Study of Various Techniques for Estimating the Generalised Extreme Value Distribution Parameters

Iqbal Hossain¹, Monzur Imteaz¹ and Anirban Khaustagir²
¹Swinburne University of Technology, Hawthorn, VIC 3122, Australia, Email: ihossain@swin.edu.au
²RMIT University, Melbourne, VIC 3000

Abstract. Generalised extreme value distribution (GEVD) remains the commonly employed technique for investigating the probability of occurrence of extreme events for given recurrence of intervals. However, the application of the GEV distribution requires the estimation of its three parameters. There are different methods presented in the literature to determine the parameters of the GEVD. Different methods have been adopted by different researchers in determining the three parameters. This paper investigates the comparison of the commonly used methods to estimate the GEVD parameters. The maximum likelihood estimation (MLE), generalised maximum likelihood estimation (GMLE) and L-moments methods were considered in this study. The analysis was performed using the monthly extreme rainfall of Tasmania, Australia. The GEVD was fitted to four different data sets using the three parameters estimation techniques. The outcomes of the analysis suggest that parameters estimation techniques have negligible impact on the magnitude of the parameters. However, length of the data series has minor impact on the parameters value of different parameters estimation techniques.

Keywords: GEVD parameters, MLE, GMLE, L-moments

1. Introduction
Rainfall is the complex atmospheric phenomena in the hydrologic cycle, which is difficult to predict with reasonable accuracy. The increase in anthropogenic activities in urban areas, which is responsible for man-made climate changes further exacerbate the complexity in rainfall prediction. The resilience of the climate change variability on rainfall could be adopted by incorporating the climate drivers (e.g. sea surface temperature anomaly) on rainfall forecasting models [1]. Regarding rainfall prediction, climate drivers are considered as the effective predictors in the literature [2-5]. Study on the effects of climate drivers on the long-term precipitation employing linear and non-linear modelling techniques are widely available in the literature. Most of the available research study prefer to the use of non-linear modelling approaches in excess of linear modelling techniques for the precise projection of long-term seasonal rainfall [4, 5]. However, some of the researchers found better predictability of combined linear models over non-linear models [6]. The statement was obtained using the small length of data. Nevertheless, investigation on the actual behaviour of extreme rainfall cannot be ascertained using the traditionally applied linear or non-linear models [2-4]. Therefore, prediction of rainfall using the tradition linear and non-linear modelling approaches could lead to the severe humanitarian catastrophe because of unanticipated floods and droughts. Consequently, accurate prediction of extreme
precipitation is critical for the design, planning, operation, and maintenance of essential infrastructures, and hence for monetary and safety purposes [7].

Although there is a possibility of decreased annual precipitation in some regions around the world due to anthropogenic climate change, the rate of recurrence of extreme climatic events (e.g. extreme rainfall) are likely to increase in near future [8]. Therefore, the incidence of rainfall induced floods are anticipated to increase around the world. Consequently, the extreme value theory was invented for the analysis and prediction of extreme climatic events. The historical formalisation and application of extreme value theory in hydrologic regimes is embedded in broad literature dating towards the back of 1940s [9]. The general extreme value distribution (GEVD) is traditionally used in modelling to characterise the rare events of natural phenomena. The GEVD is considered as the continuous probability distribution which has been developed within the extreme value hypothesis of statistics. The application of extreme distribution is of widespread interest to many academic disciplines, e.g. hydrology, finance etc. For example, Yilmaz et al. [10] used the generalised extreme value distribution in sub-daily extreme rainfall prediction. Embrechts et al. [11] applied GEVD technique for insurance and finance modelling. McNeil et al. [12] used the method in modelling extreme quantitative risk management.

The GEVD is an adaptable distribution combining three distributions inside a specific framework. The distribution is parameterised with the three parameters: location, scale, and shape. The single structure of the GEVD consists of the Gumbel, Fréchet and Weibull distributions that permit to reflect an array of potential shapes. The Gumbel, Fréchet and Weibull distributions are also known as type I, type II, and type III distribution, respectively. General introduction of the GEVD analysis is provided by [13]. The detailed evaluation of the method in the hydrological boundaries can be found in Katz et al. [14]. The magnitude of the GEVD parameters has the potential to influence the projected values of the extreme precipitation, which is used for the design and construction expensive hydraulic infrastructures. However, the uncertainty in the estimation of the GEVD parameters encounter significant challenges in fitting the distribution [7]. Complicated topography and changes in climatic conditions may further increase the difficulty in estimating the GEVD parameters for precipitation data.

The popular approaches for the estimation of the GEVD parameters are the maximum likelihood estimation (MLE) and the L-moments. Hosking et al. [15] favoured MLE in estimating parameters of the GEVD for small sample sizes due to small variances to the estimated quantiles. Due to the changes in the magnitude of the GEVD parameters, Ragulina and Reitan [16] emphasis on the regional estimation of the GEVD parameters. Because of the large spatial variability of precipitation, catchment scale analysis of extreme was also highlighted by Dyrrdal et al. [7]. However, there is paucity in the literature on detailed comparison of the parameters in identifying the superior parameters estimation technique. From the motivation of the better understanding on the magnitude of the GEVD parameters, this research investigated the influence of the GEVD parameters estimation techniques in extreme event modelling. Three different parameters estimation techniques (MLE, GMLE and L-moments) were investigated in this research. As there are regional differences on the magnitude of the GEVD parameters [16], Tasmanian rainfall in Australia was considered in this research as a case study.

2. Study Area and Data Collection
The area considered in this study is Tasmania, which is the island state of Australia. Twelve meteorological stations located in Tasmania were used and analysed to estimate the parameters of the GEVD distribution. Rainguage stations are selected based on long-term data availability and spatial distribution across Tasmania. Specific information of the meteorological stations is shown in Table 1. Geographical locations of the rainfall stations are shown in Figure 1. Historical rainfall data from the twelve selected rainfall stations were collected from the SILO database. The daily data were collected and used from 1965 to 2018.
3. Methodology

In this research, the collected rainfall from twelve meteorological stations were used and analysed using the extreme value statistics. Monthly maximum of the daily rainfall data was considered as the extreme
data in this paper. Like other research studies on extreme event modelling, it was assumed that the extreme rainfall data sets follow GEVD. The GEVD was fitted to the extreme data sets. The traditional GEV distribution can be explained by Equation 1 as follows:

\[
G(Y; \theta) = \exp \left\{ - \left[ 1 + \xi \left( \frac{Y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} 
\]

Where, \( Y \) represents the extreme rainfall data (in this case monthly maximum from daily rainfall data) and \( \theta \) represents the three parameters (\( \mu, \sigma, \xi \)) of the GEV distribution. From the three parameters, \( \mu \) denotes the location parameter, \( \sigma \) denotes the scale parameter and \( \xi \) denotes the shape parameter.

Application of the GEVD requires the selection of appropriate parameter estimation techniques. In this paper, three GEV parameters estimation techniques were used to compare the variation on the magnitude of the parameters. The three methods used are: maximum likelihood estimation (MLE), generalised maximum likelihood estimation (GMLE) and L-moments. As these methods are widely used in hydrologic research, their influences on the magnitude of the of parameters were assessed in this research. Influence of data series length on the magnitude of the parameters were also assessed in this research. The parameters of the GEV distribution were determined for four different extreme data sets: the whole study period (1965 – 2018), before millennium drought (1965 – 1996), during millennium drought (1997 – 2009) and after millennium drought (2010 – 2018). The parameters were estimated for each of the data series separately.

4. Results and Discussion

The estimated parameters for the GEV distribution are shown in Table 2 to Table 5 for the whole study period (1965 – 2018), before millennium drought (1965 – 1996), during millennium drought (1997 – 2009) and after millennium drought (2010 – 2018) respectively. Table 2 shows the estimated GEV parameters using the extreme rainfall data for the whole study period (1965 – 2018) for all the GEV parameters estimation techniques. It was observed from Table 2 that the shape parameters for all the stations is positive except one station (station #97020). This indicated the type II (Fréchet) GEV distribution is appropriate to model Tasmanian extreme rainfall for the whole study period. For station #97020, Weibull distribution is appropriate to determine the probability of occurrence of daily extreme rainfall. There is minor change in the magnitude of the estimated parameters for different GEV parameters estimation techniques.

| Station # | MLE | GMLE | L-moments |
|-----------|-----|------|-----------|
|           | \( \mu \) | \( \sigma \) | \( \xi \) | \( \mu \) | \( \sigma \) | \( \xi \) | \( \mu \) | \( \sigma \) | \( \xi \) |
| 91039     | 14.883 | 8.205 | 0.107 | 14.768 | 9.502 | 0.132 | 14.921 | 9.594 | 0.1 |
| 92019     | 12.546 | 8.189 | 0.358 | 12.542 | 8.207 | 0.36 | 12.535 | 8.145 | 0.354 |
| 92027     | 11.001 | 6.554 | 0.401 | 11.005 | 8.19 | 0.399 | 11.443 | 8.896 | 0.288 |
| 93014     | 9.384 | 8.574 | 0.251 | 9.347 | 6.541 | 0.264 | 9.659 | 7.033 | 0.163 |
| 94025     | 12.837 | 7.467 | 0.241 | 12.794 | 8.564 | 0.253 | 12.997 | 8.86 | 0.199 |
| 94061     | 9.855 | 6.16 | 0.321 | 9.841 | 6.648 | 0.326 | 10.099 | 7.059 | 0.242 |
| 95016     | 12.443 | 8.36 | 0.097 | 12.315 | 7.457 | 0.133 | 12.564 | 7.794 | 0.061 |
| 98000     | 15.056 | 10.813 | 0.11 | 14.951 | 8.327 | 0.137 | 15.102 | 8.455 | 0.098 |
| 94027     | 20.302 | 7.359 | 0.121 | 20.179 | 10.781 | 0.145 | 20.302 | 10.813 | 0.12 |
| 94041     | 15.958 | 12.521 | 0.078 | 15.848 | 7.328 | 0.109 | 15.926 | 7.308 | 0.085 |
| 97020     | 34.44 | 9.032 | -0.014 | 34.55 | 12.63 | -0.03 | 34.55 | 12.63 | -0.03 |
| 99000     | 13.214 | 7.644 | 0.204 | 13.169 | 9.03 | 0.216 | 13.109 | 8.811 | 0.227 |
The estimated GEV parameters before the millennium drought (1965 – 1996) are shown in Table 3. Like the whole study period, all the meteorological stations (except one station) follow type II (Fréchet) distribution for the millennium drought extreme data sets. The same station (station #97020) as observed for the whole study period extreme data follow type III (Weibull) distribution. The magnitude of the GEV parameters before the millennium drought data sets followed the same pattern of the whole study period data sets regardless of the method used to estimate the parameters.

Table 3. Estimated GEV parameters using rainfall data during drought (1965 – 1996)

| Station # | MLE | GMLE | L-moments |
|-----------|-----|------|-----------|
|           | μ   | σ    | ξ         | μ   | σ    | ξ         | μ   | σ    | ξ         |
| 91039     | 15.25 | 8.804 | 0.082 | 15.064 | 9.449 | 0.124 | 15.286 | 9.482 | 0.078 |
| 92019     | 12.968 | 9.022 | 0.369 | 12.965 | 8.806 | 0.37  | 12.859 | 8.539 | 0.387 |
| 92027     | 11.899 | 6.782 | 0.346 | 11.88  | 9.018 | 0.351 | 12.316 | 9.712 | 0.247 |
| 93014     | 9.417  | 9.027 | 0.264 | 9.36   | 6.761 | 0.283 | 9.774  | 7.399 | 0.154 |
| 94025     | 13.702 | 7.639 | 0.229 | 13.623 | 9.009 | 0.25  | 13.864 | 9.332 | 0.186 |
| 94061     | 10.641 | 6.284 | 0.259 | 10.584 | 7.163 | 0.278 | 10.91  | 7.659 | 0.178 |
| 95016     | 12.432 | 8.39  | 0.138 | 12.296 | 7.327 | 0.178 | 12.559 | 7.673 | 0.097 |
| 98000     | 15.078 | 10.545 | 0.123 | 14.931 | 8.35  | 0.161 | 15.126 | 8.496 | 0.11  |
| 94027     | 20.563 | 7.194 | 0.123 | 20.397 | 10.512 | 0.158 | 20.521 | 10.459 | 0.131 |
| 94041     | 17.813 | 12.465 | 0.035 | 17.603 | 7.146 | 0.097 | 17.717 | 7.033 | 0.061 |
| 97020     | 34.57 | 9.538 | -0.036 | 34.607 | 12.591 | -0.045 | 34.607 | 12.591 | -0.045 |
| 99000     | 13.934 | 7.875 | 0.155 | 13.82  | 9.524 | 0.182 | 13.868 | 9.412 | 0.169 |

Table 4 demonstrates the approximated GEV parameters for the extreme rainfall data during the millennium drought (1997 – 2009). During the millennium drought data sets, two stations followed type III (Weibull) distribution. All other stations followed type II (Fréchet) distribution. Nevertheless, not much variation on the magnitude of the parameters were observed depending the parameters estimation techniques used.

Table 4. Estimated GEV parameters using rainfall data during drought (1997 – 2009).

| Station # | MLE | GMLE | L-moments |
|-----------|-----|------|-----------|
|           | μ   | σ    | ξ         | μ   | σ    | ξ         | μ   | σ    | ξ         |
| 91039     | 13.46 | 7.53 | 0.15 | 13.21 | 8.67 | 0.217 | 13.534 | 8.991 | 0.122 |
| 92019     | 12.214 | 7.03 | 0.288 | 12.128 | 7.51 | 0.316 | 12.455 | 7.977 | 0.214 |
| 92027     | 9.828 | 6.161 | 0.433 | 9.864 | 7.034 | 0.419 | 10.318 | 7.86 | 0.284 |
| 93014     | 9.133 | 8.226 | 0.167 | 8.913 | 6.064 | 0.246 | 9.381 | 6.676 | 0.078 |
| 94025     | 11.744 | 7.38 | 0.206 | 11.536 | 8.154 | 0.264 | 11.999 | 8.768 | 0.133 |
| 94061     | 8.823 | 6.239 | 0.352 | 8.806 | 5.94 | 0.36 | 8.921 | 6.119 | 0.303 |
| 95016     | 12.238 | 7.618 | 0.005 | 12.344 | 8.569 | -0.031 | 12.344 | 8.569 | -0.031 |
| 98000     | 14.181 | 11.649 | 0.092 | 13.902 | 7.577 | 0.175 | 14.142 | 7.621 | 0.096 |
| 94027     | 20.062 | 6.754 | 0.099 | 19.635 | 11.573 | 0.182 | 20.016 | 11.669 | 0.101 |
| 94041     | 13.011 | 12.843 | 0.104 | 12.775 | 6.717 | 0.183 | 12.991 | 6.766 | 0.105 |
| 97020     | 34.334 | 7.941 | -0.012 | 34.599 | 13.081 | -0.051 | 34.599 | 13.081 | -0.051 |
| 99000     | 11.432 | 6.97 | 0.274 | 11.366 | 7.956 | 0.297 | 11.291 | 7.682 | 0.301 |
Table 5 shows the estimated GEV parameters for the extreme rainfall data after the millennium drought (2010 – 2018). Type II (Fréchet) distribution was observed for all the rainfall stations for the after-millennium drought data sets. Again, there was minor variation of the parameter estimation technique on the magnitude of the parameter. However, length of the extreme data sets has some impact in selecting the type of GEV distribution.

| Station # | MLE   | GMLE   | L-moments |
|-----------|-------|--------|------------|
|           | μ     | σ      | ξ          | μ     | σ      | ξ          | μ     | σ      | ξ          |
| 91039     | 15.736 | 7.081  | 0.137      | 15.357 | 10.674 | 0.222      | 15.695 | 10.745 | 0.137      |
| 92019     | 11.706 | 6.75   | 0.402      | 11.715 | 7.075  | 0.397      | 11.812 | 7.268  | 0.352      |
| 92027     | 9.822  | 6.349  | 0.571      | 9.988  | 6.759  | 0.499      | 10.229 | 7.378  | 0.43       |
| 93014     | 9.794  | 7.364  | 0.285      | 9.724  | 6.557  | 0.316      | 9.768  | 6.327  | 0.278      |
| 94025     | 11.732 | 6.733  | 0.301      | 11.669 | 7.376  | 0.325      | 11.593 | 7.206  | 0.329      |
| 94061     | 9.034  | 5.514  | 0.435      | 9.065  | 5.857  | 0.419      | 9.159  | 6.12   | 0.375      |
| 95016     | 12.772 | 9.307  | 0.113      | 12.469 | 6.624  | 0.217      | 12.824 | 6.935  | 0.083      |
| 98000     | 16.448 | 10.188 | 0.071      | 15.989 | 9.26   | 0.183      | 16.534 | 9.539  | 0.047      |
| 94027     | 19.545 | 5.734  | 0.181      | 19.19  | 10.085 | 0.261      | 19.775 | 10.769 | 0.125      |
| 94041     | 13.753 | 12.248 | 0.428      | 13.787 | 5.739  | 0.412      | 14.134 | 6.42   | 0.286      |
| 97020     | 34.243 | 8.532  | 0.047      | 33.563 | 12.186 | 0.171      | 34.203 | 12.214 | 0.05       |
| 99000     | 13.571 | 7.758  | 0.267      | 13.495 | 8.582  | 0.296      | 13.262 | 7.991  | 0.325      |

Visual comparison of the goodness fit test for a sample meteorological station (station #91039) is shown Figure 2. The plots for the goodness of fit test were performed for all the GEV parameters estimation techniques (i.e. MLE, GMLE and L-moments). The visual fits from one example station is shown in this manuscript for the whole study period (1965 to 2018). The probability plot of the extreme data sets forms an approximate straight line indicating that the GEV distribution is appropriate for our extreme data. Similarly, the quantile plot also shows that the extreme data sets came from the GEV distribution. However, some points are out of the line as shown in Figure 2. As outliers were not separated from our extreme data sets, some of the points in the quantile plots were out of the line. Likewise, the density plot of the GEV distribution shows that the empirical and modelled data match each other.

The estimated return levels for the selected station lied within the 45⁰ line up to 100 years return period. This indicated that the GEV models were good approximator to predict the daily extreme rainfall up to 100 years return period. Although the return levels did not lie within the 45⁰ line after the 100 years return periods, the estimated values are within the 95% confidence interval as shown in Figure 2. Similar outcomes were observed for all the parameter’s estimation techniques of the GEV distribution. Therefore, the GEV distribution is appropriate to determine the probability of occurrence of daily extreme rainfall in Tasmania.

5. Conclusions and Recommendations

This paper investigated the comparison of the GEVD parameters estimation techniques to be used for extreme event modelling. Long-term historical daily rainfall data from 1965 to 2018 were collected for twelve rainfall stations located in Tasmania, Australia. Parameters of the GEVD were determined using three different methods: MLE, GMLE and L-moments. For each of the methods, parameters were determined four different data series: whole study period, before millennium drought, during millennium drought and after millennium drought.
From the analysis, it was found that GEVD parameters estimation technique has negligible impact on the magnitude of the parameters. Moreover, the selection of the GEVD type does not depend on the parameter estimation technique. However, the length of the data series has minor impact on the magnitude and type of GEVD. Nevertheless, rigorous analysis on Tasmania as well as other states should be performed before making a generic conclusion. It should be noted that the analysis was performed based on the monthly maximum of the daily data; and hence it is likely to vary the outcomes from annual maximum data series.

![Graphs showing the goodness of fit test for meteorological station #91039](image)

Figure 2. Visual comparison of the goodness of fit test for meteorological station #91039

6. References

[1] Gado D A, Seidou O, Karambiri H, Sittichok K, Paturel J E and Saley H M, 2015, Development and assessment of non-linear and non-stationary seasonal rainfall forecast models for the Sirba watershed, West Africa. *J. Hydrol.: Reg. Stud.*, 4 134-152.

[2] Hossain I, Esha R and Imteaz M A, 2018, An Attempt to Use Non-Linear Regression Modelling Technique in Long-Term Seasonal Rainfall Forecasting for Australian Capital Territory. *Geosci.*, 8 282.
[3] Hossain I, Rasel H M, Imteaz M A and Mekanik F, 2018, Long-term seasonal rainfall forecasting: efficiency of linear modelling technique. *Environ. Earth Sci.*, 77 280.

[4] Hossain I, Rasel H M, Imteaz M A and Mekanik F, 2020, Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. *Meteorol. Atmospheric Phys.*, 132 331-341.

[5] Mekanik F, Imteaz M A, Gato-Trinidad S and Elmahdi A, 2013, Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *J. of Hydrol.*, 503 11-21.

[6] Djibo A, Karambiri H, Seidou O, Sittichok K, Philippon N, Paturel J and Saley H, 2015, Linear and Non-Linear Approaches for Statistical Seasonal Rainfall Forecast in the Sirba Watershed Region (SAHEL). *Climate*, 3 727-752.

[7] Dyrrdal A V, Skaugen T, Stordal F and Førland E J, 2016, Estimating extreme areal precipitation in Norway from a gridded dataset. *Hydrol Sci J*, 61 483-494.

[8] IPCC, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. 2012.

[9] Serinaldi F and Kilsby C G, 2014, Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resour. Res.*, 50 336-352.

[10] Yilmaz A G, Hossain I and Perera B J C, 2014, Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne. *Hydrol. Earth Syst. Sci.*, 18 4065-4076.

[11] Embrechts P, Klüppelberg C and Mikosch T, *Modelling Extremal Events: for Insurance and Finance*. 1997: Springer.

[12] McNeil A J, Frey R and Embrechts P, *Quantitative risk management: concepts, techniques and tools-revised edition*. 2015: Princeton university press.

[13] Coles S, *An Introduction to Statistical Modeling of Extreme Values*. 2001: Springer-Verlag: New York, NY.

[14] Katz R W, Parlange M B and Naveau P, 2002, Statistics of extremes in hydrology. *Adv Water Resour.*, 25 1287-1304.

[15] Hosking J R M, Wallis J R and Wood E F J T, 1985, Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27 251-261.

[16] Ragulina G and Reitan T, 2017, Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach. *Hydrol Sci J*, 62 863-879.