location. In this study total eight probability distributions (Gumbel, Van Te chow (V. T. Chow), (LP-III), Log Normal (LN), Exponential (EXP), GEV, Generalized Pareto distribution (GP) and Gamma distributions)which are most frequently used or recommended in extreme rainfall analyses are presented. **Table 1** shows the description of selected probability distribution functions viz. probability density function (PDF), cumulative density function (CDF), quantile function and parameter estimates. The parameters of the chosen distributions are estimated using the method of moments (MOM) and the L-moments.

3.2 Goodness of fit criteria

3.2.1 Chi-square test

The validity of selected probability distribution models is checked using goodness-of-fit test statistics, which quantify the compatibility of a random sample with the theoretical probability distribution. The Ch-square test is used to test the degree of agreement between observed data and those expected upon a given null and alternative hypothesis. The null hypothesis is usually set up as the uninteresting hypothesis and hence, we wish to reject it. The Chi-square test is applied for testing the null hypothesis: the rainfall data follows the selected probability distribution adequately. If the calculated values of the Chi-square test statistic are less than the critical values at the chosen significance level, the null hypothesis is accepted, and the chosen distribution is assumed to be suitable for rainfall estimation. For this study the critical value of Chi-Square distribution is obtained 15.5073 at significant level of 0.05 and 8 degree of freedom.

The Chi-Square values, χ^2 can be calculated as Eq. (1):

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \qquad (1)$$

Where

O_i = Observed frequency

E_i = Expected frequency

i = number of observations (1, 2, …….n)

3.2.2 Nash-Sutcliffe efficiency (NSE)

NSE is a normalised statistic that is used to calculate the relative magnitude of residual variance (Moriasi et al., 2015 and Son et al., 2019). It is widely used and offers a better option for dimensionless goodness of fit (Green and Stephenson, 1986, Pretorious et al., 2013). NSE quantifies how well a model simulation can predict the outcome variable. It lies between 1.0 (perfect fit) and −∞. A NSE value of zero has the same predictive power as the mean, whereas an NSE value less than zero indicates that the observed time series' mean value would have been a better predictor than the model. NSE goodness of fit ratings for model calibration is presented in Table 2. NSE is calculated using Eq. (2):

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(E_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \qquad (2)$$

Where:

$\bar{O}$ = Mean of observed values

4 Results And Discussion

In this section, the eight selected probability distributions were used to compute the annual maximum rainfall for one day to five consecutive days. The results obtained from the fitting of the eight probability distributions are presented from Table 3 to 26. The goal is to identify which distribution provides the best fit for the month and year to daily precipitation data extracted from 1984 to 2021. The Chi-Square test results, NSE values, rating and ranking based on the NSE values and estimated parameters values of all the distributions are presented from Table 3 through Table 27.

4.1 Annual maximum rainfall

To evaluate the quality of the fit of the previously mentioned distributions, the Chi-Square test statistic and NSE coefficient must be calculated. From Tables 3 and 5, it can be seen that the Chi-Square test values obtained for all the distributions were more than the critical value of Ch-Square distribution (15.5073 at significant level of 0.05 and 8 degree of freedom) which indicated that they have not passed the test and none of them are considered to be a good one.However, according to NSE ranking V T Chow, LN and GEV distributions performed better to estimate one day annual maximum rainfall while Gumbel, LN and V T Chow distributions performed better to estimate three days annual maximum rainfall. It is observed from Table 4 and Table 6 that only Gumbel distribution has passed the Ch-Square test. However, Gumbel, LN and V T Chow distributions performed better with rank 1, 2 and 3 respectively based on NSE statistics to estimate two days annual maximum rainfall whereas LN, Gumbel and GEV distributions performed better with rank 1, 2 and 3 respectively based on NSE statistics to estimate four days annual maximum rainfall. For five days annual maximum rainfall (Table 7) only Gumbel distribution has passed the Ch-Square test. LN, Gumbel and GEV were identified as good performance distribution with 1, 2 and 3 rank based on NSE statistics. Overall, the best performance was found by Gumbel distribution to estimate two days annual maximum rainfall (passed the Chi-Square test, NSE value 0.9166 with rank 1).The worst result obtained by EXP distribution with NSE value − 0.2447 and poor rating for three days annual maximum rainfall. Figure 2 was created to aid in the spatial visualisation of the chosen distribution for estimating the annual maximum rainfall over two days. It is expressed see that the Gumbel distribution was the best fit to the data; it underestimated the two days consecutive rainfall up to 15 years return period. After that it overestimated the two days consecutive rainfall.

4.2 Monthly maximum rainfall (June)

The selected probability distributions were adjudge by comparing the average of Chi-Square and NSE values obtained for these distributions corresponding to return period 3, 4, 5, 10, 15, 20, 25, 30 and 35 years respectively for the month of June as shown in Tables 8 through 12. From the Chi-Square test and NSE values it was concluded that one day to five days consecutive maximum rainfall of June was best fitted by Gumbel and GEV distributions as compared to other distributions. Except three days maximum rainfall of June Gumbel distribution stood on 2nd rank based on NSE values; rest of the cases Gumbel distribution attained 1st rank. The best result was obtained for five days consecutive maximum rainfall of June by Gumbel distribution with NSE value 0.9933.The worst result obtained by GP distribution with NSE value − 8.0486 and poor rating for one day annual maximum rainfall. By looking at Fig. 3 it is possible to identify that Gumbel distribution had the highest intimacy of best fit among all for five days maximum rainfall of June.

4.3 Monthly maximum rainfall (July)

For July, Gumbel and GEV were passed Chi-Square test for all one to five days maximum rainfall while LN was passed Chi-Square test for only three days maximum rainfall with NSE value 0.8461 and attained 3rd rank (Table 13 to Table 17). Gumbel distribution was attained 1st rank four times and attained rank 4 in two-day maximum rainfall of July. The GEV distribution was reached to 2nd rank four times and 1st rank once in two days maximum rainfall of July. The Gumbel distribution consistently performed better for three to five days maximum rainfall of July but the GEV distribution performed best for two days maximum rainfall of July with NSE value 0.9899. The GP distribution had shown nastiest performance with NSE value − 13.2470 and poor rating for one day maximum rainfall of July. Figure 4 shows the best performance of the GEV distribution for two days maximum rainfall of July. It can be seen that the GEV distribution mostly under estimated while the Gumbel distribution slightly overestimated the rainfall after 15 years of return period.

4.4 Monthly maximum rainfall (August)

The goodness of fit test and the overall ranking based on NSE of the selected distributions for one to five days maximum rainfall of August are presented from Table 18 to Table 22. From the values that are shaded in the above tables, it is clear that the LN and Gumbel distributions consistently passed goodness of fit test for all one to five days maximum rainfall of August. Based on the comparison of the NSE values, LN distribution was attained 1st rank for all cases except for two days maximum rainfall of August. Gumbel distribution was attained 1st rank for two days, 2nd rank for three days, 3rd rank for four and five days and 5th rank for one day maximum rainfall of August. V T Chow distribution received 2nd rank for one- and four-days maximum rainfall of August. Exponential distribution was found unreliable as it was not passed goodness of fit test and shown poor rating with NSE value − 1.2754 for two days maximum rainfall of August. LN distribution was emerged as the best fit distribution with NSE value 0.9946 for four days maximum rainfall of August (Fig. 5).

4.5 Monthly maximum rainfall (September)

It is noticed from Table 23 through Table 27 that the selection of the Gumbel distribution showed an interesting behaviour. Only Gumbel distribution was passed Chi-Square test for all five cases followed by LN distribution which was passed all the cases except for five days maximum rainfall of September. Gumbel distribution was showed good fitting potential with 2nd rank for one day, 3rd rank for two days and 1st rank for three to five days maximum rainfall of September with NSE values 0.8740, 0.9126, 0.9284, 0.9531 and 0.9603 respectively. LN distribution was showed reasonable fitting potential with 1st rank for one and two days, 2nd rank for three days and 3rd rank for four days maximum rainfall of September with NSE values 0.9176, 0.9492, 0.9188 and 0.8998 respectively. The Gumbel distribution gave the best fit for five days maximum rainfall of September and the EXP distribution gave the poorest rating with NSE value − 0.4871for three days maximum rainfall of September. Comparative plot of selected distributions showing observed and estimated five days maximum rainfall of September is presented in Fig. 6.

5 Conclusions And Recommendations

Regional rainfall analysis was made by eight different probability distributions for one to five days maximum monthly and annual rainfall of Junagadh region. Daily rainfall data from 1984 to 2021 were used to determine the best fit probability distribution for the study region. The findings of this study allowed us to draw the following conclusions and recommendations.

1. Gumbel distribution showed the best fits for two days, four days and five days consecutive annual maximum rainfall while others were not passed goodness of fit test for any cases.
2. The best fit performance of Gumbel distribution was found for two days consecutive annual maximum rainfall with NSE value 0.9166 in the Junagadh region.

3. Gumbel and GEV distributions were fitted well to the rainfall data as compared to other distributionsfor one to five days days consecutive maximum rainfall of June and July. However, Gumbel distribution gave the best result with NSE value 0.9933for five days consecutive maximum rainfall of June and the GEV distribution performed the best with NSE value 0.9899for two days maximum rainfall of July.
4. LN and Gumbel distributions presented the overall best fits to the datafor all one to five days maximum rainfall of August. LN distribution was appeared as the best fit distribution with NSE value 0.9946 for four days maximum rainfall of August
5. Only Gumbel distribution was passed Chi-Square test for all one to five days maximum rainfall of September.Gumbel distribution showed the best fit with NSE values 0.9603 for five days maximum rainfall of September.

It is recommended that Gumbel distribution should be considered in the final selection of optimum probability distribution for one to five days maximum monthly and annual rainfallin Junagadh (Gujarat-India) region.

Declarations

Acknowledgement The author expresses deep respect and gratitude to Junagadh Agro-meteorological Cell, Department of Agronomy, Junagadh Agricultural University, Junagadh, for making hydro-meteorological data available on its website. Author Declarations Funding: Not applicable Conflicts of interest/Competing interests: The authors have no relevant financial or non-financial interests to disclose. The author also declares that he does not have any competing interests. Ethics approval/declarations: This material is the authors' own original work, which has not been previously published elsewhere. The manuscript is not currently being considered for publication elsewhere. The manuscript reflects the authors' own research and analysis in a truthful and complete manner. Data availability: The hydro-meteorological datasets analysed during the current study are available on the website of Junagadh Agro-meteorological Cell, Department of Agronomy, Junagadh Agricultural University, Junagadh. http://www.jau.in/index.php/annual-weather-reports–weather-data. Consent to participate: Not applicable. Consent for publication: Not applicable.

References

1. Amin MT, Rizwan M, Alazba AA (2016) A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. Open Life Sci., 11, 432–440.
2. Arora K, Singh VP (1989) A comparative evaluation of the estimators of the log Pearson type 3 distributions. J. Hydrol, 105, 19–37.
3. Bhakar SR, Bansal AN, Chhajed N and Purohit RC (2006) Frequency analysis of consecutive day’s maximum rainfall at Banswara, Rajasthan, India, ARPN Journal of Engineering and Applied Sciences, **1**, 3, 64–67.

4. Bhakar SR, Iqbal M, Devanda M, Chhajed N, Bansal AK (2008) Probability analysis of rainfall at Kota. Ind. J. Agric. Res., 42, 201–206.
5. Eslamian SS, Feizi H (2007) Maximum Monthly Rainfall Analysis Using L-Moments for an Arid Region in Isfahan Province, Iran. J. Appl. Meteorol. Climatol, 46, 494–503.
6. Green I. and Stephenson D (1986) Criteria for Comparison of Single Event Models, Hydrological Sciences Journal, 31 (3): 395–411.
7. Hanson LS, Vogel R (2008) The probability distribution of daily rainfall in the United States. In World Environmental and Water Resources Congress, Ahupua'A (pp. 1–10).
8. Hosking JRM, Wallis JR (1997) Regional frequency analysis: an approach based on L-moments. Cambridge: Cambridge University Press, England.
9. Kumar A (2000) Prediction of annual maximum daily rainfall of Ranichauri (Tehri Garhwal) based on probability Analysis. Ind. J. Soil Conserv, 28, 178–180.
10. Kumar R, Singh RD, and Sharma KD (2005) Water resources of India, Curr Sci, 89 (5): 794–811.
11. Kwaku XS, Duke O (2007) Characterization and frequency analysis of one day annual maximum and two to five consecutive days maximum rainfall of Accra, Ghana. ARPN J. Eng. Appl. Sci, 2, 27–31.
12. Lee C (2005) Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan. Crop Environ. Bioinf, 2, 31–38.
13. Moriasi DN, Gitau MW, Pai N, and Daggupati P (2015) Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. Transactions of the American Society of Agricultural and Biological Engineers, Vol. 58(6): 1763–178.
14. Ogunlela AO (2001) Stochastic Analysis of Rainfall Evens in Ilorin, Nigeria. J. Agric. Res. Dev., 39–50.
15. Olofintoye OO, Sule B F, & Salami AW (2009) Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria. New York Science Journal, 2(3), 1–12.
16. Pretorius H, James W and Smit J (2013) A Strategy for Managing Deficiencies of SWMM Modeling for Large Undeveloped Semi-Arid Watersheds, Journal of Water Management Modeling, 21: R246-01. doi: 10.14796/JWMM.R246-01.
17. Sen Z, Eljadid AG (1999) Rainfall distribution functions for Libya and Rainfall Prediction. Hydrol. Sci. J., 44, 665–680.
18. Sharma MA, & Singh JB (2010) Use of probability distribution in rainfall analysis, New York Science Journal, 3(9), 40–49.
19. Singh RK (2001) Probability analysis for prediction of annual maximum rainfall of Eastern Himalaya (Sikkim mid hills), Ind. J. Soil Conserv, 29, 263–265.
20. Son K, Lin L, Band L, and Owens EM (2019) Modelling the interaction of climate, forest ecosystem, and hydrology to estimate catchment dissolved organic carbon export, *Hydrological processes*. 33:1448–1464, 2019.https://doi.org/10.1002/hyp.13412
21. Zalina MD, Desa MNM, Nguyen VTA, & Kassim AHM (2002) Selecting a probability distribution for extreme rainfall series in Malaysia, Water science and technology, 45(2), 63–68.

Tables

Table 1 to 27 are available in the Supplementary Files section.

Figures

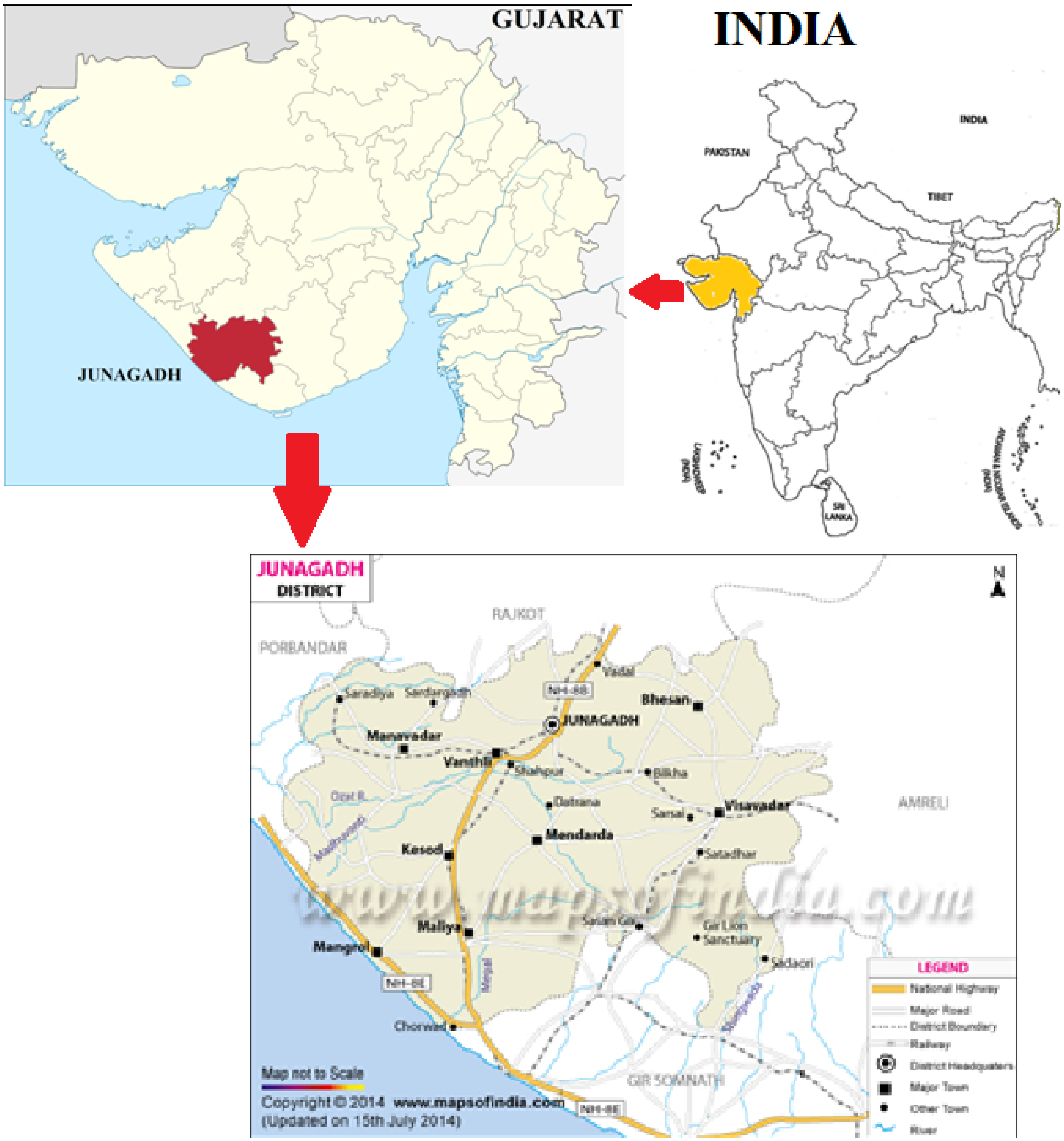

Figure 1

Junagadh (Gujarat-India) region

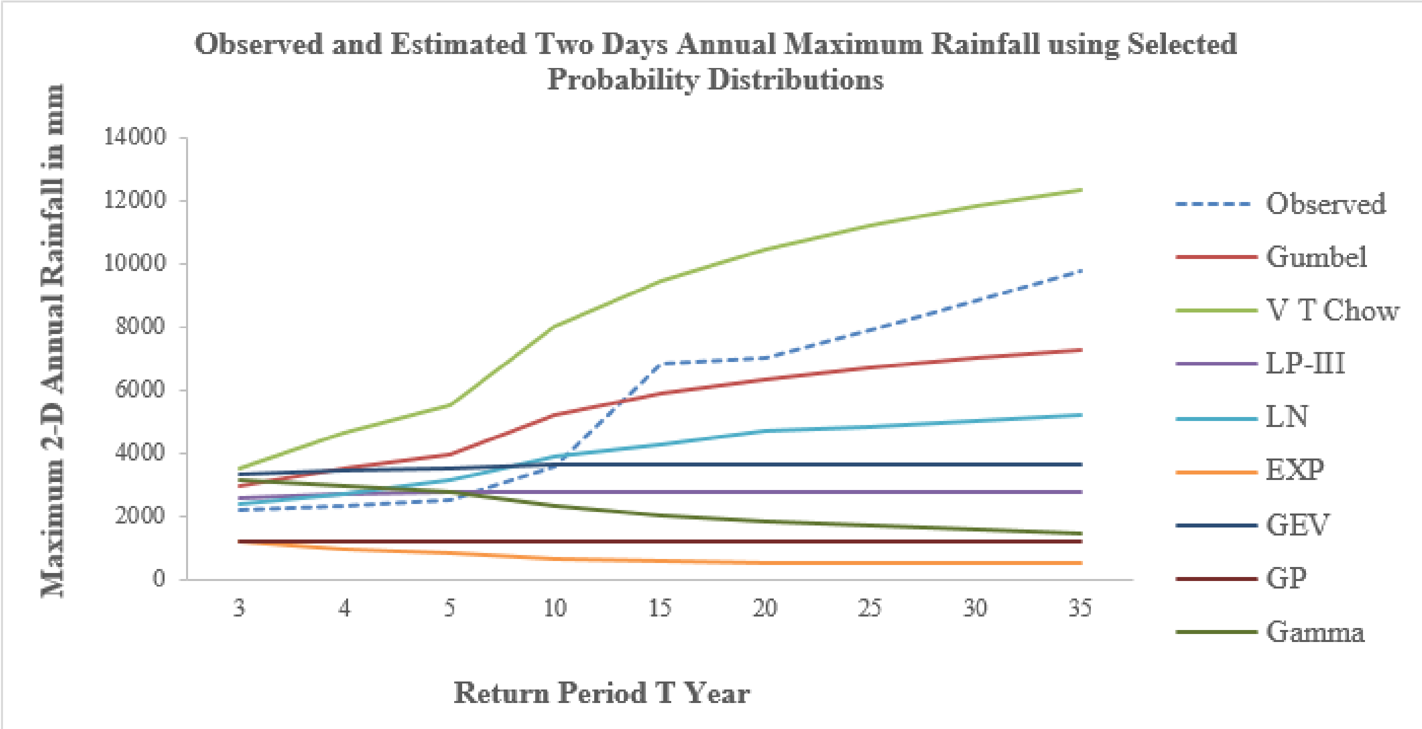

Figure 2

Observed and estimated two days annual maximum rainfall using selected probability distributions

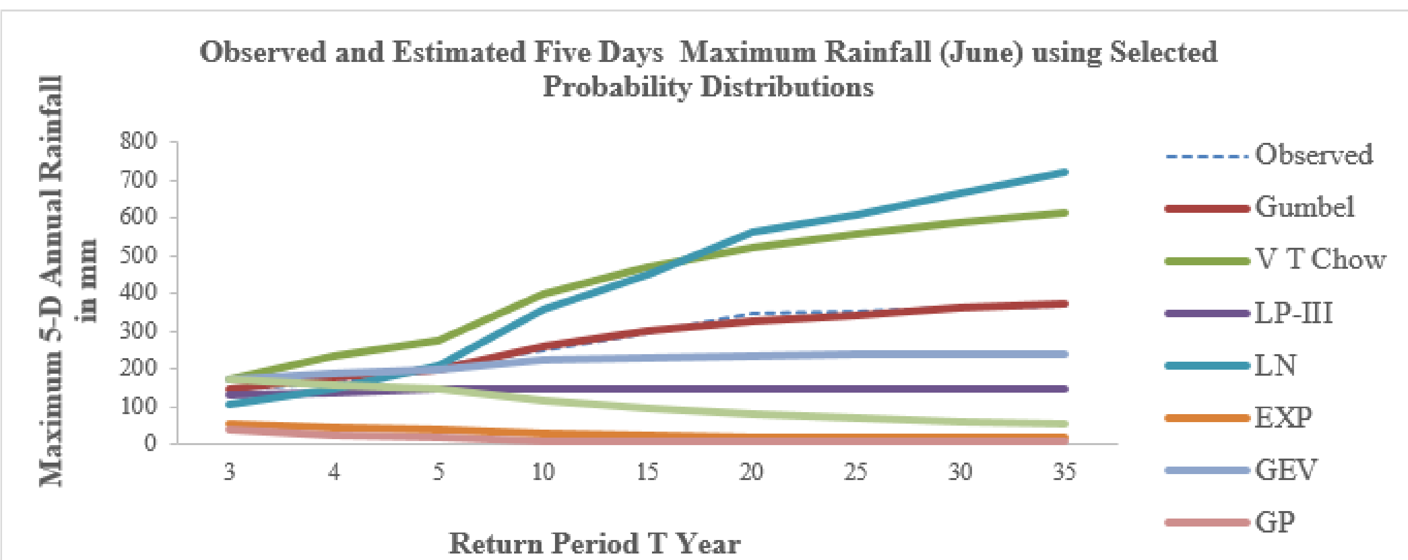

Figure 3

Observed and estimated five days maximum rainfall (June) using selected probability distributions

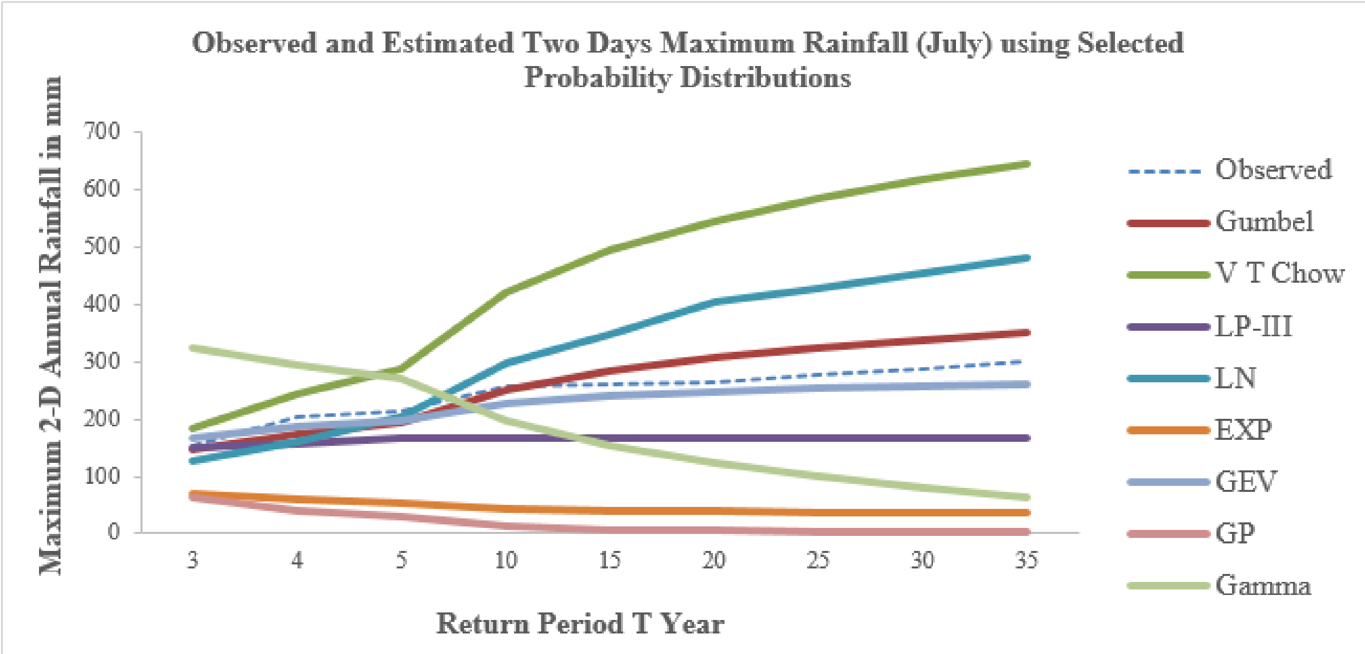

Figure 4

Observed and estimated two days maximum rainfall (July) using selected probability distributions

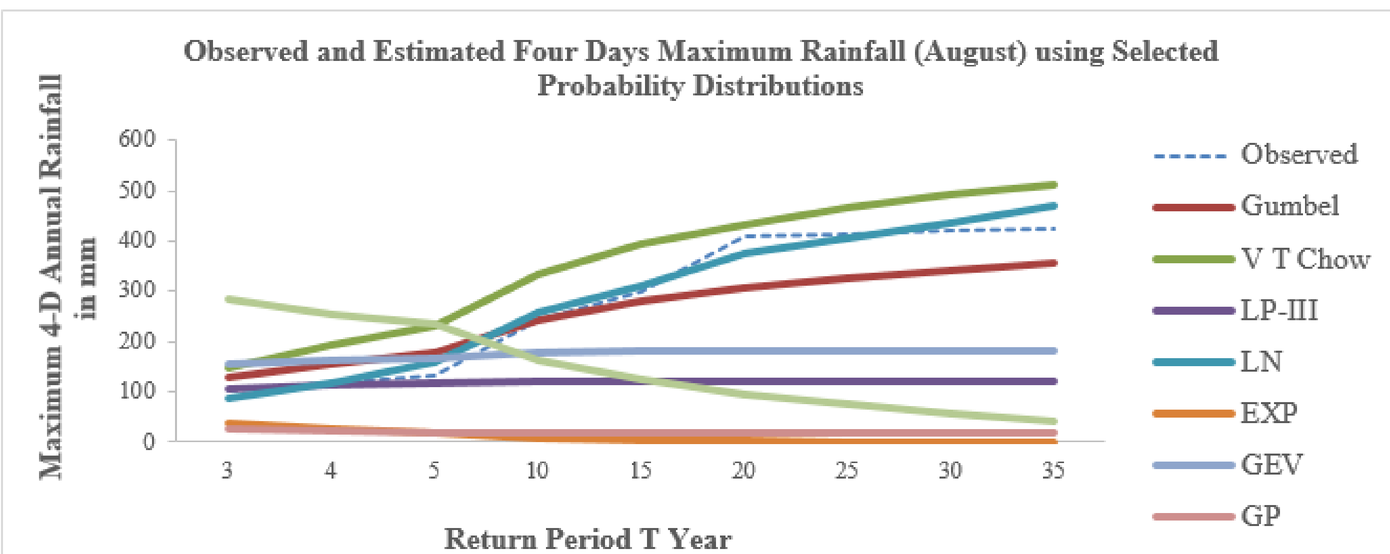

Figure 5

Observed and estimated four days maximum rainfall (August) using selected probability distributions

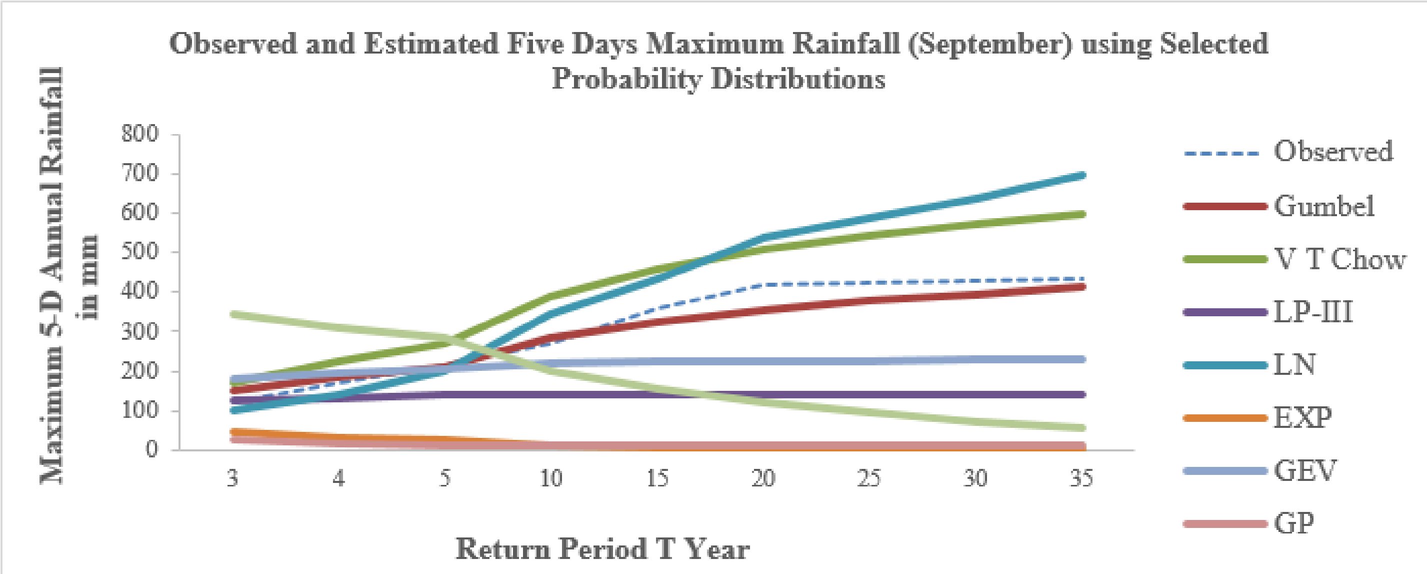

Figure 6

Observed and estimated five days maximum rainfall (September) using selected probability distributions

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx