Rainfall Analysis by Power Transformation

HAMEED RASHEED, A. S. ALDABAGH AND MURUR V. RAMAMOORTHY

Department of Irrigation and Drainage, University of Mosul, Mosul, Iraq

(Manuscript received 6 November 1981, in final form 15 February 1983)

ABSTRACT

Power transformation was used to normalize the peak daily and peak monthly rainfall at various raingage stations in Iraq. Excellent correlations were found between the coefficient of skewness (C_s) and a parameter for power transformation (λ), coefficient of kurtosis (C_k) and λ, and between C_s and C_k. The relationship between C_s and λ is used to develop an estimation procedure for calculation of the transformation parameter. The method eliminates the use of trial and error for estimating λ. The method has been used for estimating T-year peak daily and peak monthly rainfall by power transformation. The results are compared with results from other methods.

1. Introduction

Proper estimates of peak rainfall are essential in the design of irrigation and drainage structures, and in soil erosion problems. Estimation of peak rainfall is much simpler when the data follow a normal distribution, but most hydrologic data are highly skewed and poorly fitted by the normal distribution function. Three alternatives are available to deal with such data: finding another suitable distribution, transformation, and nonparametric or distribution-free methods (Cooper and Clarke, 1980).

The first alternative is to use a method appropriate to other best-fitting distribution functions. Various distributions have been suggested for this purpose (Subcommittee of Hydrology, 1966; Yevjevich, 1972), but since records are generally short, often the best-fitting distribution cannot be identified for either rainfall or flood series (Benson, 1968). However, the log-Pearson Type III distribution has been recommended (Hydrology Committee, 1967) to U.S. Federal agencies for flood analysis. The other widely used distribution is the first asymptotic distribution of extreme values, commonly called the Gumbel distribution (Gumbel, 1958).

The second alternative is to transform the data to normality before applying methods appropriate to normal distributions. A normalizing transformation may be sought for any of three reasons: 1) to utilize simple properties of normal function, 2) to obtain a satisfactory fit to data, and 3) to avoid the use of functions that cannot be justified, though they give a good fit (Yevjevich, 1972).

Bethlahmy (1977) suggested a SMEMAX transformation based on the trigonometric solution of a right-angled triangle whose three vertices represent the smallest, median and the largest observed values of a sample. While points along the base and the height of the triangle represent observed values, suitably projected points of these values on the hypotenuse represent the transformed values. These transformed values have a nearly normal distribution for flood flows of 65 rivers in the northwestern United States.

Subash Chander *et al.* (1978) have shown that the power transformation suggested by Box and Cox (1964) transforms the data to a near-normal series better than the SMEMAX transformation. The power transformation suggested is

$$Y_i = \begin{cases} (X_i^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \log X_i, & \lambda = 0 \end{cases}, \qquad (1)$$

where the X_i are variates of the given series, Y_i are the transformed variates, and λ is a transformation parameter, generally $-1.0 < \lambda < +1.0$.

Because Eq. (1) cannot be solved in a closed form, Subash Chander *et al.* (1978) suggest estimation of the desired value of λ by the maximum likelihood method or by trial and error. The least-square fitting method is very popular in parameter estimation and is used in almost all regression analysis problems. In this paper a method based on this principle is suggested for the estimation of λ.

Rainfall data are available for 38 years (1923–60) for eight raingage stations and for less than 30 years in two stations in Iraq. Two series were considered for processing, peak monthly rainfall in millimeters per month (maximum of the 12 monthly values), and peak daily rainfall in millimeters per day in a year (maximum of 365 daily values). The year is taken as the calendar year, January to December.

2. Iterative process

A computer program transformed the daily extremes, with the transformation constant λ varying from -1.0 to $+1.0$ in steps of 0.1, and gave the av-

TABLE 1. Effect of various λ values on the parameters of transformed variables for peak daily rainfall at Mosul.

Transformation	λ	Mean $\bar{Y}$	Standard deviation σ	Coefficient of skewness C_s	Coefficient of kurtosis C_k	Coefficient of variation C_v
Untransformed	—	36.3316	12.0986	0.9502	3.7106	0.3330
Reciprocal	−1.000	0.9697	0.0091	−0.1795	2.2109	0.0094
	−0.900	1.0635	0.0130	−0.1331	2.2146	0.0122
	−0.800	1.1743	0.0184	−0.0857	2.2227	0.0157
	−0.700	1.3061	0.0262	−0.0372	2.2366	0.0201
	−0.600	1.4643	0.0373	0.0123	2.2367	0.0255
	−0.500	1.6557	0.0532	0.0630	2.2831	0.0321
	−0.400	1.8892	0.0758	0.1147	2.3614	0.0401
	−0.300	2.1767	0.1082	0.1675	2.3568	0.0497
	−0.200	2.5334	0.1546	0.2214	2.4048	0.0610
	−0.100	2.9797	0.2211	0.2764	2.4607	0.0742
Logarithmic	0.000	1.5385	0.1374	0.3325	2.5249	0.0893
	0.100	4.2583	0.4535	0.3897	2.5978	0.1065
	0.200	5.1749	0.6505	0.4479	2.6798	0.1257
	0.300	6.3596	0.9341	0.5072	2.7713	0.1469
	0.400	7.8941	1.3425	0.5676	2.8728	0.1701
Square-root	0.500	9.9035	1.9371	0.6289	2.9845	0.1950
	0.600	12.5478	2.7822	0.6913	3.1069	0.2217
	0.700	16.0484	4.0114	0.7546	3.2402	0.2500
	0.800	20.7084	5.7896	0.8189	3.3853	0.2796
	0.900	26.9446	8.3650	0.8841	3.5419	0.3105
	1.000	35.3316	12.0986	0.9502	3.7106	0.3424

erage $\bar{Y}$, standard deviation σ, coefficient of skewness C_s, coefficient of kurtosis C_k, and the coefficient of variation C_v of the transformed series. As λ varies from −1 to +1, the coefficient of skewness C_s changes from negative to positive (Table 1) in all series tried. For example, C_s changes from −0.0372 to 0.0123 when λ goes from −0.7 to −0.6. Obviously, the desired value of λ lies between −0.7 to −0.6 and must be found by further trials. This is called the iterative process for estimating λ.

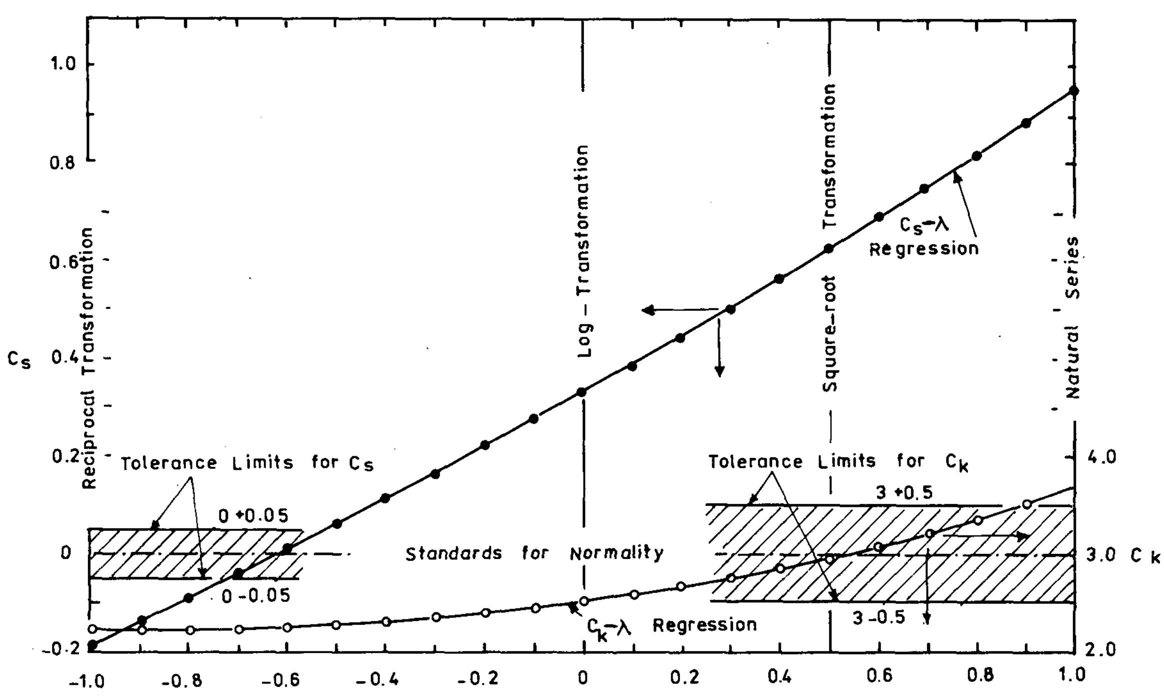

FIG. 1. Variations of C_s and C_k with λ for peak daily rainfall at Mosul.

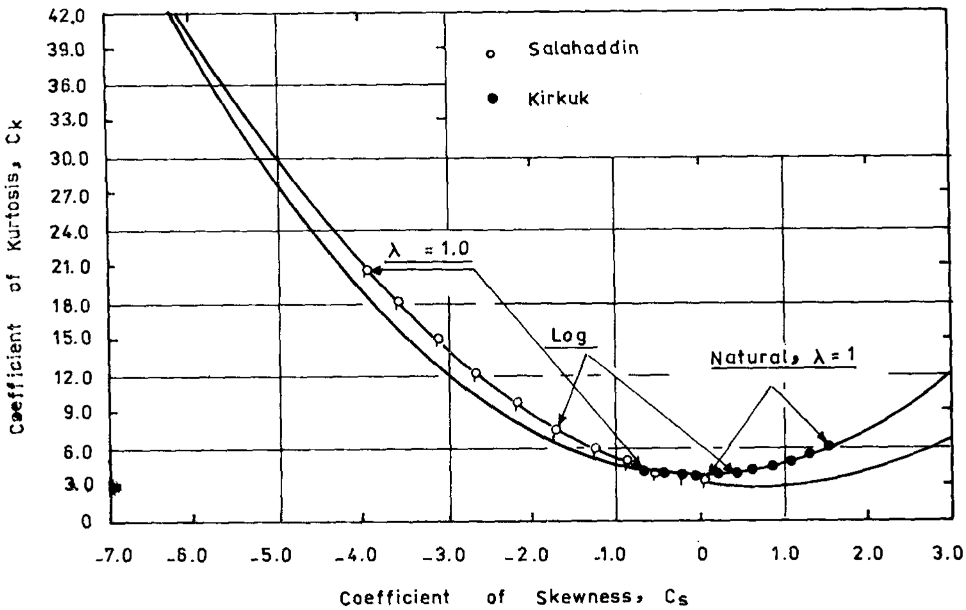

FIG. 2. C_s–C_k regression curves for peak daily rainfall at Salahaddin and Kirkuk.

A polynomial regression of C_s on λ of the form

$$C_s = b_0 + b_1\lambda + b_2\lambda^2 + \cdots + b_n\lambda^n \qquad (2)$$

was carried out for various values of n. An excellent correlation was observed for the second-degree polynomial (Fig. 1). Similar results were obtained for the regressions of C_k on λ (Fig. 1) and C_k on C_s for peak daily rainfall at Mosul. Tolerance limits as suggested by Yevjevich (1972) for acceptability of normality are also shown. Regressions of C_k on C_s for peak daily rainfall at Salahaddin and Kirkuk are shown in Fig. 2.

The regression lines reveal that both C_s and C_k cannot simultaneously satisfy the condition of normality. In other words, a value of λ cannot be chosen to suit the standard relating to either C_s or to C_k. Chow (1964) does not recommend moments of order higher than 3 for statistical analysis of hydrologic data because reliable estimates require record lengths longer than usually available. Therefore, the λ value is chosen to make the parameter C_s zero. In addition, the transformed values of the data were tested for normality of distribution using the chi-squared test of goodness of fit. The test indicated that there is no significant difference at the 5% probability level.

3. Estimation of λ by least-square fitting

Polynomial regression of C_s on λ (Fig. 1) reveals an excellent correlation. Since the curvature is small, the relation can be expressed alternatively as

$$\lambda = B(0) + B(1)C_s + B(2)C_s^2, \qquad (3)$$

and used to estimate the normalizing values of λ by simply substituting $C_s = 0$ in it. This substitution leads to

$$\lambda = B(0). \qquad (4)$$

The constants $B(0)$, $B(1)$, $B(2)$ in Eq. (3) were first estimated by least-squares using 11 values of C_s corresponding to λ = −1, −0.8, −0.6, −0.4, −0.2, 0, 0.2, 0.4, 0.6, 0.8 and 1 (Table 1). Such estimates are called 11-point estimates. Similarly 5-point estimates correspond to λ values of −1, −0.5, 0, +0.5 and 1 and 3-point estimates to λ values of −1, 0 and 1. Table 2 shows the values of $B(0)$, $B(1)$ and $B(2)$ and the correlation coefficient for peak daily rainfall at Mosul. Similar results were obtained for all other stations.

Values of $B(0)$, $B(1)$ and $B(2)$ vary little with sample size, with the correlation coefficient being nearly unity in all cases. Similar results were found for all other stations. Independence of $B(0)$, $B(1)$ and $B(2)$ from sample size, and the fact that the correlation coefficient is nearly unity, indicate a true functional relationship between λ and C_s which needs to be theo-

TABLE 2. Comparison of polynomial constants obtained with different sample sizes for peak daily rainfall at Mosul.

Estimate	11-point	5-point	3-point
$B(0)$	−0.6283	−0.6301	−0.6318
$B(1)$	1.9842	1.9894	1.9984
$B(2)$	−0.2907	−0.2911	−0.2958
Correlation	0.999964	0.999964	1.0000*
λ	−0.6283	−0.6301	−0.6318

* Since three constants are calculated from just three points, the curve necessarily passes exactly through the three points, and there are no degrees of freedom for estimating the correlation coefficient.

TABLE 3. Comparison of values of λ estimated by least-square fitting with different sample sizes and by iterative process [peak monthly rainfall (mm $month^{-1}$)].

Name of station	Years of record	11 points		5 points		3 points	Iterative process
		λ	Correlation	λ	Correlation	λ	λ
Mosul	1941–78	0.4256	0.999992	0.4261	0.999992	0.4262	0.424
Sinjaar	1941–78	0.3365	0.999805	0.3388	0.999803	0.3396	0.327
Salahaddin	1941–42 1953–78	0.8640	0.997720	0.8804	0.997636	0.9048	0.880
Kirkuk	1941–78	−0.1354	0.999997	−0.1355	0.999998	−0.1342	−0.135
Baghdad	1941–78	−0.3506	0.999601	−0.3540	0.999606	−0.3535	−0.336
Al-Rathba	1941–78	−0.1097	0.999900	−0.1091	0.999908	−0.1120	−0.115
Al-Najaff	1957–78	0.7224	0.997238	0.7373	0.997116	0.7619	0.697
Al-Haye	1941–78	0.1548	0.999997	0.1546	0.999998	0.1558	0.156
Al-Nasiriya	1941–78	0.6707	0.999969	0.6722	0.999970	0.6736	0.668
Basrah	1941–78	−0.0695	0.999984	−0.0694	0.999985	−0.0681	−0.068

retically proved. Therefore, the 3-point estimate is considered sufficient to evaluate λ.

Tables 3 and 4 show the comparison of the values of λ by the 3, 5, and 11-point estimates and the iterative process for peak monthly and daily rainfalls, respectively.

4. Comparison of rainfall estimates

SMEMAX and power transformations assume a normal distribution for the transformed data. Exceedance probability based on the normal distribution for various probability levels is used in estimating peak rainfall for various recurrence intervals. Table 5 shows the peak monthly and daily rainfall with 1% exceedance probability (recurrence interval of 100 years) for various stations. Different trends of similarity appear in the estimates.

Estimates of peak monthly rainfall by power transformation are, with one exception, closest to those from the log-Pearson Type III distribution for stations with more than 30 years of record. The 30-year value is generally accepted as a good minimum for analysis of hydrologic data. For stations with fewer years of record, power transformation and SMEMAX tend to give strikingly similar results. In 8 out of 10 stations, the Gumbel distribution gives higher values of estimates than others.

A slightly different picture is shown by peak daily rainfall estimates. While Gumbel and log-Pearson Type III distribution results are close to those obtained by power transformation, SMEMAX offers consistently lower estimates than power transformation.

5. Conclusions

From polynomial regressions of λ on C_s and on C_k and of C_k on C_s, the following conclusions can be drawn:

1) A second-degree polynomial equation represents the relations between these pairs of statistical parameters with excellent correlation.

2) To estimate the parameter λ of the power transformation, C_s is computed using three values of λ for

TABLE 4. Comparison of values of λ estimated by least-square fitting with different sample sizes and by iterative process [peak daily rainfall (mm day^{-1})].

Name of station	Years of record	11 points		5 points		3 points	Iterative process
		λ	Correlation	λ	Correlation	λ	λ
Mosul	1941–78	−0.6283	0.999964	−0.6301	0.999964	−0.6318	−0.625
Sinjaar	1941–78	−0.3825	0.999969	−0.3815	0.999970	−0.3797	−0.388
Salahaddin	1941–42 1953–78	0.9298	0.997690	0.9465	0.997994	0.9757	0.969
Kirkuk	1941–78	−0.4014	0.999895	−0.4030	0.999907	−0.4053	−0.394
Baghdad	1941–78	−0.3332	0.999938	−0.3340	0.999946	−0.3364	−0.328
Al-Rathba	1941–78	−0.2781	0.999957	−0.2780	0.999939	−0.2816	−0.274
Al-Najaff	1957–78	0.6129	0.999749	0.6169	0.999777	0.6206	0.604
Al-Haye	1941–78	−0.2960	0.999993	−0.2963	0.999993	−0.2970	−0.295
Al-Nasiriya	1941–78	0.4657	0.999766	0.4684	0.999793	0.4728	0.455
Basrah	1941–78	−0.3349	0.999882	−0.3369	0.999896	−0.3356	−0.327

TABLE 5. Peak monthly and daily rainfalls (mm) with 1% exceedance probability.

Station	Years	Gumbel Type I	SMEMAX	Log-Pearson Type III	Power trans-formation
Monthly					
Mosul	38	214.75	185.56	188.05	188.15
Sinjaar	38	296.80	251.03	268.57	265.53
Salahaddin	28	516.09	408.02	347.82	407.46
Kirkuk	38	292.57	271.80	281.43	281.20
Baghdad	38	156.25	146.80	161.11	166.27
Al-Rathba	38	127.72	116.98	131.89	132.94
Al-Najaff	22	119.05	96.13	79.60	95.80
Al-Haye	38	149.28	133.82	140.60	139.48
Al-Nasiriya	38	105.99	90.28	91.76	89.29
Basrah	38	153.98	143.36	.151.01	151.95
Daily					
Mosul	38	86.90	74.21	77.84	86.44
Sinjaar	38	112.39	106.14	113.86	124.57
Salahaddin	28	128.54	99.51	85.32	100.69
Kirkuk	38	103.95	96.59	104.80	106.82
Baghdad	38	81.68	74.26	83.76	88.24
Al-Rathba	38	68.88	63.57	75.83	79.16
Al-Najaff	22	65.96	51.43	55.03	54.10
Al-Haye	38	90.73	86.89	95.74	100.72
Al-Nasiriya	38	59.95	50.99	53.69	52.45
Basrah	38	79.51	77.43	79.11	80.01

fitting λ as a second-degree polynomial in C_s. The constant term of the second-degree polynomial gives the normalizing value of λ.

3) Invariability in the values of normalizing λ with sample size and the fact that the correlation coefficient is nearly unity in all cases indicate a true functional relationship of second degree between λ and C_s. This needs to be further investigated for a theoretical proof.

Rainfall estimates of various return periods were found using some popular distribution functions and some transformation techniques. These estimates lead to the following conclusions:

1) The annual series of monthly peak rainfall data shows that the log-Pearson Type III distribution offers estimates of rainfall similar to those of a power transformation. However, for data of less than 30 years period, a SMEMAX transformation gives estimates similar to those of a power transformation. A Gumbel distribution gives higher estimates.

2) For daily peak rainfall data, log-Pearson Type III and Gumbel distributions give estimates close to those obtained by power transformation. SMEMAX transformation offers consistent lower estimates.

Acknowledgment. The authors appreciate the useful suggestions offered by Dr. Sudesh Kumar Spolia while preparing the paper. The authors are also grateful to the reviewers for their valuable suggestions in revising the original manuscript.

APPENDIX

Notation

The following symbols are used in this paper:

$B(0)$, $B(1)$, $B(2)$, b_0, b_1, b_2 are polynomial constants
C_k coefficient of kurtosis
C_s coefficient of skewness
C_v coefficient of variation
X rainfall of specified probability
X_i ith value of original series
$\bar{X}$ mean of the original series
$\bar{Y}$ mean of the transformed series
Y_i ith value of transformed series
λ parameter for power transformation
σ standard deviation.

REFERENCES

Benson, M. A., 1968: Uniform flood-frequency estimating methods for federal agencies. *Water Resour. Res.*, **4**, 891–898.
Bethlahmy, N., 1977: Flood analysis by SMEMAX transformation. *J. Hydraul. Div., ASCE,* **HY-1,** 69–80.
Box, G. E. P., and D. R. Cox, 1964: An analysis of transformations. *J. Roy. Statist. Soc.*, **B26,** 211–252.
Chow, V. T., 1964: *Handbook of Applied Hydrology.* McGraw-Hill 8.8–8.9.
Cooper, D. M., and R. T. Clarke, 1980: Distribution-free methods for estimating flood and streamflow exceedance probabilities by correlation. *Water Resour. Res.*, **16,** 1121–1127.
Gumbel, E. J., 1958: *Statistics of Extremes.* Columbia University Press.
Hydrology Committee, 1967: A uniform technique for determining flood flow frequencies. *Water Resour. Comm.*, Bull. No. 15.
Chander, S., S. K. Spolia and A. Kumar, 1978: Flood frequency analysis by power transformation. *J. Hydraul. Div., ASCE,* **HY-11,** 1495–1504.
Subcommittee of Hydrology, 1966: Methods of flow frequency analysis. Inter-Agency Comm. Water Resour., *Notes Hydrol. Act,* Bull. No. 13.
Yevjevich, V., 1972: *Probability and Statistics in Hydrology.* Water Resources Publications, Fort Collins, 118–181.