Original Research Article

Estimating Extreme Temperature at Agricultural Engineering College and Research Institute, Kumulur Station by Using Generalized Extreme Value Distribution

M. Vaidheki¹ and U. Arulanandu²

¹Agricultural Statistic, KRISAT, Madurai district, Tamil Nadu, India
²Statistics, ADAC&RI, Tiruchirapalli, Tamil Nadu, India
*Corresponding author

A B S T R A C T

The data are derived from daily maximum temperatures recorded at AEC&RI, Kumulur Meteorological Observatory. Extreme maximum temperature using 25 years of data is studied. The maximum temperatures are modelled using the generalized extreme value distribution. The results showed that twenty five year maximums are significant to be fitted to the GEV model and thus are used as our selection periods. The Mann-Kendall test detects no monotonic trends for the maximum temperatures. The Kwiatkowski-Phillips-Schmidt-Shin test indicates that the data are stationary. Parameter values are estimated using the method of maximum likelihood. Both the Kolmogorov-Smirnov and Anderson-Darling goodness of fit tests show that the GEV distribution gives reasonable fit to the maximum quarterly temperatures. Estimates of the T-year return levels for the return periods 5, 10, 25, 50, 100, 110 and 200 years reveal that the air temperature for the AEC&RI Kumulur will be increasing over the next 200 years.

Keywords
GEV, Extreme temperature, Kwiatkowski-Phillips-Schmidt-Shin and Mann-Kendall.

Introduction

Extreme value Theory differs from other typical statistical techniques in its objective to quantify the stochastic behaviour of a process at unusually large or small levels. It is based on the analysis of the maximum (or minimum) value in a selected time period.

In general, EVT usually requires estimation of the probability of events that are more extreme than any that have already been observed. Extreme value theory has been widely used and studied by many researchers. The earliest recorded application of extreme value was by astronomers in rejecting outlying observations.

There is growing concern around the world about increased emissions by the industrialised nations of greenhouse gases (GHG), which will cause increase in the global temperature and changes of other climatic variables such as rainfall and evaporation (Raskin and Kemp-Benedict, 2004; Parida; Moalafhi and Dube, 2005; Parida and Moalafhi, 2008). In its Fifth Assessment Report, the International Panel on Climate Change (IPPC) concludes, among others, that future climate extreme events would indicate significant changes in terms of the frequency, intensity and duration around the world. In vulnerable regions, extreme
weather and climate can lead to disasters with significant impacts on human and natural systems (Ebi and Bowen, 2016). Historically, according to Ebi and Bowen, extreme events were generally rare in any one location, with time between events when human and natural systems could recover from the impacts experienced. However, as climate change increases the frequency, intensity and duration of some extreme weather and climate events (IPCC, 2013), the time between extreme events will shorten across this century. This means that as the frequency, intensity and duration of extreme weather and climate events increase, there will be greater decreases in the return periods of such events. And, shorter return periods could affect the resilience of the affected communities to subsequent extreme events, especially in communities or countries at high risk of experiencing extreme weather and climate events like droughts. Precipitation and temperature extremes are considered to be the most important climate events and, have been extensively explored over the past several decades, according to Wen; Fang; Qi; Zhou and Gao (2015). Although deficient rainfall is considered the chief architect of droughts, heat waves and temperature extremes, though underestimated, often play crucial roles in drought development and intensification (Trenberth and Qian, 2004), (Bandyopadhyay; Bhuiyan and Saha, 2016). Regions with arid and semi-arid climate are more susceptible to drought, since they are more sensitive to rainfall deficiency and temperature extremes (Bhuiyan, 2008; Bandyopadhyay; Bhuiyan and Saha, 2016). The present study focuses on temperature extremes.

Materials and Methods

The present investigation was conducted at AEC&RI, Kumulur Tiruchirappalli district, 25 years daily maximum temperature data was analyzed and data was collected from the Meterological observatory, AEC&RI Kumulur.

Statistical analysis was carried out by 25 years extreme temperature was descriptive statistics, Mann Kendal trend Test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Stationary Test, fitting of probability distributions, estimating the parameter by using maximum likelihood method, Normality test (Q-Q Plot, P-P Plot, Probability difference plot), Goodness of fit test and estimating the return level. Different statistical software was used for different statistical analysis.

Study area

The study area lies between latitude 100 56’ 34.05’’N and longitude 780 49’ 34.89’’ E at an elevation of 72.2376 m above mean sea level. The total area of the campus is 280 acres. Annual average rainfall of Kumulur is 881.412 mm. The average monthly relative humidity is 60.5%. Different types of crops such as rice, maize, sugarcane, ragi and vegetables are growing in the campus.

Method of determination of the extreme value

Block maxima

Block Maxima Method Data is divided into blocks of a specific time period.

Each block is further specified period formed the highest value. Highest data is the sample of extreme values.

Peaks over threshold

This method uses standard or threshold value. Data that exceeds standard or threshold value is the sample of extreme value.
Methodology

The brief historical development of extreme value is given in the encyclopedia [KJR85] states that the basic theory of extreme value was first developed by Frechet in 1927 and by Fisher & Tippet in 1928 but was formalized by Gnedenko in 1943. Suppose there exists an independent distributed sequence of random variables X, the pdf function is given by

Mann Kendall trend test

The Mann-Kendall test is described in detail in Hirsch, et al., (1982, 1984), Gilbert (1987) and Helsel and Hirsch (2002). It is used to test for a monotonic trend of the variable of interest when the data collected over time are expected to have seasonality and serial correlation. The test assumes that when no trend is present, the observations in the time series are not serially correlated.

Adopting the notation of Hirsch, et al., (1982, 1984), let \( X_{ij} \) denote the observation obtained in season \( i \) in year \( j \), where a season may be a day, week, month, quarter or any other period of time. In the present paper, the season is a quarter of a year. Hence, let \( X = (x_1, x_2, x_3, x_4) \) represent the entire dataset collected over years consisting of data subsets \( X_i, x_2, x_3, x_4 \) where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{in}), i = 1, 2, 3, 4 \), denotes the set of data for the \( i \)th quarter for \( n \) years.

The null hypothesis \( H_0 \) for the two-sided (homogeneity) MK test is that there is no monotonic trend in the series and the alternative hypothesis \( H_A \) is that for one or more season there is an upward and downward trend over time.

The null hypothesis is \( z \) equal to zero. On the other hand, a positive (negative) value of \( z \) is an indicator of an increasing (decreasing) trend.

Mann Kendall test was carried out by using XLSTAT.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Stationary Test

The KPSS test tests the null hypothesis that a time series is stationary against the alternative that the series is non-stationary due to the presence of a unit root. The test is derived by starting with the model

\[
\begin{align*}
X_i &= \beta' D_i + \mu_i + \epsilon_i, \\
\mu_i &= \mu_{i-1} + \epsilon_i, \\
W &= WN(0, \sigma^2)
\end{align*}
\]

where \( \beta \) is a vector of regression coefficients, \( D_i \) contains deterministic components (constant or constant plus time trend), \( \mu_i \) is a pure random walk with innovation variance \( \sigma^2 \), and \( \epsilon_i \) is a random error term of \( X_i \) and may be heteroskedastic., \( \epsilon_i \) denotes the error term of \( \mu_i \) and is assumed to be a series of identically distributed independent random variables of expected value equal to zero and constant variance \( \sigma^2 \).

The KPSS test is a one-sided test with the null hypothesis of stationarity being equivalent to the assumption that the variance \( \sigma^2 = 0 \), which implies that \( \mu_i \) is a constant and, hence, that \( X_i \) is trend stationary. Thus, the KPSS test statistic is the Lagrange multiplier (LM) or score statistic for testing \( \sigma^2 = 0 \) against the alternative that \( \sigma^2 > 0 \) and is given by Equation (4) as
\[
K = \frac{1}{T^2} \sum_{t=1}^{T} \hat{S}_t^2
\]

(4)

Where \( \hat{S}_t = \sum_{j=1}^{t} \hat{u}_j \), \( \hat{u}_j \) is the residual of a regression of \( x_i \) on \( D_t \) and \( \hat{\lambda}^2 \) is a consistent estimate of the long-run variance of \( u_t \).

As the KPSS stationary test is a one-sided right-tailed test, one rejects the null of stationarity at the 100\( \alpha \)% level if the KPSS test statistic \( K \) is greater than the 100(1 - \( \alpha \))% quantile from the appropriate asymptotic distribution.

The MLEs of \( \mu, \sigma \) and \( \xi \) are determined by maximising Equation (5) with respect to the parameter vector \( (\mu, \sigma, \xi) \). See Hosking (1985) for details. No analytical solution to Equation (5) exists and numerical optimisation algorithms have been used in practice. In this study, the Newton-Raphson method was used to solve the likelihood equations above, following Hosking (1985) and MacLeod (1989).

Fitting of distribution and parameter estimation was carried out by EASYFIT software.

**Model assessment**

**Diagnostic plots**

Though it is impossible to check the validity of an extrapolation based on a GEV model, assessment can be made with reference to the observed data (Coles, 2001). In this paper, we use the probability plot, quantile plot, return level plot and density plot for model checking. However, graphical diagnostic techniques are subjective in nature and should, therefore, be supported by more rigorous statistical tests such as the well-known Kolmogorov-Smirnov and Anderson-Darling as they allow for quantitative assessment of model fit.

**Goodness-of-fit and model selection**

The two commonly used goodness-of-fit tests for extreme value models are the Kolmogorov-Smirnov test and the Anderson-Darling test, both of which are based on the empirical distribution function (edf).

**Kolmogorov-Smirnov test**

The Kolmogorov-Smirnov (K-S) test is defined as the largest vertical difference between the empirical cumulative distribution

\[
1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) > 0, \text{ for } i = 1, 2, ..., n.
\]
function (edf) $F_n(x)$ to and the theoretical cumulative distribution function (cdf) $F(x)$. It is used for testing the null hypothesis that a sample of size $n$ comes from a population with a specified distribution.

That is, if $X(1), X(2), \ldots, X(n)$ denote an ordered sample of independent observations from a population with distribution function $F(x)$, the K-S test statistic is given by Equation (6).

$$D = \max_{1 \leq i \leq n} \left( \frac{i}{n} - u_i, u_i - \frac{1}{n+1} \right)$$

(6)

Where $u_i = F(x_i)$, the cdf evaluated at the $i$th ordered observation of the ordered sample, $i = 1, 2, 3, \ldots, n$, and $n$ is the number of observations.

The null hypothesis of the test is rejected in favour of the alternative if the calculated value of $D$ is improbably large.

**Anderson-Darling test**

The Anderson-Darling (A-D) test due to Anderson and Darling (1952) is used for testing the null hypothesis that a sample of size $n$ comes from a population with a specified distribution. It is based on the discrepancy between the empirical cumulative distribution function $F_n(x)$ to and the theoretical cumulative distribution function $F(x)$.

This test gives more weight to the tails of the distribution than the Kolmogorov-Smirnov test. Let $X(1), X(2), \ldots, X(n)$ denote an ordered sample of independent observations from a population with distribution function $F(x)$. The test statistic of the Anderson-Darling test is

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left[ \ln(u_i) + \ln\left(1-u_{n+i-1}\right) \right]$$

(7)

Where, as before, $u_i = F(x_i), \ i = 1, 2, 3, \ldots, n$.

The distribution function of the A-D statistic is complicated, even asymptotically.

However, Marsaglia and Marsaglia (2004) provide a computational method to evaluate the accuracy of the cdf for any sample size.

And, the null hypothesis of the test is rejected in favour of the alternative if the calculated value of $A_n^2$ is improbably large.

**Return level**

The fitted distribution then can be used to estimate the $t$-year return level, which represents the high quantile for which the probability that the annual maximum exceeds this quantile is $1/t$.

The return period of a particular event is the inverse of the probability that the event will be exceeded in any given year, i.e. the $t$-year return level is associated with a return period of $t$ years.

The $T$-year return level is obtained by inverting the GEV distribution function given by

$$F(q_T) = 1 - \frac{1}{T}$$

(8)

$$q_T = \begin{cases} \mu - \frac{\sigma}{\xi} \left(1 - \log\left(1 - \frac{1}{T}\right)\right)^{-\frac{1}{\gamma}}, & \gamma \neq 0 \\ \mu - \sigma \log\left(1 - \frac{1}{T}\right), & \gamma = 0 \end{cases}$$

(9)
Fig. 1 PDF of various distributions

Fig. 2 PDF of various distributions for Extreme Maximum temperature

Probability density function
PDF of GEV distribution

Fig.3 Q-Q plot for fitted distributions to assessing normality
Fig. 4 P-P Plot for assessing the normality

Fig. 5 Probability difference chart for the fitted distribution model
Fig. 6 Estimated return level

KPSS (Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Stationary Test)

| KPSS test (Level / Lag Short / temperature): |
|---------------------------------------------|
| Eta (Obs) | 0.228 |
| Eta (Crit) | 0.440 |
| p-value | 0.242 |
| alpha | 0.05 |

Test interpretation:
H0: The series is stationary.
Ha: The series is not stationary.
As the computed p-value is greater than the significance level alpha = 0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 24.20%.

Table 1 Descriptive statistics of extreme temperature

| Statistic           | Value    | Percentile   | Value  |
|---------------------|----------|--------------|--------|
| Sample Size         | 25       | Min          | 36     |
| Range               | 6.5      | 5%           | 36.3   |
| Mean                | 40       | 10%          | 38.2   |
| Variance            | 2        | 25% (Q1)     | 39.5   |
| Std. Deviation      | 1.4142   | 50% (Median) | 40     |
| Coef. of Variation  | 0.03536  | 75% (Q3)     | 41     |
| Std. Error          | 0.26284  | 90%          | 42     |
| Skewness            | -0.88869 | 95%          | 42.35  |
| Excess Kurtosis     | 2.185    | Max          | 42.5   |
Table 2: Parameters of fitted extreme value distribution

| Distributions | Location | Shape      | Scale  |
|---------------|----------|------------|--------|
| GEV           | 39.624   | -0.42469   | 1.4032 |
| Gumbel Max    | 39.624   |            | 1.1027 |
| Frechet       |          | 39.132     | 28.192 |
| Pareto        |          | 36         | 9.5468 |
| Welbull       |          | 40.572     | 30.483 |

Table 3: Kolmogorov-Smirnov test

| Distributions | Statistics | Critical value | P value | Reject |
|---------------|------------|----------------|---------|--------|
| Gumbel Max    | 0.2533     | 0.2640         | 0.06692 | No     |
| GEV           | 0.20877    | 0.2640         | 0.196   | No     |
| Weibull       | 0.20284    | 0.26404        | 0.22269 | No     |
| Frechet       | 0.30402    | 0.31657        | 0.01511 | No     |

Table 4: Anderson-Darling test

| Distributions | Statistics | Critical Value | Significance level |
|---------------|------------|----------------|--------------------|
| Gumbel Max    | 2.4912     | 2.5018         | 0.05               |
| GEV           | 0.96712    | 1.374          | 0.05               |
| Weibull       | 0.0124     | 2.5018         | 0.05               |
| Frechet       | 2.441      | 2.5018         | 0.05               |

Table 5: Estimated return level

| Return period T(Years) | Estimated return level (°C) |
|------------------------|----------------------------|
|                        |                            |
| 5                      | 41.72                      |
| 10                     | 42.03                      |
| 50                     | 42.48                      |
| 100                    | 42.59                      |
| 200                    | 42.68                      |
Results and Discussion

The study uses 25 years of yearly maximum temperature data of Kumulur. A quick glance at Figure 2 will not give much insight to long term change. The summary statistics of extreme rainfall data is illustrated in Table 1, which revealed that data have slight flat tail at right side i.e. positively skewed and platikurtic with high variability. Maximum likelihood estimates of the Gumbel Max, Frechet and Weibull distribution are given in Table 2 and probability density functions (PDF) of Gumbel Max, Frechet, Weibull and GEV distributions are indicated in Figure 3, respectively.

In the fitted distributions Gumbel Max, Frechet and Weibull distributions are two parameters distribution and GEV is three parameters distribution, so that in Table 2 Gumbel Max, Frechet and Weibull have only two parameters but GEV has all the three parameters.

The Table 2 indicates that two parameter distribution having the comparatively larger value of their parameter than GEV distribution.

Mann Kendal trend test

\[ P=0.338 > q = (-41.000) \]

The results show the decreasing trend in the temperature series. There is has not been a significant monotonic trend in the overall temperature data.

The observed KPSS test statistics is \( k=0.228 \) with the \( p \) value greater than the 0.1, indicating that the test is not significant.

Conclude that the observed data of AEC&RI, Kumulur (temperature data), indicate that the data are in stationarity.

Model assessment

Normality test (Q-Q plot) Quantile plot

Normality of the fitted distributions is assessed by Q-Q plot indicated in the Figure 4. It is obvious form the figure that quintiles of GEV distribution [Figure 4 (D)] mostly occupied the reference line compare to other distributions. In Figure 5 Probability difference chart revealed that inner most (sky color) quintiles are much closer to horizontal reference line belongs to GEV distribution. So, on the basis overall statistical selection criteria GEV is the best fitted distributional model to assess the extreme maximum temperature in Kumulur station.

Normality of the fitted distributions is assessed by Q-Q plot indicated in the plots.

Goodness of fit

Goodness of fit for the fitted models is tested by K-S Test and A-D test illustrated in the Table 3. It is reported from the table that all the models are successfully passed the goodness of fit test (K-S Test and A-D test) at 5% level of significant (Tables 3 and 4). That means all the models can easily occupy the uncertainties pattern in the data.

All the models are successfully passed the goodness of fit test at 5% level of significant. So on the basis overall statistical selection criteria GEV is the best fitted distributional model to assess the extreme maximum temperature in Kumulur.

Return level

Once the best model for the data has been selected, the interest is in deriving the return levels of extreme maximum temperature. The return level for 5, 10, 50, 100 and 200 years are indicated in the Table 5.
This analysis shows extreme value Theory (EVT) to be a reliable tool for climate extreme. Generalized extreme value distribution is used to model maximum temperatures using data from AEC&RRI Kumulur for 1991 to 2015. The KPSS test of stationarity on the series reveals that the maximum temperatures are stationary. Mann Kendall trend test reveals no presence of monotonic trend. Parameters of the GEV distribution using the MLE method. Model diagnostics, which includes the probability plot, quantile plot and density plot shows a good fit of the GEV model to the AEC&RRI Kumulur. Kolmogorov-Smirnov and Anderson-Darling goodness of fit test is test the data is good for fit. T year return levels for the return periods, 5, 10, 25, 50, 100, 110 and 200 years respectively are estimated. AEC&RRI Kumulur temperature will be increase for the next 200 years. On the basis of available data, it is also revealed that the observed maximum temperature of 42.68°C may not be exceeded even in the next 200 years. For further study, researcher can make a long term prediction for other weather parameter which are directly or indirectly affect to the sector like agriculture, production industry and other sectors on which human life is dependent.

References

Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain ‘goodness-of-fit’ criteria based on stochastic processes. The Annals of Mathematical Statistics, vol. 23, No. 2, 193-212.

Arreyndip NA, Joseph E (2015) Extreme temperature forecast in Mbonge, Cameroon, through return level analysis of the generalized extreme value (GEV) distribution. International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering vol. 9, No. 6, 343 – 348.

Bandyopadhyay, N., Bhuiyan, C., and Saha, A. K. (2016), Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards, vol. 82, issue 1, 367-388.

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and Garcia-Herrera, R. (2011). The hot summer of 2012: Redrawing the temperature record map of Europe, Science 332 (6026), 220-224.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). Statistics of extremes. Theory and Applications, with contributions from de Waal, D.J. and Ferro, C. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Bhuiyan, C. (2008). Desert Vegetation during Droughts: Response and Sensitivity. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B8. Beijing.

Bouza-Deano, R., Ternero-Rodriguez, M., and Fernandez-Espinosa, A.J. (2008). Trend study and assessment of surface water quality in the Ebro River (Spain). Journal of Hydrology 361, 227–239.

Byakatonda, J., Parida, P.B., and Kenabatho, K.P. (2015). Climate Variability and Trends in Meteorological Time Series in Semi-Arid Botswana. 10th Alexander von Humboldt International Conference. Addis Ababa | Ethiopia | 18 – 20 November 2015, AvH10-54.

Ebi, K.L. and Bowen, K. (2016). Extreme events as sources of health vulnerability: Drought as an example. Weather and Climate Extremes 11 (2016), 95-102.

Hirsch, R.M. and J.R. Slack. (1984). A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 20 (6):727-732.

Hirsch, R.M., Slack, J.R., and Smith, R.A. (1982). Techniques of trend analysis for
monthly water quality data. Water Resources. Research 18, 107-121.
IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B, Tignor, M. and Miller, H.L. (editors). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.
IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
Jenkins, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. Quarterly Journal of the Royal Meteorological Society 81, 158-172.
Kwiatkowski, D., Phillips, P.C.B., and Shin, P.S., Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54, 159-178. North-Holland.
Parida, B.P. and Moalafhi D.B. (2008), Regional rainfall frequency analysis for Botswana using L-Moments and radial basis function network. Physics and Chemistry of the Earth, Parts A/B/C, 33(8), 614-620.
Wang, W., Vrijling, J.K., Van Gelder, P.H.A.J.M., and Ma, J (2006) Testing for nonlinearity of stream flow processes at different timescales. Journal of Hydrology, 322 (1–4), 247–268.
Wen, Fang, Qi, Zhou and Gao (2015). Changes of temperature and precipitation extremes in China: past and future. Theoretical and Applied Climatology, 10.1007/s00704-015-1584-x, 369-383.

How to cite this article:

Vaidheki, M. and Arulanandu, U. 2017. Estimating Extreme Temperature at Agricultural Engineering College and Research Institute, Kumulur Station by Using Generalized Extreme Value Distribution. Int.J.Curr.Microbiol.App.Sci. 6(11): 3874-3886.
doi: https://doi.org/10.20546/ijcmas.2017.611.454