International Journal of Applied Research 2019; 5(11): 288-291

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2019; 5(11): 288-291
www.allresearchjournal.com
Received: 20-09-2019
Accepted: 26-10-2019

Vishakha Tiwari
Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Pratyasha Tripathi
Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Vishal Vincent Henry
Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Correspondence Author:
Vishakha Tiwari
Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Grubbs-back test and multiple Grubbs beck test compared for the Godavari basin: A case study

Vishakha Tiwari, Pratyasha Tripathi and Vishal Vincent Henry

Abstract
In this paper, a few catchment areas of the Godavari river basin, Maharashtra, India, are used to compare the effectiveness of the Grubbs-Beck test and the multiple Grubbs-Beck test to detect discordant observations. It is found that the Multiple Grubbs-Beck test, is more effective while used for the Weibull distribution, than the original Grubbs-Beck test. It is observed that the upper river station, no lower discordant observation was found when using the method of Grubbs test and in Indravati river station also no lower discordant observation was detected while using the method of Multiple Grubbs Beck test (MGBT) five survey catchments.

Keywords: Grubb's test, discordant observation, MGB test, Weibull sample

Introduction
Flood frequency analysis is the method to estimate the design flood required for design Bridges, culverts, flood embankments and other water infrastructure, with various water resource management tasks such as flood management and flood insurance research.
Griffs and Stedinger (2008) [18] found that flood magnitude and frequency estimates using river monitoring stations with short annual peak flow data records have higher standard error or uncertainty compared to current levels of stream gauge with longer annual peak flow records. In flood frequency analysis, probability distributions are often selected based on statistical tests or graphical methods are convenience role in selection of proper distribution. An important step in flood frequency analysis is the detection of lower discordant in flood data, lower discordant is a very small observation of flood data and deviates significantly from the trends of the rest of the data. Identification and treatment of lower discordant is an important issue in flood frequency analysis, such observations can have a significant impact on the estimation of extreme flood quantiles.
Flood estimation extract as much information as possible from the available data, be robust to the distribution model, and finding a low discordant. In addition, unusually small values can lead to relatively poor estimates of large inundation quantiles.
Discordant observation is extra variable that disturb the whole observation in flood data series. This arises due to different causes, when the observation is different from the selected samples, measurement error or missing of observation, mostly observation is from the different population of most of the data.
Grubb's test is the method of detection of discordant observation. It was introduced by Frank E. Grubbs in the year 1950 after that it was also generalization of the Grubb beck test in the different test statistic in the year 1969, 1972. It also called as maximum normed residual test or extreme studentized deviate (ESD) test. It is a statistical test that exist in a single variate of data for detecting outliers which follows the probability distribution. Mean, standard deviation, sample are used for the analysis.
Many methods discussed in the Bulletin 17 B for the detection of discordant observation that are Z score test, Graphical method- Box plot, Histogram, Run sequence plot, Probability plot technique, Lag plot, Quality control test, Stedigner test, Grubb's-beck test and Multiple Grubb's-beck test. The 10% significance test with a single outlier threshold was applied in Bulletin 17B. This also recommended to determine to classify one or more of the smallest observations as low outliers in order to improve the frequency analysis robustness.

The $p(k;n)$ function of Cohn *et al.* (2013) [17] accurately explains whether the k_{th} smallest observation in a normal sample of n variates is uncommon. The masking effect is the effect of a real concern is that a flood record may contain more than one low discordant, causing the original GB test statistic is to fail to distinguish the smallest observation as a discordant, due to inward sweep tests are more sensitive to masking, an outward sweep is preferable to prevent the problem. A two-sided outward sweep is used by Rosner. The MBG test is also based on a one-sided critical value of 10% significance of a normally distributed sample, but this test is designed to examine ordered groups of data and are excluded from the dataset when calculating the critical values. The three parameters Weibull probability distribution used in this analysis for single or multiple discordant detection techniques. The three parameters of Weibull Distribution are the continuous distribution, this distribution mainly used in the extreme field events like flood data, survival analysis, etc. Weibull distribution used in the regional flood frequency analysis (2001) was performed by using the sample properties of quantile estimates based on the maximum likelihood (ML), Method of moments (MOM), and probability weighted moments (PWM) methods. Empirical performance assessment of outlier detection algorithms for Weibull or extreme–value distributions (2007) were used. In testing for lower discordant, the Grubb statistic G' performed well. Mann's W was significantly worse than the others for the labelled slippage.

The objectives of this paper are the detection of discordant observation in the five rivers station of the Godavari basin by using the Grubb's test and the multiple Grubb's-beck test using Weibull distribution.

Methodology

Weibull distribution

The three parameters Weibull distribution is mainly used for this analysis. The Weibull distribution was first distinguished by (Frechet Maurice 1927). This distribution is a continuous probability distribution named given by Swedish mathematician Waloddi Weibull. It can work efficiently, even with small sample sizes, and can be used precisely in frequency analysis and obtaining parameter estimates.

Let $X_1, \dots, X_n$ be independent random variables from a Weibull distribution with location, scale and shape parameters as $\mu, \sigma, and\ \gamma$ respectively.

For case I: When the observations are smaller than the location parameter μ, $i.e. \forall, x_i < \mu; i = 1, \dots, n$, then *pdf* is given by

Probability density function *(pdf)*

$$f(x_i;\mu,\sigma,\gamma) = \frac{\gamma}{\sigma}\left(\frac{\mu-x_i}{\sigma}\right)^{\gamma-1} e^{-\left(\frac{\mu-x_i}{\sigma}\right)^{\gamma}} \forall\ i = 1, \dots, n$$

Cumulative distribution function (*cdf)*

$$F(x_i) = e^{-\left(\frac{\mu-x_i}{\sigma}\right)^{\gamma}} \forall\ i = 1, \dots, n, \mu < \mathrm{x_i} < \infty, -\infty \leq \mu \leq +\infty, \sigma > 0, -\infty \leq \gamma \leq \infty$$

For case II: When the observations are larger than the location parameter $\mu, i.e. \forall, x_i > \mu; i = 1, \dots, n$ then the pdf is given by

Probability density function *(pdf)*

$$f(x_i;\mu,\sigma,\gamma) = \frac{\gamma}{\sigma}\left(\frac{x_i-\mu}{\sigma}\right)^{\gamma-1} e^{-\left(\frac{x_i-\mu}{\sigma}\right)^{\gamma}} \forall\ i = 1, \dots, n$$

Cumulative distribution function (*cdf)*

$$F(x_i) = 1 - e^{-\left(\frac{x_i-\mu}{\sigma}\right)^{\gamma}} \forall\ i = 1, \dots, n, \mu < \mathrm{x_i} < \infty, -\infty \leq \mu \leq +\infty, \sigma > 0, -\infty \leq \gamma \leq \infty$$

where, μ is the location parameter, σ is the scale parameter that decides the appearance or shape of the distribution and γ is the shape parameter.

Proposed test statistic

The original Grubbs-Beck tests uses the field logarithms of the peak flow data, to calculate a critical value at a significance level of 10% on one side of a normally distributed sample. Multiple peak flows recorded by the flow meter can be smaller than the critical Grubbs-Beck test, but usually only non-zero peak flows recorded in the test are identified as lower discordant. The original GB test recommended by Bulletin 17B. The original GB test identifies only one discordant from a particular dataset, but the data may have more discordant available. A method has been developed to statistically detect multiple discordant using the generalized Grubbs-Beck test. Grubbs test (1950) is a test for detecting a single discordant observation.

$$G = max\left|\frac{\mathrm{X_i}-\bar{\mathrm{X}}}{s}\right|$$

$\mathrm{X_i}$ is the i_{th} observation and $\overline{X}$ is the sample mean, s is the sample standard deviation.

Multiple Grubb's-beck test is a generalization of Grubb's-beck test, it is used for detecting the multiple discordant observation that is based on the extremes value distribution or approximate normal distribution. In this k different discordant was detected by using any data set. Implementing the recommended MGB test for Bulletin 17C consisting of two steps, firstly test each observation at the significance level α_{out} of the MGB test, starting from the median and outward toward the smallest observation. When the k-minimum observation is identified as a low outlier, the outward sweep stops and everything Observations smaller than k-smallest (that is, $j = 1, \dots, k$) are also identified as low outliers. Inward sweeps always start with the smallest observation and move towards the median. The significance level is α_{in}. If the observed value m ≥ 1 cannot be identified by the sweep, the sweep will stop. In that case, the total number of low discordant identified by the MGB test will be the maximum of k and m-1. The algorithm has two parameters that need to be specified, Outward sweep significance level α_{out} for each comparison. Inward sweep significance level α_{in} for each comparison.

Table 1 explains the comparison of no of discordant observation found by applying the Grubbs test, and Multiple Grubbs beck test for the five river stations of Godavari basin

In Multiple Grubbs beck test, Weibull distribution are used for the estimation of quantile or critical value of the data of river stations. It can be observed that multiple Grubbs-Beck is performing

Well as compared to that of Grubbs-beck test for detecting single discordant. Evidence can be seen for the Upper river station where multiple Grubbs-Beck test detected 20 discordant while Grubbs-beck test detected none of them.

Table 1: Comparison of single Grubbs test and multiple Grubbs Beck test in five rivers station of Godavari Basin

S. No	River Station	No of discordant observation for Grubbs test	No of discordant observation for MGB Test
1	Polavaram	1	13
2	Konta	1	7
3	Upper	0	20
4	Nandgaon	1	16
5	Indravati	1	0

Table 2: Estimated Quantile value by Weibull Distribution of five river station for Godavari Basin using Grubbs test and Multiple Grubbs Beck test with outlier or without outliers at 10% level of significance.

Estimated quantiles (m3 /s) using the Weibull distribution (Discordant removed by original GB test)					Percentage difference between the Weibull distribution with MGB test and the Weibull distribution with original GB test)		
S.No.	River Station	With discordant for Grubbs test (1)	Removing discordant by Grubbs test (2)	Removing discordant by MGBT (3)	% Difference between (1) and (2)	% Difference between (3) and (2)	% Difference between (1) and (3)
1	Polavaram	0.918062	0.928655	0.93722	-0.01 (-1%)	-0.01 (-1%)	0.02 (2%)
2	Konta	2.013576	0.894355	0.99823	1.12 (112%)	-0.10 (-10%)	0.01 (1%)
3	Upper	0.875678	0.881809	0.88282	0.01 (-1%)	0 (0%)	0.01 (1%)
4	Nandgaon	0.867439	0.870401	0.89566	0 (0%)	-0.03 (-3%)	0.03 (3%)
5	Indravati	0.897656	0.86543	0.87865	0.03 (3%)	-0.01 (-1%)	-0.02(-2%)

In Table 2, the comparison of estimated quantile values of single discordant observation and multiple discordant observation by using the single Grubbs test method and Multiple Grubbs beck test method for the five rivers station of Godavari basin are estimated. In Konta river station, high quantile value 2.013576 estimated with discordant by Grubbs test, but in Polavaram Godavari basin estimated the high quantile value 0.928655 with removing the discordant by Grubbs test. When using the Weibull distribution in MGB test in different river discharge, it also estimated the high quantile value 0.99823 in Konta river station.

It also shows the flood quantiles using Weibull distribution where the lower discordant are identified and shown by the original GB test and MGB test. It was found that there are% difference between the Weibull distribution with MGB test and the Weibull distribution with original GB test between the five different river station of Godavari Basin. It shows the variation between -1% to 3% percentage difference of flood quantiles between the methods of Grubbs test and Multiple Grubbs beck test. Table 2 shows the variation between the flood quantiles estimated by two methods: Weibull distribution with discordant by GB test, Weibull distribution without discordant by GB test and Weibull distribution without discordant by MGB test. It is found that for Station Konta, Weibull distribution overestimates flood quantile by 112%, underestimates the flood quantiles by 10%, 1%, 3%, and -1% respectively.

Result and Discussion

This paper used maximum discharge flow of rainfall in five river station of Godavari Basin. In Polavaram, Konta, Nandagaon and Indravati single discordant detected and in Upper Godavari no discordant detected by Grubbs test but when using the test of multiple grubbs beck test, different number of discordant detection for the different river station likewise, Polavaram, Konta, Nandgaon, Upper, and Indravati. Quantile values with discordant, without discordant for Grubbs test and quantile value without discordant for MGBT using Weibull distribution are also estimated for the Polavaram, Konta, Upper, Nandgaon and Indravati river station of Godavari Basin. 0.918062, 2.0135756, 0.875678, 0.867439, and 0.897656 quantile value for Polavaram, Konta, Nandgaon, Upper, and Indravati using Grubbs test with discordant. 0.928655, 0.894355, 0.87568, 0.870401, and 0.86543 estimated the quantile value for Polavaram, Konta, Nandgaon, Upper, and Indravati using Grubbs test without discordant. 0.93722,0.99823,0.88282, 0.89566, and 0.87865 quantile value for Polavaram, Konta, Nandgaon, Upper, and Indravati using Multiple Grubbs- beck test without discordant in Weibull distribution.

Conclusion

This paper estimates the discordant observation for the five rivers station of Godavari basins. To find out the single discordant and multiple discordant, two methods Grubbs test and Multiple Grubbs beck test was used for comparison and its quantile value was also estimated. For the upper river station, no lower discordant observation was found when using the method of Grubbs test and in also Indravati river station also no lower discordant observation was detected while using the method of Multiple Grubbs Beck test (MGBT). It also concluded that the when the variation between the flood discharge is low then the Grubbs test and Multiple Grubbs Beck test not found any discordant observation.

References

1. Bobee B, Cavidas Cavidas G, Ashkar Ashkar F, Bernier J, Rasmussen P. Towards a systematic approach to comparing distributions used in flood frequency analysis. Journal of Hydrology. 1993;142:121-136.
2. Rosner B. On the detection of many outliers. Technometrics. 1975;17(2):221-227.
3. Saf B. Assessment of the effects of discordant sites on regional flood frequency analysis. Journal of Hydrology. 2010;380(3-4):362-375.
4. Spencer C, McCuen R. Detection of PILFs in Pearson type III data. Journals of Hydrology Engineering. 1996;1:2-10.
5. Cunnane C. Factors affecting choice of distribution for flood series, Hydrol Sci J. 1985;30:25-36.
6. Cunnane C. Statistical distributions for flood frequency analysis, in Proc. Operational hydrological Report No. 5/33, World Meteorological Organization (WMO), Geneva, Switzerland, 1989.

7. Grubbs FE. Procedures for Detecting Outlying Observations in Samples, Technometrics. 1969;11(1):1-21.
8. Tietjen G, Moorie R. Some Grubbs-type statistics for the detection of several outliers, Technometrics. 1972;14(3):583-597.
9. Interagency Advisory Committee on Water Data (IAWCD), Guidelines for Determining Flood Flow Frequency: Bulletin 17-B. Hydrol. Sub comm., Washington, DC, 1982.
10. Interagency Advisory Committee on Water Data (IAWCD), Robust National Flood Frequency Guidelines: What is an Outlier? Bulletin 17- C, 2013.
11. Lamontagne JR, Stedinger Cohn TA, Barth NA. Robust national flood frequency guidelines: What is an outlier? In Proceeding World Environmental and Water Resources Congress, ASCE, 2013.
12. Beard LR. Statistical methods in hydrology, Civil works investigation project CW-151, U.S. Army Corps of Engineers, Sacramento, California, 1962.
13. Benson MA. Uniform flood-frequency estimating methods for federal agencies, Water Resource Research. 1968;4(5):891-908.
14. Vogel RM, McMahon TA, Chiew FHS. Flood flow frequency model selection in Australia. J Hydrol. 1993;146(421):449.
15. Merz R, Bloschil G, Humer G. National flood discharge mapping in Austria, Nat Hazards. 2008;46:53-72.
16. Nathan RJ, Weinmann PE. Application of at-site and regional flood frequency analyses, in Proc. International Hydrology Water Resources Symposium, Perth, 1991, 769-774.
17. Cohn TA, England JF, Berenbroc CE, Mason RR, Stedinger JR, Lamontagne. A generalized Grubbs-Beck test statistic for detecting multiple potentially influential outliers in flood series, Water Resources Research. 2013;49:5047-5058.
18. Griffs VW, Stedinger JR. The LP3 distribution and its application in flood frequency analysis, 2. Parameter estimation methods, Journal of Hydrology Engineering. 2007;12(5):492-500.
19. Thomas WO. A uniform technique for flood frequency analysis, Journals Water Resources Planning Management. 1985;111(3):321-337.