Journal of Mathematical Problems, Equations and Statistics 2022; 3(2): 01-05

Journal of Mathematical Problems, Equations and Statistics

E-ISSN: 2709-9407
P-ISSN: 2709-9393
JMPES 2022; 3(2): 01-05

www.mathematicaljournal.com
Received: 02-04-2022
Accepted: 03-05-2022

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

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

A technique to detect a discordant observation in Rainfall data of Shahid Veer Narayan Singh Dam (Kodar Reservior) of Mahasamund District of Chhattisgarh

Vishakha Tiwari, Pratyasha Tripathi and Vishal Vincent Henry

Abstract
In this paper, work has been done for the technique to discordant observation in rainfall data of Shahid Veer Narayan Singh Dam (Kodar reservoir) of Mahasamund district of Chhattisgarh. It has been demonstrated that discordant observation can significantly mislead the analysis, which is particularly dangerous when discussing sensitive issues such as extreme rainfall. A few such methods for identifying discordant observations have been found. As a result, a new test statistic for detecting a discordant observation from the distribution is proposed here. Its efficiency in terms of the probability of detecting a discordant observation is obtained. Its efficiency is also being especially in comparison to a sample of maximum rainfall data of Shahid Veer Narayan Singh Dam (Kodar reservior) of Mahasamund District of Chhattisgarh.

Keywords: discordant observations, extreme rainfall

Introduction

Shahid Veer Narayan Singh Dam (Kodar Reservior) is located at 210 11' 50" N to 820 10' 40" E in the Mahanadi basin of Chhattisgarh Plains. Figure 1 depicts the dam's location. The Shahid Veer Narayan Singh Dam project was initiated by the Chhattisgarh Water Resources Department in 1976-77 and done in 1998-99. The dam's goal is to irrigate a task area of at least 24876.16 Ha in 52 villages in Chhattisgarh's Mahasamund District. The river flows almost entirely through plains in Chhattisgarh. The Shahid Veer Narayan Singh Dam has a catchment area of 317.17 square Kilometres. The terrain in the area is flat.

The highest rainfall is an extreme value event; it is involved with probabilistic and statistical particular example to extremely high or extremely low values in variable sequences and stochastic processes. Several of the major application areas of the extreme value process also include finance, environment, internet traffic, athletics, asymptotic theory and statistical inference/tests. Rainfall is a crucial different climatic scale that must be precisely calculated. One of the critical weather issues affecting the Mahasamund region of Chhattisgarh is the change in rainfall rate.

Discordant detection approaches aid in the identification of features that may have been overlooked while using traditional statistics and mathematical models. When dealing with extreme events, the problem of discordance is a key problem. In statistics, a discordant is a single observation that is "far away" from entire data and can result in impossible findings, particularly when extrapolation of quantiles of the variables. Smith (2003) [24], for example, examined environmental datasets from this perspective, and a played crucial in track performances reached by a Chinese athlete, along with a world record, which raised instant doubts.

Robinson and Tawn (1995) [23] investigated the mentioned earlier collection to determine discordant performance. Non-stationarity is frequently visible in environmental processes caused by seasonal effects, possibly due to various weather trends in different months, or in the type of trends, plausibly subject to prolonged trends (Coles, 2001) [5] presented various methods in the literature review on discordant detection in high dimensional and functional data. Specifically, it introduced helpful graphs for visualizing uni variate instances of functional data and designed to detect functional discordant in the functional setting of discordant detection data techniques.

To detect discordant functional data graphically, such suggested graphical methods, such as functional bag plots and functional highest density region plots are displayed suggested the enhanced boxplot. This tool can visually recognize discordant(s) which may not be noticeable in the actual data plot and can observe functional data. Furthermore, illustrated techniques for displaying huge quantities of the functional dataset as well as designed to discordant functional data.

As a result, numerous functional methods of discordant detection have already been effectively utilized in a range of application areas, including, which implemented the functional depth technique for detecting of discordant(s) in urban gas emissions. For examining the discordant(s) of the daily flood-flow series, used functional discordant detection methods such as rainbow plots and the functional bag-plot and box-plot.

Recently, used rainbow plots to visualize huge quantities of hydrological data, as well as the functional/bivariate bag-plot and box-plot for graphically detecting discordant(s) and used the functional highest density region box-plot in the density-based approach to identify functional discordant(s) in flood data stream flow hydrographs.

Discordant detection methods are crucial in analysing data collected and designed to detect extreme points that may affect the validity of the results when evaluating rainfall data. In comparison to traditional analyses, the functional structure can be viewed as a suitable measure for gaining insight. Furthermore, it allows for a thorough overview of rainfall data by performing a single analysis of the entire data visuals instead of multiple analyses. The present study may be helpful to researchers investigating future research. The findings also could contribute significantly to the effective management of water. This work is based on the maximum rainfall of the Mahasamund region from 1981 to 2022 in order to recognize the discordant event of rainfall variation over long time intervals. The functional methods have an effective way of identifying discordant graphically, which might not be visible through the original data plot in classical analysis. This study's main objective is to detect the extreme rainfall events using discordant(s) detection methods. The entire range of possible limit distributions for extreme realizations of a random variable is given by the extremely type's theorem. The families of extreme value and generalized Pareto distributions are more widely used to perform extreme value analyses.

Coles (2001) [5] focuses on statistical inference for extreme ends and employs all of the aforementioned methods to analyse various data sets and create statistical inference for extreme ends. Once dealing with extreme events, the problem of discordant is a major concern. A discordant in statistics is a single observation that is "far away" from the entire data and can lead to unrealistic conclusions, particularly once trying to extrapolate to high sufficient quantiles of the variables under study. Smith (2003) [3], for example, investigated environmental datasets from this perspective, and a played crucial of track showings achieved by a Chinese athlete, including a new world record, which also raised instant doubts.

Robinson and Tawn (1995) [23] investigated the mentioned earlier series to determine the discordant performance. Non-stationary is an issue in environmental processes due to seasonal effects, possibly due to the different climatic change trends in different months, or in the form of trends, possibly subject to prolonged trends. Coles (2001) [5]. Davison and Smith (1990) [6] mentioned covariate modelling in the sense of the threshold excess model. The model's location and scale parameters have been assumed to be distributed as polynomial functions of time or as sinusoidal terms, which can be treated as entirely separate or closely related. They can also be expressed as follows from multiple other parameters: the frequency and severity of extreme events. The deviance statistic D is a statistical parameter used to identify the significance of such trends. Data were extracted from the tropical rainfall measuring mission and the analysis has been processed by R software.

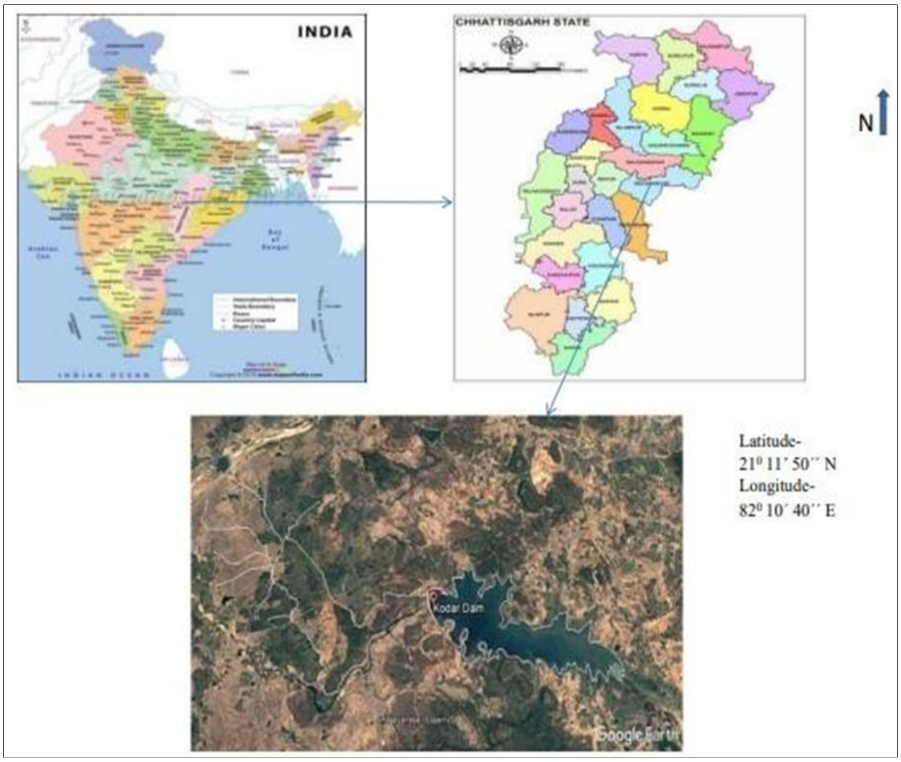

Fig 1: Location map of shahid veer Narayan Singh dam (Kodar Reservior) of Mahasamund district of Chhattisgarh

Table 1: Monthly average climate data of Mahasamund district

S. No.	Month	Max. Temp ()	Min Temp ()	Humidity ()	Wind Speed ()	Sun Shine ()	()
1.	January	23.86	9.86	65.27	2.09	8.3	3.89
2.	February	26.55	11.73	56.43	2.05	8.7	3.67
3.	March	35.21	19.01	35.57	2.34	9	3.49
4.	April	39.28	24.83	29.33	3.20	8.8	2.91
5.	May	43.53	29.23	22.93	3.37	5.8	2.09
6.	June	36.78	26.77	54.14	4.35	3.5	2.09
7.	July	32.07	24.69	76.29	4.23	4	2.19
8.	August	30.86	24.07	82.68	3.48	5.3	2.78
9.	September	33.57	21.74	54.76	3.08	7.7	3.59
10.	October	33.27	21.54	56.34	2.98	8.1	4.30
11.	November	30.57	16.86	57.78	2.07	8.3	4.40
12.	December	29.30	14.27	58	2.05	7.2	4.03
Average		23.86	9.86	65.27	2.09	8.3	3.29

Proposed test statistic

The new test statistic proposed here to test the null hypothesis that there is no upper discordant observation is present in the data set against the slippage alternative using the Weibull distribution in the maximum rainfall of Kodar reservior, is defined as

$$\left(\overline{\quad}\right),$$

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

The critical values

The critical values of the test statistic for the distribution under consideration are obtained by simulation technique as finding the theoretical density is a tedious job. Thus, simulation technique is applied to obtain critical values. As Lalitha and Tripathi (2016), shows that for extreme value distribution, critical values are not affected by location and scale parameters. But nothing can be said about the third parameter, the shape parameters which may affect the critical value of Z. Thus, in this paper the critical values are obtained for different values of at different levels of significances and sample size up to 50.

Computation of Z when Parameters are not known.

In practical situations, generally parameters are not known, in such a situation estimates of the parameter are used for computation of the test statistic. While dealing with the Weibull sample, estimates of all the three parameters can be obtained by using inbuilt functions in *R*- programming and the critical values of Z are obtained by simulation technique. Thus, the critical values will remain same for both the cases *i.e.* whether parameters are known or unknown.

Table 2: The critical values of Z at.

Parameters		1				1				
Sample Size		1%	5%	10%	1%	5%	10%	1%	5%	10%
=3	1	3.145	3.047	2.924	3.145	3.047	2.924	3.145	3.047	2.924
	2	2.435	2.309	2.151	2.019	1.869	1.682	0.491	0.451	0.401
	3	0.620	0.618	0.616	1.204	1.112	0.995	2.152	2.037	1.893
=4	1	4.112	3.776	3.358	4.112	3.776	3.358	4.112	3.776	3.358
	2	2.676	2.600	2.505	1.720	1.550	1.337	1.039	0.968	0.879
	3	1.927	1.892	1.847	0.871	0.799	0.708	0.966	0.855	0.716
=5	1	2.664	2.446	2.173	2.664	2.446	2.173	2.664	2.446	2.173
	2	3.854	3.727	3.569	3.977	3.586	3.096	3.688	3.430	3.108
	3	2.232	2.130	2.002	1.525	1.400	1.242	3.581	3.150	2.612
=6	1	0.864	0.844	0.819	0.864	0.844	0.819	0.864	0.844	0.819
	2	2.892	2.590	2.214	1.171	0.967	0.711	2.572	2.270	1.893
	3	2.373	2.244	2.082	1.057	0.931	0.773	1.606	1.467	1.294
=7	1	3.545	3.120	2.589	3.545	3.120	2.589	3.545	3.120	2.589
	2	4.619	4.601	4.577	0.656	0.569	0.461	2.542	2.282	1.957
	3	2.007	1.966	1.914	0.482	0.393	0.282	2.856	2.566	2.203
=8	1	3.349	3.303	3.245	3.349	3.303	3.245	3.349	3.303	3.245
	2	2.083	1.806	1.461	2.591	2.253	1.831	1.773	1.746	1.711
	3	3.141	3.001	2.826	2.658	2.234	1.705	2.775	2.511	2.181
=10	1	1.791	1.442	1.006	1.791	1.442	1.006	1.791	1.442	1.006
	2	2.050	1.867	1.639	2.917	1.856	0.528	3.339	2.609	1.697
	3	1.495	1.177	0.779	3.309	2.572	1.650	2.098	1.990	1.844
=12	1	2.631	1.938	1.117	2.631	1.938	1.117	2.631	1.938	1.117
	2	3.764	3.157	1.108	2.871	2.404	1.185	1.804	1.706	1.169
	3	6.572	4.209	1.169	1.970	1.652	1.011	1.257	1.201	1.112
=15	1	2.267	2.029	2.029	2.267	2.029	2.029	2.267	2.029	2.029
	2	3.517	2.756	2.038	4.146	3.858	2.488	2.815	2.700	2.455
	3	4.400	3.582	2.920	1.130	1.020	0.863	2.090	2.027	1.768
=20	1	2.864	2.696	1.592	2.864	2.696	1.592	2.864	2.696	1.592
	2	4.982	3.973	2.168	6.557	5.881	2.426	2.398	2.341	2.204

	3	4.328	4.055	2.872	0.173	0.216	0.240	2.944	2.734	2.627
=25	1	3.345	2.185	2.060	3.345	2.185	2.060	3.345	2.185	2.060
	2	4.133	3.200	2.140	6.573	6.115	4.554	3.727	2.201	1.493
	3	2.689	2.485	2.387	3.840	3.695	1.560	2.270	1.874	1.464
=30	1	2.831	2.559	2.163	2.831	2.559	2.163	2.831	2.559	2.163
	2	5.259	2.597	2.277	8.054	5.481	4.650	3.459	2.126	1.376
	3	4.630	2.786	2.648	5.257	3.867	2.935	2.801	2.387	2.054
=50	1	3.770	2.863	2.305	3.770	2.863	2.305	3.770	2.863	2.305
	2	4.030	2.765	2.208	9.555	6.668	4.637	3.491	3.165	2.846
	3	6.314	3.820	2.809	39.442	18.125	7.481	2.702	2.588	2.075

In Table 1, the critical values are shown, which are obtained for the test statistic by fixing two parameters namely scale parameter, location parameter at unity and for different values of shape parameter, varies from 1 to 3, for sample sizes = 3,4,5,…,10,15,20,25,30, and 50.

Example

In this section, an example, to validate our findings is been considered, for that the data set of 41 years of the maximum rainfall of the Shahid Veer Narayan Singh Dam (Kodar reservoir) of Mahasamund District of Chhattisgarh from the year of 1981-2022 have been taken from the office of sub divisional office Kodar head work, water resources sub division Mahasamund, Chhattisgarh.

Table 2: Maximum Rainfall of the Saheed Veer Narayan Singh Bandh (Kodar reservoir) Project

S. No.	Year	Maximum Rainfall	
		Total Rainfall (m.m.)	**Z value**
1.	1981-82	1156	1.0208
2.	1982-83	1361	1.9321
3.	1983-84	820	1.2355
4.	1984-85	1137	1.9551
5.	1985-86	1368	1.9703
6.	1986-87	1136	1.9517
7.	1987-88	741	1.1458
8.	1988-89	470	1.0100
9.	1989-90	794	1.2028
10.	1990-91	1336	1.7997
11.	1991-92	810	1.2225
12.	1992-93	1220	1.2643
13.	1993-94	1131	1.9350
14.	1994-95	1726	4.6723
15.	1995-96	988	0.5352
16.	1996-97	847	0.4631
17.	1997-98	955	0.1232
18.	1998-99	716	0.2730
19.	1999-2000	608	0.1596
20.	2000-2001	716	0.0520
21.	2001-2002	755	0.1232
22.	2002-2003	608	0.0520
23.	2003-2004	1117	0.8891
24.	2004-2005	824	0.2408
25.	2005-2006	1149	0.9963
26.	2006-2007	863	0.2970
27.	2007-2008	1184	1.1230
28.	2008-2009	686	0.0993
29.	2009-2010	1200	1.1844
30.	2010-2011	1172	1.1784
31.	2011-2012	1164	1.1493
32.	2012-2013	1328	1.7586
33.	2013-2014	1692	4.3467
34.	2014-2015	1415	2.2401
35.	2015-2016	1195	1.1650
36.	2016-2017	1065	0.7319
37.	2017-2018	1044	0.6741
38.	2018-2019	1322	1.7282
39.	2019-2020	1109	0.8636
40.	2020-2021	1266	1.4615
41.	2021-2022	1138	0.9585

Using R software, 278.1039, 3.042324 871.9575 are the estimated values obtained for location, scale and shape parameters respectively using 41-year rainfall data of Shahid Veer Narayan Singh Dam (Kodar reservior) of Mahasamund district of Chhattisgarh. It is very convincing to say that our proposed test statistics has successfully detected an upper discordant observation in the data set under consideration with probability 0.5853 at 10% level of significance.

Conclusion

In this paper, a new test statistic is proposed to detect a discordant observation from the rainfall data for known parameters and the case of unknown parameters is also discussed. The critical values of the proposed test statistic are obtained up to 50 sample size at different level of significances namely 1%, 5% and 10%. A real-life data is used to check the effectiveness of the statistic, maximum rainfall data of Shahid Veer Narayan Singh Dam (Kodar reservior) of Mahasamund district of Chhattisgarh is considered and suggested test statistic is performing good as at 10% level of significance.

The probability of successfully detecting the discordant observation is 0.5853 at 10% level of significance for the data under consideration.

Reference

1. Butler A. Statistical modeling of synthetic Oceano graphic extremes. Ph.D. thesis, Lancaster University, 2005.
2. Castillo E, Hadi AS, Balakrishnan N, Sarabia JM. Extreme value and related models with applications in Engineering and Science, Wiley Series in Probability and Statistics, 2005.
3. Chebana F, Dabo-Niang S, Ouarda TBMJ. Exploratory functional flood frequency analysis and outlier detection. Water Resources Research. 2012;48:W04514.
4. Coles SG, Tawn JA. A Bayesian analysis of extreme rainfall data, Applied Statistics. 1996;45(4):463-478.
5. Coles S. An Introduction to statistical modeling of extreme Values, Springer Series in Statistics, 2001.
6. Davison AC, Smith RL. Models for exceedances over high thresholds (With discussion J R. Stat. Soc., Ser. B. 1990;52:393-442.
7. Dixon MJ, Tawn JA. The effect of non-stationarity on sea-level estimation, Applied Statistics. 1999;48:135-151.
8. Febrero M, Galeano P, González-Manteiga W. A Functional Analysis of NOx levels: location and scale estimation and outlier detection. Computational statistics. 2007;22:411-427.
9. Febrero M, Galeano P, González-Manteiga W. Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. Environmetrics. 2008;19:331-345.
10. Ferraty F, Vieu P. Non parametric functional data analysis: theory and practice, Springer, New York, 2006.
11. Filzmosera P, Maronna R, Werner M. Outlier identification in high dimensions, Computational Statistics & Data Analysis. 2008;52:1694-1711.
12. Gaetan C, Grigoletto M. Smoothing sample extremes with Dynamic Models, Extremes. 2004;7:221-236.
13. Hael MA. Modeling of rainfall variability using functional principal component Method: A Case Study of Taiz Region, Yemen. Modeling Earth Systems and Environment, 2020.
14. Hyndman RJ. Computing and graphing highest density regions. The American Statistician. 1996;50:120-126.
15. Hyndman RJ, Shang HL. Rainbow Plots, bagplots, and boxplots for functional data. Journal of Computational and Graphical Statistics. 2010;19:29-45.
16. Huerta G, Sanso B. Time-varying models for extreme values, Department of Mathematics and Statistics, University of New Mexico.
17. Martínez Torres J. Detection of outliers in gas emissions from urban areas using functional data analysis. Journal of Hazardous Materials. 2010;186:144-149.
18. Pandey MD, Van Gelder PHAJM, Vrijling JK. The estimation of extreme quantiles of wind velocity using L-moments in the peaks-over-threshold approach, Journal of Structural Safety. 2001;23(2):179-192.
19. Ramsay JO. Functional Data Analysis (Springer Series in Statistics). Springer, New York, 2005.
20. Lopez-Pintado S, Romo J. On the concept of depth for functional data. Journal of the American Statistical Association. 2009;104:718-734.
21. Narisetty NN, Nair VN. Extremal depth for functional data and applications. Journal of the American Statistical Association, 2016;111:1705-1714.
22. Reiss RD, Thomas M. Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields, Second Edition, 2001.
23. Robinson ME, Tawn JA. Statistics for exceptional athletic records, Applied Statistics. 1995;44:499-511.
24. Smith RL. Statistics of extremes with applications in Environment, Insurance and Finance, University of North Carolina, 2003.
25. Suhaila J. Application of functional data analysis in stream flow hydrograph. Proceedings of the third international conference on computing, mathematics and statistics (iCMS2017), Springer, Berlin, 2019, 245-251.
26. Sun Y, Genton MG. Functional boxplots. Journal of Computational and Graphical Statistics. 2011;20:316-334.
27. Stephenson A, Tawn J. Bayesian inference for extremes: accounting for the three extremal types. 2004;7(4):291-307.
28. Shang HL. Rainbow: An R package for visualizing functional time series. The R Journal. 2011;3:54-59.