PROBABILITY DISTRIBUTION OF ANNUAL FLOOD DATA IN THAILAND

Chavalit Chaleeraktrakoon, Wiwat Sarochnanjeen and Kriengkrai Saengratwatchara
Department of Civil Engineering, Faculty of Engineering, Thammasat University
Klong Luang, Pathumthani 12121
Phone (66-2) 564-3001-9 Ext. 3100
Fax (66-2) 564-3001-9 Ext. 3039
E-mail: cchava@engr.tu.ac.th

Abstract

The main objective of this paper is to search for a practical distribution of annual flood records in Thailand using popularly linear moment diagrams. The diagrams of 163 sequences of instantaneously annual flood data (Thailand) with the total record period ranging from 20 to 77 years were constructed. The theoretical moment relationships of various distributions were fitted to the observed diagrams. Results have indicated that the linear moment relationships of generalized extreme value (GEV) distribution can approximate best to the observed moment diagrams. The GEV is shown to be adequate for describing the data at several tested sites. In addition, the existing log-Pearson type 3 distribution generally gives greater flood-quantile estimates than the GEV model does.

1. Introduction

Probability distribution of annual flood records is fundamental information for planning and design of various water resources engineering projects. However, the true probability distribution of the flood data is usually unknown. Even if it were, its mathematical expression would be too complicated to use. One common approach is to select a simple and reasonable distribution function that may appear to fit the observed frequencies of the flood data at a considered site. Unfortunately, the distribution function obtained from the at-site frequency analysis is too sensitive because of the sampling variation of the limited flood data.

The preferred alternative approach is to assess the goodness of fit of various probability distributions to regional samples of annual flood using linear moment diagrams [1-4]. The diagrams are advantageous to the other goodness-of-fit statistics (e.g., Chi square, Kolmogorov Smirnov, and probability plot correlation coefficient). They can compare the fit of several distributions to a number of sample data by only few graphs, and indicate the true distribution of a phenomenon interested [5]. While, the other mentioned statistics usually accept several tested distributions, and thus cannot identify the represented distributions of the phenomenon [1]. Further, the estimates of linear moment statistics are almost unbiased for all distributions, as compared with those of significantly biased product moments [4].

Consequently, many researchers have applied them to select the probability distribution of annual maximum flows for several regions of the world; such as Australia, New Zealand, Canada and U.S.A. {e.g., [4] and [6-8]}. These studies often chose a generalized extreme value (GEV) distribution for modeling the empirical frequencies of annual flood records within the regions. Unfortunately, Thailand and the referred regions have different factors generating flood phenomenon (e.g., physical basin properties and characteristics of rainfall process). It is thus uncertain whether the suggested GEV is consistent with the flood data in the region. Moreover, the linear moment diagrams have never been used in any previously flood frequency analysis in Thailand. Most investigations used Chi square and Kolmogorov-Smirnov statistics to evaluate the goodness-of-fit of hypothetically distribution functions, and selected a log-Pearson type 3

RECEIVED 8 MAY, 2002
ACCEPTED 20 AUGUST, 2002

(LP3) distribution as the model of annual flood data {see, e.g., [9-10]}.

The main objective of this study is to search for the practical probability distribution of annual flood data in Thailand using linear moment diagrams. The diagrams were constructed based on the instantaneously annual flood records of over 163 locations for 25 major river basins around Thailand. The total periods of the flood data considered range from 20 to 77 years. The observed diagrams were described using the moment relationships of various distributions [e.g., families of log normal (LN), GEV and Pearson type 3 (PE3); and generalized logistic (GLOG)]. It will be shown in this paper that the GEV is acceptable for modeling the annual flood data (Thailand), like the results obtained from the other regions.

2. Linear Moment Diagrams

Let $X_{i,j}$ be the ranked j-th flow record in descending order at station i for $i = 1,\ldots,m$ (m = the number of stations); and $j = 1,\ldots,n_i$ (n_i = the period in years of available flow records at site i). The construction of linear moment diagrams begins with estimating the r-th order probability-weighted moment (PWM) $\hat{b}_{i,r}$ of $b_{i,r}$ for r = 0, 1, ..., 3. Hence, the unbiased PWM estimate of order zero -- $\hat{b}_{i,0}$ -- is

$$\hat{b}_{i,0} = \overline{X}_i \qquad (1)$$

in which $\overline{X}_i$ = the mean of the flow data. For $r \geq 1$, the unbiased estimate $\hat{b}_{i,r}$ is generally given as

$$\hat{b}_{i,r} = \frac{1}{(r+1)} \sum_{j=1}^{n_i - r} \frac{\binom{n_i - j}{r} X_{i,j}}{\binom{n_i}{r+1}} \qquad (2)$$

Next, calculate the $(r + 1)$-th order linear-moment estimate $\hat{\lambda}_{i,r+1}$ of $\lambda_{i,r+1}$ using the following relationship [11]:

$$\hat{\lambda}_{i,r+1} = \sum_{k=0}^{r} q_{r,k} \hat{b}_{i,k} \qquad (3)$$

in which $q_{r,k}$ is the coefficient of PWM linear combination. The coefficient $q_{r,k}$ can be expressed by

$$q_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \qquad (4)$$

It should be noted that the linear moment statistics are similar to the product moment ones. In particular, the first-order linear moment $\lambda_{i,1}$ is a measure of location, and the second-order one $\lambda_{i,2}$ is a measure of scale. Notice also in **(2)** and **(3)** that the $(r + 1)$-th order linear-moment estimate is the linear combination of the ranked data. It does not involve squaring or cubing observations as usually do in product moment calculation. Thus, they are almost unbiased for various distributions, no matter how high the order is [4].

Now, denote $\tau_{i,2}$ as the linear coefficient of variation, and $\hat{\tau}_{i,2}$ as its sample estimate. The following step is to estimate the linear coefficient of variation $\hat{\tau}_{i,2}$ as

$$\hat{\tau}_{i,2} = \frac{\hat{\lambda}_{i,2}}{\hat{\lambda}_{i,1}} \qquad (5)$$

Then, compute the sample linear-skewness estimate $\hat{\tau}_{i,3}$ of $\tau_{i,3}$, and the sample linear-kurtosis one $\hat{\tau}_{i,4}$ of $\tau_{i,4}$ by

$$\hat{\tau}_{i,r} = \frac{\hat{\lambda}_{i,r}}{\hat{\lambda}_{i,2}} \qquad (6)$$

Repeat all calculation steps for every site, i = 1, 2, ..., m.

The next step is to plot all sample estimates of $\hat{\tau}_{i,2}$ versus $\hat{\tau}_{i,3}$ and $\hat{\tau}_{i,3}$ against $\hat{\tau}_{i,4}$. Draw the observed relationships using locally weighted scatter smoothes – LOWESS – {see [12]}. Now, the step is to fit the theoretical relationships of the linear moment ratios (τ_r versus τ_{r+1} for $r = 2, 3$) for several considered distributions [e.g., two- and three-parameter lognormal (LN2 and LN3), Gumbel, GEV, gamma (GAM), PE3, LP3, and GLOG]. Note that, for any given probability distribution function, τ_r and τ_{r+1} are usually expressed in terms of the parameters of the distribution which one can use the expressions to determine the theoretical relationships between τ_r and τ_{r+1} in exact form. However, to facilitate the construction of linear moment diagrams, polynomial approximation to such a relationship is commonly used as

$$\tau_{r+1} = \sum_{k=0}^{s} c_k \tau_r^k \qquad (7)$$

in which c_k is the constant for each distribution as shown in Table 1 {see [2]}, and $s = 6$ and 8 for the referred two- and three-parameter distributions respectively. The distribution having its theoretical relationship agrees best with the observed one is then chosen for describing the empirical frequencies of the regional flow data.

3. Annual Flood Distribution

The available records of naturally instantaneous annual flood with the minimum record length of 20 years have been considered in this study. Figure 1 illustrates the locations of 163 chosen sites within 25 major river basins across Thailand. The record length of the observed flood data ranges from 20 to 77 years with the average of 29 years. (1) – (7) as presented earlier were applied to the selected flood data. Results of the illustrated example can be shown as follows.

Figure 2 compares the theoretical relationships between linear coefficients of variation $\tau_{i,2}$ and skewness $\tau_{i,3}$ of LN2, GAM and Gumbel with the appropriate observed one [see numerical results in Table 2]. It is seen that the 2-parameter distributions do not describe the observed relationship adequately. Figure 3 presents the observed and fitted relationships of linear skewness and kurtosis coefficients ($\tau_{i,3}$ and $\tau_{i,4}$) for several 3-parameter distributions (LN3, PE3, GLOG and GEV) [numerical results illustrated in Table 3]. The comparison of observed and theoretically LP3 relationships between $\tau_{i,3}$ and $\tau_{i,4}$ is shown in Figure 4 [numerical results presented in Table 4]. They appear that the theoretical GEV relation agrees best with the observed one, as compared with the other distributions. It is therefore selected as the distribution for the annual flood data, like the results obtained from the other mentioned regions. The GLOG and LN3 can describe only some parts of the observed relationship curve. In particular, the GLOG model is accurate for the upper part, while the LN3 function is precise for the lower one. The LP3 distribution function suggested by [10] yields the poorest description of the observed relation of $\tau_{i,3}$ versus $\tau_{i,4}$.

Since the GEV distribution is chosen based on the accuracy for approximating the regional observed relationships $\tau_{i,3}$ against $\tau_{i,4}$, it is worth to investigate whether the GEV model is adequate for calculating flood quantile estimates at a given site. Hence, probability plots {i.e., $X_{i,j}$ against $\hat{X}_{i,j}$, where $\hat{X}_{i,j}$ = the GEV flood estimate [1]} were used to show its adequacy graphically. The plots would fall approximately on a straight line if the postulated GEV distribution were adequate for modeling the data at the location.

Figure 5 shows the GEV probability plots of the annual flood data at station P.21 (Ping River), E.23 (Chi River), and Y.3A (Yom River). These flood quantiles at the return periods of 5, 10, 20 and 50 years are also compared with the existing LP3 model [10]. The plots of the flood data for the remaining stations can be seen in [13]. It indicates that the

GEV distribution can model adequately the annual flood records. The flood quantiles of existing LP3 model are generally greater than those of recommended GEV distribution.

4. Summary and Conclusions

The main objective of this study is to choose the practical distribution function of annual flood records in Thailand based on linear moment diagrams. The diagrams are preferable to the other goodness-of-fit statistics because they can indicate the best fitted distribution to the data at many locations in a region by only few graphs. Further, the linear-moment sample estimates are approximately unbiased for various distributions.

The instantaneously annual flood records of 163 stations in 25 major river basins across Thailand were used to construct the linear moment diagrams. The total periods of such flood series range from 20 to 77 years. Results of fitting several distributions to the observed diagrams indicate that the GEV distribution has its moment relationships that can describe best the observed diagrams. The GEV distribution is shown to be adequate through the probability plots of flood data at several locations. Moreover, the flood quantile estimates using the suggested GEV are generally smaller than those applying the existing LP3.

5. Acknowledgments

The authors would like to express their appreciation to Royal Irrigation Department, The Electricity Generating Authority of Thailand and The Department of Energy Development and Promotion for providing the annual flood data used in this study.

References

[1] J. R., Stedinger, R. M., Vogel, and E., Foufoula-Georgiou. "Frequency Analysis of Extreme Events." *Handbook of Hydrology*, D. R. Maidment, ed., McGraw Hill Book Co., Inc., New York, N.Y., pp. 18.1-18.66, 1993.

[2] J. R. M., Hosking, and J. R., Wallis. "Some Statistics Useful in Regional Frequency Analysis." *Water Resources Research*, 29(2), pp. 271-281, 1993.

[3] J. U., Chowdhury, J. R., Stedinger, and L. H., Lu. "Goodness-of-Fit Tests for Regional Generalized Extreme Value Flood Distributions" *Water Resources Research*, 27(7), pp. 1765-1776, 1991.

[4] R. M., Vogel, and N. M., Fennessey. "L-Moment Diagrams Should Replace Product Moment Diagrams." *Water Resources Research*, 29(6), pp. 1745-1752, 1993.

[5] R. M., Vogel, and I., Wilson. "Probability Distribution of Annual Maximum, Mean, and Minimum Streamflows in the United States." *Journal of Hydrologic Engineering*, ASCE, 1(2), pp. 69-76, 1996.

[6] C. P., Pearson. "New Zealand Regional Flood Frequency Analysis Using L-moments." *Journal of Hydrology – New Zealand*, 30(2), pp. 53-64, 1991.

[7] D., Gringas, and K., Adamowski. "Coupling of Nonparametric Frequency and L-moment Analysis for Mixed Distribution Identification." *Water Resource Bulletin*, 28(2), pp. 263-272, 1992.

[8] R. J., Nathan. and P. E., Weinmann. "Application of At-site and Regional Flood Frequency Analysis." *International Hydrology and Water Resource Symposium*, pp. 769-774, 1991.

[9] M. A., Sabur. "Regional Flood Frequency Analysis of Thailand." *M. Eng. Thesis*, Asian Institute of Technology, Bangkok, Thailand, 1982.

[10] S., Compliew. "Magnitude and Frequency of Flood in the Northern and North-Eastern Regions of Thailand." *M. Eng. Thesis*, Chulalongkorn University, Bangkok, Thailand, 1999.

[11] J. R. M., Hosking. "L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics." *Journal of Royal Statistical Soceity, Serial B*, London, England, 52, pp. 105-124, 1990.

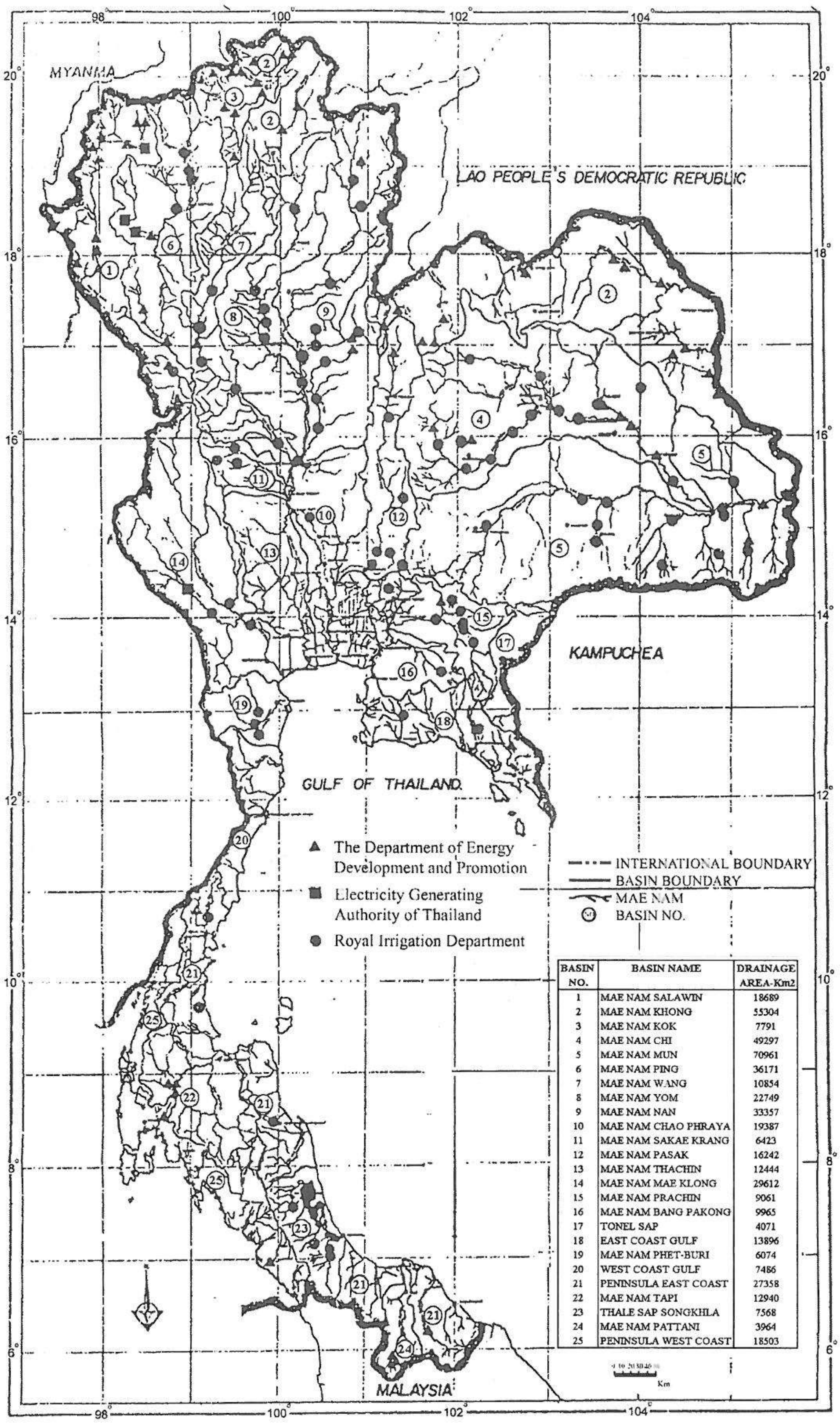
[12] D. R., Helsel, and R. M., Hirsch, *Statistical Methods in Water Resources*, Elsevier, Amsterdam, 1992.

[13] W., Sarochnanjeen. "Frequency Analysis Annual Maximum and Minimum Flows in Thailand" *M. Eng. Thesis*, Thammasat University, Bangkok, Thailand, 2001.

Table 1 Coefficients of approximate relationships between τ_2 and τ_3, and τ_3 and τ_4 for several considered distributions.

Coefficient, c_k	Two-parameter Distribution (τ_2 and τ_3)			
	LN2	GAM	Gumbel	
c_0	0	0	0.1699	
c_1	1.16008	1.74139	0	
c_2	-0.05325	0	0	
c_3	0	-2.59736	0	
c_4	-0.10501	2.09911	0	
c_5	0	0	0	
c_6	-0.00103	-0.35948	0	
Coefficient, c_k	Three-parameter Distribution (τ_3 and τ_4)			
	GEV	GLO	LN3	PE3
c_0	0.10701	0.16667	0.12282	0.12240
c_1	0.11090	0	0	0
c_2	0.84838	0.83333	0.77518	0.30115
c_3	-0.06669	0	0	0
c_4	0.00567	0	0.12279	0.95812
c_5	-0.04208	0	0	0
c_6	0.03763	0	-0.13638	-0.57488
c_7	0	0	0	0
c_8	0	0	0.11368	0.19383

Note: The coefficients of PE3 can be used to approximate the relationship of log-transformed LP3 variable.



BASIN NO.	BASIN NAME	DRAINAGE AREA-Km2
1	MAE NAM SALAWIN	18689
2	MAE NAM KHONG	55304
3	MAE NAM KOK	7791
4	MAE NAM CHI	49297
5	MAE NAM MUN	70961
6	MAE NAM PING	36171
7	MAE NAM WANG	10854
8	MAE NAM YOM	22749
9	MAE NAM NAN	33357
10	MAE NAM CHAO PHRAYA	19387
11	MAE NAM SAKAE KRANG	6423
12	MAE NAM PASAK	16242
13	MAE NAM THACHIN	12444
14	MAE NAM MAE KLONG	29612
15	MAE NAM PRACHIN	9061
16	MAE NAM BANG PAKONG	9965
17	TONEL SAP	4071
18	EAST COAST GULF	13896
19	MAE NAM PHET-BURI	6074
20	WEST COAST GULF	7486
21	PENINSULA EAST COAST	27358
22	MAE NAM TAPI	12940
23	THALE SAP SONGKHLA	7568
24	MAE NAM PATTANI	3964
25	PENINSULA WEST COAST	18503

Figure 1 Locations of chosen flow gauging stations within 25 major river basins in Thailand

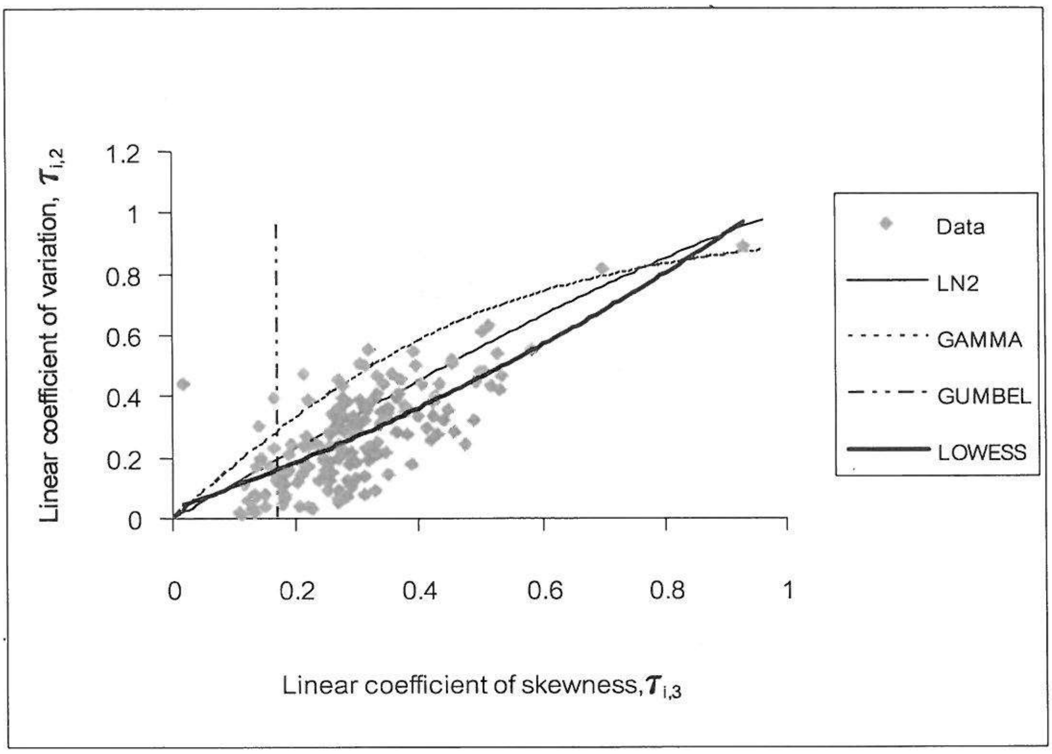


Figure 2 Observed and theoretical relationships between the linear coefficients of variation, $\tau_{i,2}$ and skewness, $\tau_{i,3}$ for annual flood records (2-parameter distributions)

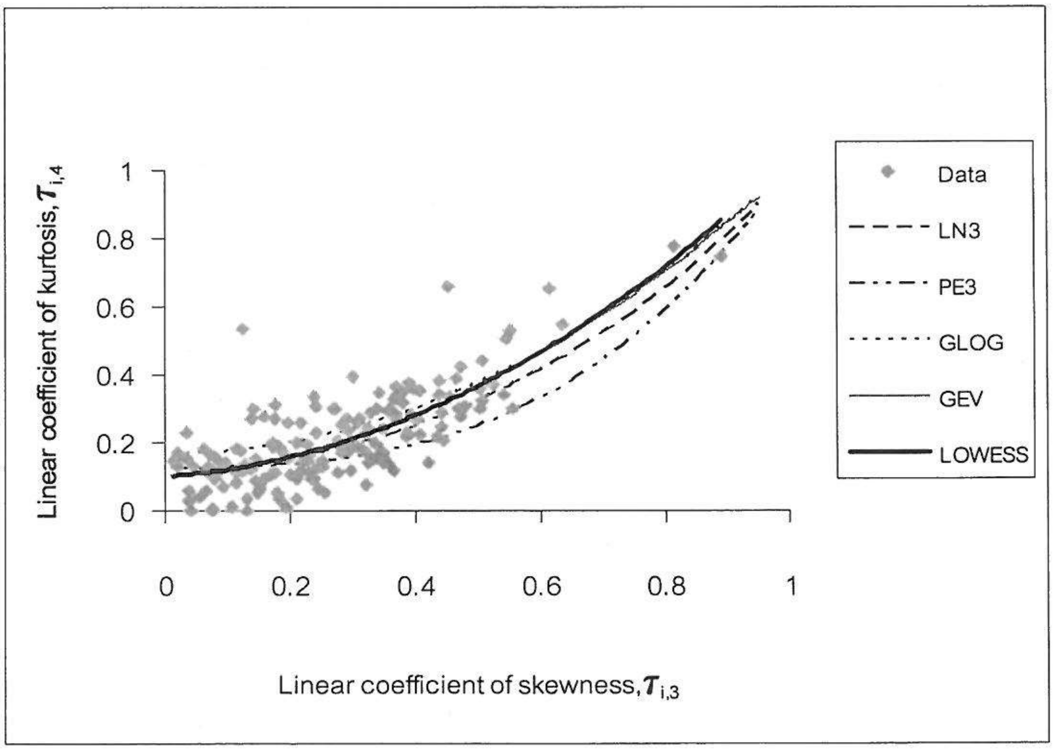


Figure 3 Observed and theoretical relationships between the linear coefficients of skewness, $\tau_{i,3}$ and kurtosis, $\tau_{i,4}$ for annual flood records (3-parameter distributions)

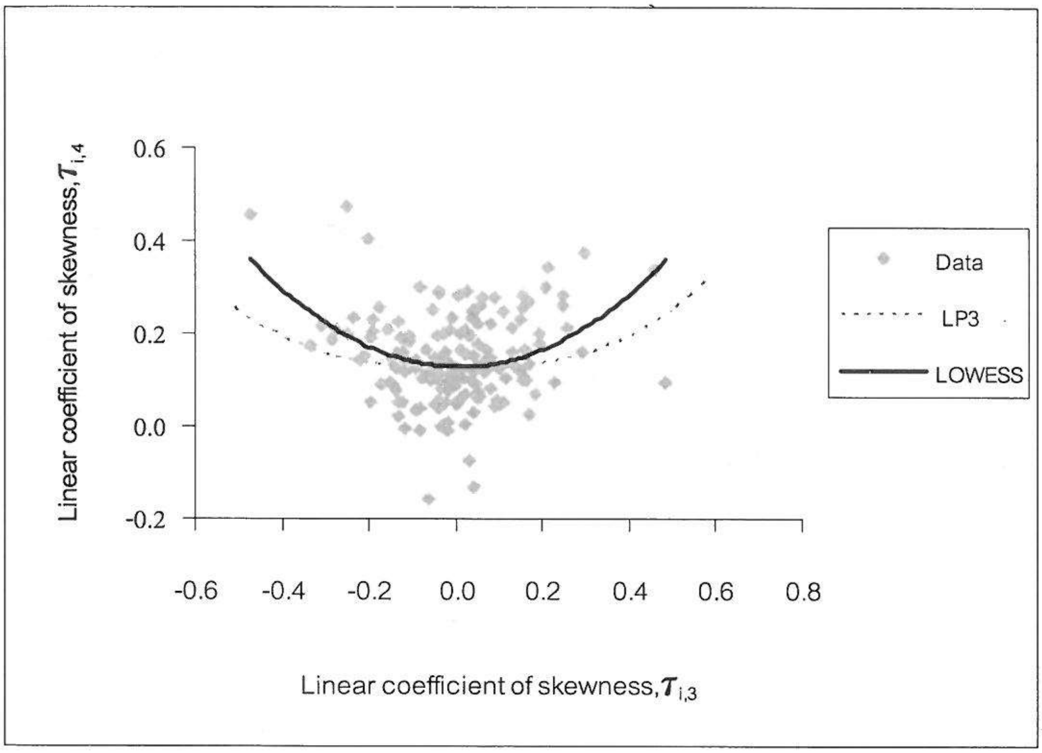


Figure 4 Observed and theoretical relationships between the linear coefficients of skewness, $\tau_{i,3}$ and kurtosis, $\tau_{i,4}$ for the logarithms of annual flood records (3-parameter distributions)

Table 2 Numerical results of observed and theoretical relationships between the linear coefficients of variation, $\tau_{i,2}$ and skewness, $\tau_{i,3}$ for annual flood records (2-parameter distributions)

coefficients	coefficients of variation, $\tau_{i,2}$		
of skewness, $\tau_{i,3}$	LOWESS	GAMMA	LN2
0.050	0.070	0.087	0.058
0.100	0.107	0.172	0.115
0.150	0.145	0.254	0.173
0.200	0.186	0.331	0.230
0.250	0.227	0.403	0.286
0.300	0.271	0.469	0.342
0.350	0.316	0.529	0.398
0.400	0.363	0.583	0.453
0.450	0.412	0.630	0.507
0.500	0.462	0.672	0.560
0.550	0.515	0.708	0.612
0.600	0.569	0.739	0.663
0.650	0.624	0.766	0.713
0.700	0.682	0.790	0.761
0.750	0.741	0.810	0.807
0.800	0.802	0.829	0.851
0.850	0.864	0.845	0.892
0.900	0.928	0.860	0.931
0.950	0.994	0.873	0.968

Note: Gumbel distribution has a constant $\tau_{i,3}$ of 0.1699 for any value of $\tau_{i,2}$

Table 3 Numerical results of observed and theoretical relationships between the linear coefficients of skewness, $\tau_{i,3}$ and kurtosis, $\tau_{i,4}$ for annual flood records (3-parameter distributions)

coefficients	coefficients of kurtosis, $\tau_{i,4}$				
of skewness, $\tau_{i,3}$	LOWESS	GEV	GLOG	PE3	LN3
0.050	0.109	0.115	0.169	0.123	0.125
0.100	0.122	0.127	0.175	0.126	0.131
0.150	0.138	0.143	0.185	0.130	0.140
0.200	0.159	0.163	0.200	0.136	0.154
0.250	0.183	0.187	0.219	0.145	0.172
0.300	0.212	0.215	0.242	0.157	0.193
0.350	0.245	0.247	0.269	0.173	0.219
0.400	0.281	0.283	0.300	0.193	0.250
0.450	0.322	0.322	0.335	0.218	0.284
0.500	0.366	0.366	0.375	0.249	0.323
0.550	0.415	0.413	0.419	0.287	0.366
0.600	0.468	0.464	0.467	0.331	0.413
0.650	0.524	0.518	0.519	0.383	0.466
0.700	0.585	0.576	0.575	0.444	0.523
0.750	0.649	0.638	0.635	0.512	0.585
0.800	0.718	0.703	0.700	0.589	0.653
0.850	0.791	0.772	0.769	0.676	0.727
0.900	0.867	0.844	0.842	0.773	0.808
0.950	0.948	0.921	0.919	0.881	0.898

Table 4 Numerical results of observed and theoretical relationships between the linear coefficients of skewness, $\tau_{i,3}$ and kurtosis, $\tau_{i,4}$ for the logarithms of annual flood records (3-parameter distributions)

coefficients of	coefficients of kurtosis, $\tau_{i,4}$	
skewness, $\tau_{i,3}$	LOWESS	LP3
-0.50	0.389	0.249
-0.45	0.340	0.218
-0.40	0.296	0.193
-0.35	0.258	0.173
-0.30	0.224	0.157
-0.25	0.195	0.145
-0.20	0.172	0.136
-0.15	0.153	0.130
-0.10	0.140	0.126
-0.05	0.132	0.123
0.00	0.128	0.122
0.05	0.130	0.123
0.10	0.137	0.126
0.15	0.149	0.130
0.20	0.166	0.136
0.25	0.188	0.145
0.30	0.215	0.157
0.35	0.247	0.173
0.40	0.284	0.193

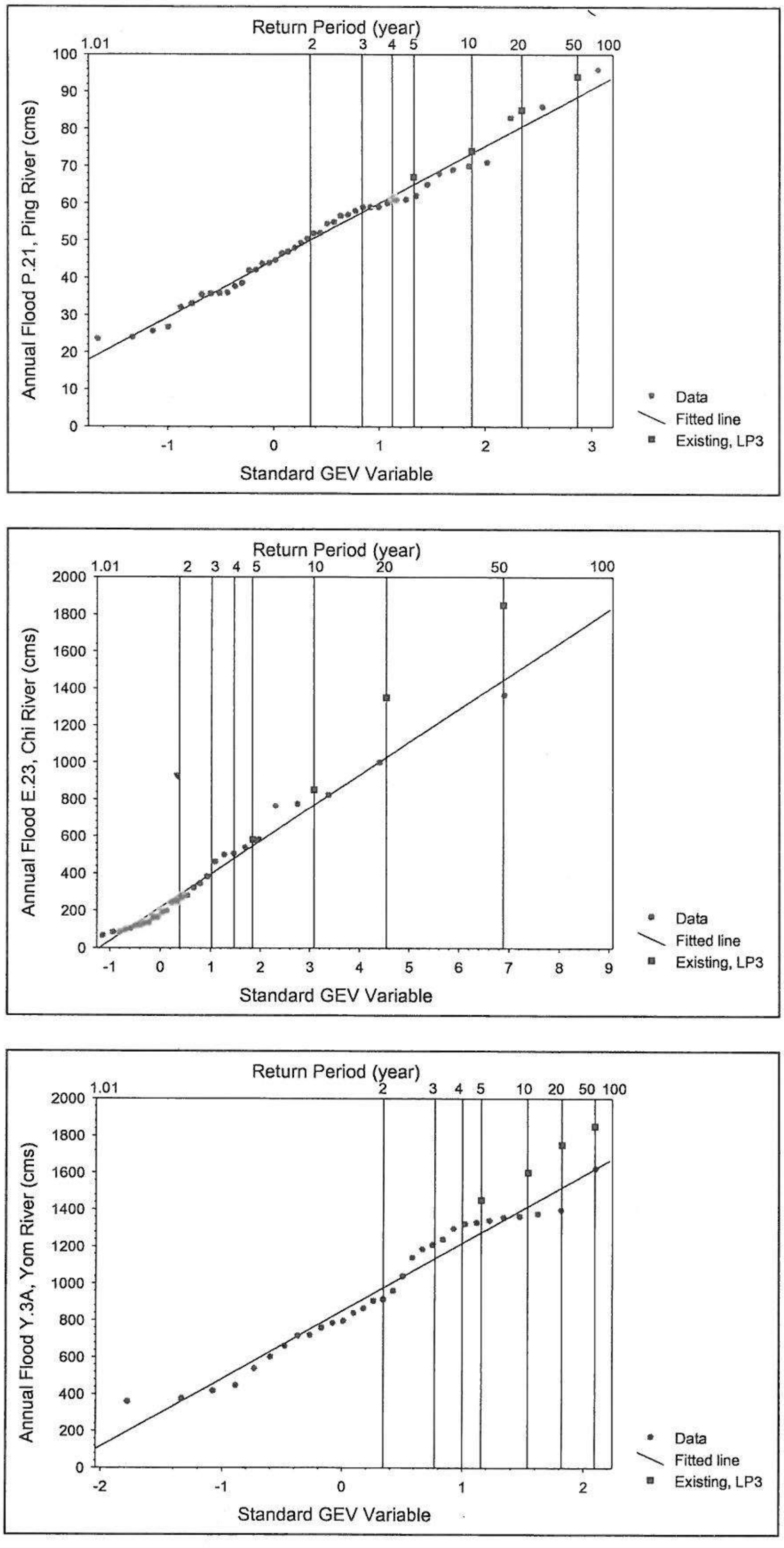


Figure 5 GEV distribution probability plots and existing LP3 model of annual flood records at Station P.21 (Ping River), E.23 (Chi River) and Y.3A (Yom River)