Matematika, 2000, Jilid 16, bil. 1, hlm. 11 30


# Statistical Flood Frequency Analysis Using Short Term Data

**Abdul Aziz Abdul Ghani & Ani Shabri**
Jabatan Matematik, Fakulti Sains
Universiti Teknologi Malaysia
81310 UTM Skudai, Johor, Malaysia

**Abstract**

The most suitable probability distribution for estimating flood risks using short-term data is presented in this paper. A simulation method using autoregressive processes to represent daily flows are used. Several distributions were fitted to the annual peaks of the autoregressive processes. The levels determined using the distributions were compared with the levels from empirical distribution determined using ordered peaks of 1000 autoregressive processes. It was found that the Lognormal III, Pearson Type III and the Weibull gave good risks estimates of flood of high return period.

**Keywords**

Autoregressive process, return period, empirical distribution.

**Abstrak** *Beberapa taburan kebarangkalian yang sesuai untuk menganggarkan kejadian banjir menggunakan data terhad dijelaskan dalam kertas ini. Kaedah simulasi berdasarkan proses autoregresi digunakan untuk mengambarkan aliran harian. Beberapa taburan dipadankan kepada puncak maksimum bagi proses autoregresi. Paras tertentu ditetapkan menggunakan taburan telah dibandingkan dengan paras menggunakan taburan empirik menggunakan puncak tertib bagi 1000 proses autoregresi. Hasil analisis didapati bahawa taburan Lognormal III, Pearson Jenis III dan Weibull memberikan anggaran terbaik untuk menganggar banjir bagi tempoh ulangan yang tinggi.*

**Katakunci**

*Proses Autoregresi, tempoh ulangan, taburan empirik.*

## 1 Introduction

Floods are extreme events, which may cause large amount of damages. Therefore defence mechanism should be designed and built to prevent floods from causing disaster such as flooding near rivers, which would destroy crops, and causing death to grazing cattle, sheep and other animals.

The relationship between flood magnitude and its return period is of great importance. The design of this defence mechanism greatly relies on the high floods exceeding certain levels. In developing countries like Malaysia, data are scarce. Therefore, a simple probability distribution that can give good estimates of flood of high return period need to be estimated from these data.

In order to find the best-fitted distribution several distributions that are widely used for frequency analysis were studied. They are Extreme Value Type I (EV1), Lognormal III (LNIII), Pearson Type III (PIII), Log Pearson Type III (LPIII), General Extreme Value (GEV) and the Weibull distribution.

## 2 Extreme Value Type I Distribution

*The Gumbel or the* Extreme Value Type I (EV1) distribution was introduced by Gumbel in 1941 [2]. It is widely used for frequency analysis in hydrology, meteorology, storms and droughts [6].

The probability density function p.d.f. of an EV1 distribution can be expressed as

$$f(x) = \frac{1}{\alpha} exp\{-\frac{x-\mu}{\alpha} - exp(-\frac{x-\mu}{\sigma})\}, \text{ for } \alpha > 0, -\infty < \mu < \infty, \tag{1}$$

where $\alpha$ is the location parameter and $\mu$ is the scale parameter.

The relationship between the event $x_p$ and its quartile value $p$ is given by

$$x_p = \mu - \alpha ln(-ln(p)). \tag{2}$$

The parameters of this distribution was estimated using the method of probability weighted moments [6].

## 3 Lognormal III Distribution

The Lognormal III distribution was first introduced by Hazen in 1914. It was used by Hoshi [3] for flood frequency analysis.

The p.d.f. of the Lognormal III distribution is given by

$$f(x) = \frac{1}{\sigma(2\pi)^{1/2}(x-\gamma)} exp(-\frac{(ln(x-\gamma)-\mu)^2}{2\sigma}) \tag{3}$$

where $\mu$ and $\sigma$ is the mean and standard deviation respectively and $\gamma$ is the location parameter.

The quartile of the Lognormall III distribution is given by

$$x_p = \gamma + \exp(\mu + \sigma z_p). \tag{4}$$

where $z_p$ is from the standard normal distribution. *The* parameters *of these distributions* were estimated using the method of ordinary moments [3].

# 4 Pearson Type III Distribution

The Pearson Type III distribution was first introduced by Karl-Pearson in 1924. This distribution is also known as the three-parameter Gamma distribution. Lall and Beard [5] found that this distribution could be used for annual peak distribution.

The p.d.f. of this distribution is given by

$$f(x) = |\beta|[\beta(x-\xi)]^{\alpha-1}\frac{exp[-\beta(x-\xi)]}{\Gamma(\alpha)}, \text{ for } \xi < x < \infty, \alpha > 0, \tag{5}$$

where $\alpha, \beta$ and $\gamma$ are the scale, shape and the location parameters respectively.

The quartile of the Pearson Type III distribution is given by

$$x_p = \mu + \sigma K_p[g(o)]. \tag{6}$$

where $\mu$ and $\sigma$ are the mean and standard deviation. $K_p$ and $g(o)$ is the frequency factor and skewness parameter respectively. The parameters of this distribution were estimated using the method of ordinary moments [5].

# 5 Log-Pearson Type III Distribution

The Log-Pearson Type III distribution was suggested by Water Resources Council USA (W.R.C) in 1967. This distribution was used for flood frequency analysis in USA and Australia [7].

The p.d.f. of this distribution is given by

$$f(x) = |\beta|[\beta(\ln|x| - \xi)]^{\alpha-1}\frac{\exp[-\beta(ln|x| - \xi)]}{x\Gamma(\alpha)} \tag{7}$$

where $\alpha$, $\beta$ and $\gamma$ are the scale, shape and the location parameters respectively.

The quartile of the Log-Pearson Type III distribution is given by

$$x_p = exp(\mu + \sigma K_p g(o)). \tag{8}$$

where $\mu$ and $\sigma$ are the mean and standard deviation. $K_p$ and $g(o)$ is the frequency factor and skewness parameter. The parameters of this distribution were estimated using the method of mixed moments [7].

# 6 Generalized Extreme Value Distribution

The Generalized Extreme Value Distribution was introduced by Jenkinson in 1955. This distribution is the combination of three distributions namely

i. Extreme Value Type I;

ii. Extreme Value Type II ;

iii. Extreme Value Type III or the Weibull distribution.

This distribution is widely used for flood frequency analysis and analysis of peaks of sea level.

The p.d.f. of this distribution is given by [4]

$$f(x) = \frac{exp(-y - exp(-y))}{\alpha(1-t)}, \tag{9}$$

where

$$y = \frac{-\ln(1-t)}{\delta},$$

and

$$t = \frac{\delta(x-\xi)}{\alpha},$$

where $\alpha$, $\delta$ and $\xi$ are the scale, shape and the location parameters respectively.

The quartile of the Generalized Extreme Value distribution is given by

$$x_p = \xi + \frac{\alpha}{\delta}(1 - (-\ln(p))^{\delta}). \tag{10}$$

The parameters of this distribution were estimated using the method of L-moments [4].

# 7 The Weibull Distribution

The Weibull distribution is widely used in life expectancy of industrial components in industry. It is also widely used in hydrology for drought analysis and rainfall [1].

The p.d.f. of the Weibull distribution is given by

$$f(x) = \frac{\delta}{\alpha^{\delta}}(x-\xi)^{\delta-1}\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{\delta}\right\}, \text{ for } \xi < x < \infty,\ \delta > 0,\ \alpha > 0, \tag{11}$$

where $\alpha$, $\delta$ and $\xi$ are the scale, shape and the location parameters respectively.

The quartile of the Weibull distribution is given by

$$x_p = \xi - \alpha(\ln(1-p))^{1/\delta}. \tag{12}$$

The parameters of this distribution were estimated using the method of probability ordinary moments[1].

# 8 Simulation Procedure

A simulation procedure using NAG routine at UTM was adopted. 1000 Autoregressive AR (1) processes with white noises and $\alpha = 0.3$ of the length 61 to represent the flood season during November and December was simulated. The highest peak of each of the 1000 series was noted. Then peaks were fitted with EV1, LNIII, PIII, LPIII, GEV and Weibull distributions. The process was repeated for $\alpha = 0.5, 0.7$ and 0.9 respectively. NAG routine G05FDF were used to generate the white noise and G05CCF were used so that different set of varieties is generated at different times. The fitted distributions are given in Figure 1, Figure 2, Figure 3 and Figure 4.

The 1000 AR(1) processes with $\alpha = 0.3$ of length 61 were again simulated. Their highest peaks were noted and arranged in ascending order of magnitudes. Their $90^{th}$ quartile was calculated. The EV1, LNIII, PIII, LPIII, GEV and Weibull distributions were fitted to the peaks of the 6 AR(1) processes. The $90^{th}$ to the $99^{th}$ quartiles were estimated. These estimates were compared with those of the empirical distribution. The processes were repeated using $8, 10, 12, 14, 16, 20, 25, 30, 35$ and 40 AR(1) processes with $\alpha = 0.5, 0.7$ and 0.9 respectively. The plot of the sample quartile estimated using the respective distributions and the Empirical Quartile are given in Figure 5a to Figure 8b.

# 9 Analysis of Real data

Riverflow data from 5 stations in Johore were analyzed. In the analysis all the six distributions were fitted to the annual peaks of each river. The $10^{th}$ to the $100^{th}$ years flood were estimated for each river. The results are given in Table 2a and 2b.

# 10 Conclusion

From the figure 5a to figure 8b it is shown that as the sample sizes increases the sample quartiles approach the empirical quartiles. Here it was found that the LNIII, PIII and the Weibull distributions gave reasonable quartile estimates of the $90^{th}$ to $99^{th}$ quartiles, while the EV1, LPIII and GEV were seen to underestimate.

From the results it was found that the LNIII, PIII and the Weibull distributions could be used to estimate floods of high return period using limited data.

In the flood frequency analysis of the rivers in Johor it is shown that there is no significant difference between floods of high return period using *each of the suggested distribution*. Generally other distributions would give higher frequency estimates.

# References

[ **1** ] A. C. Cohen *Parameter Estimation in Reliability and Life Span Model*, Marcell Dekker, **Inc. USA** (1988).

[ **2** ] E. J. Gumbel, *Statistics of Extremes.* , Colombia Univ. Press (1958).

[ **3** ] K. Hoshi, J.R. Stedinger & S. J. Burges, *Estimation of Log-Normal Quantiles : Monte Carlo Results And First Order Approximations,*, Journal of Hydrology, **Vol. 71,** (1984), 1 30.

[ **4** ] J. R. M. Hosking & J. R. Wallis,*Estiamtion of the Generalized Extreme Value Distribution by Method of Probability Weighted Moments*, Technometrics, **Vol. 27, No. 3** (1985), 251 261

[ **5** ] U. Lall & Beard, *Estimation of Pearson Type 3 Moments*, Water Resources Research, **Vol. 18** (1982), 1563 1569.

[ **6** ] J. M. Landwehr, N. C. Matalas & J. R. Wallis, *Probability Weighted Moments Compared With Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles*, Water Resources Research. **Vol. 15** (1979), 1055 1064.

[ **7** ] H. N. Phien & M. A. Hira, *Log-Pearson Type 3 Distribution; Parameter Estimation*, Journal of Hydrology. **Vol. 64** ( 1983.), 25 37.

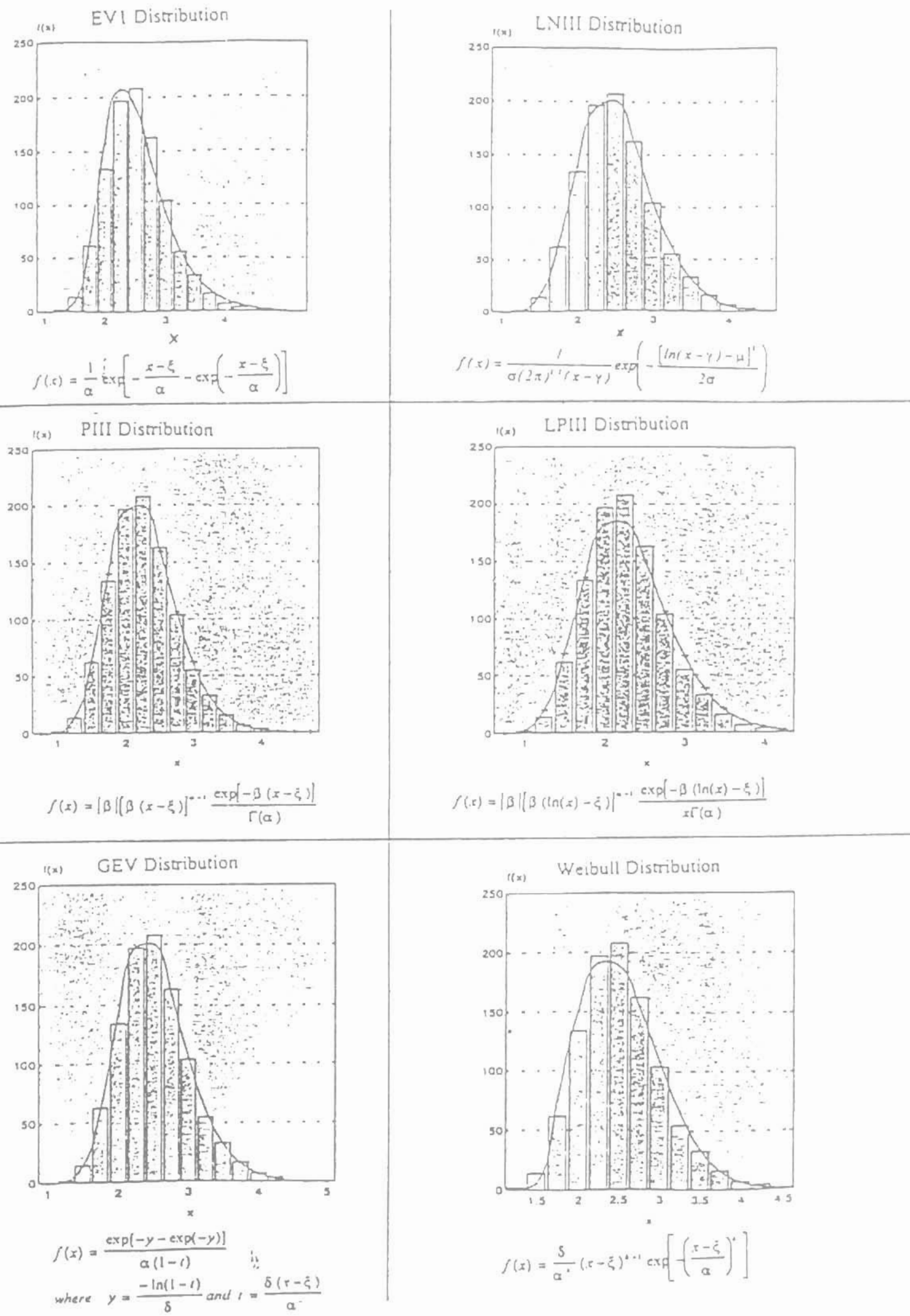



Figure 1 : Histogram And The EV1, LNIII, PIII, LPIII, GEV and Weibull Distributions of Maxima Peaks of AR(1) Process, $\alpha = 0.3$

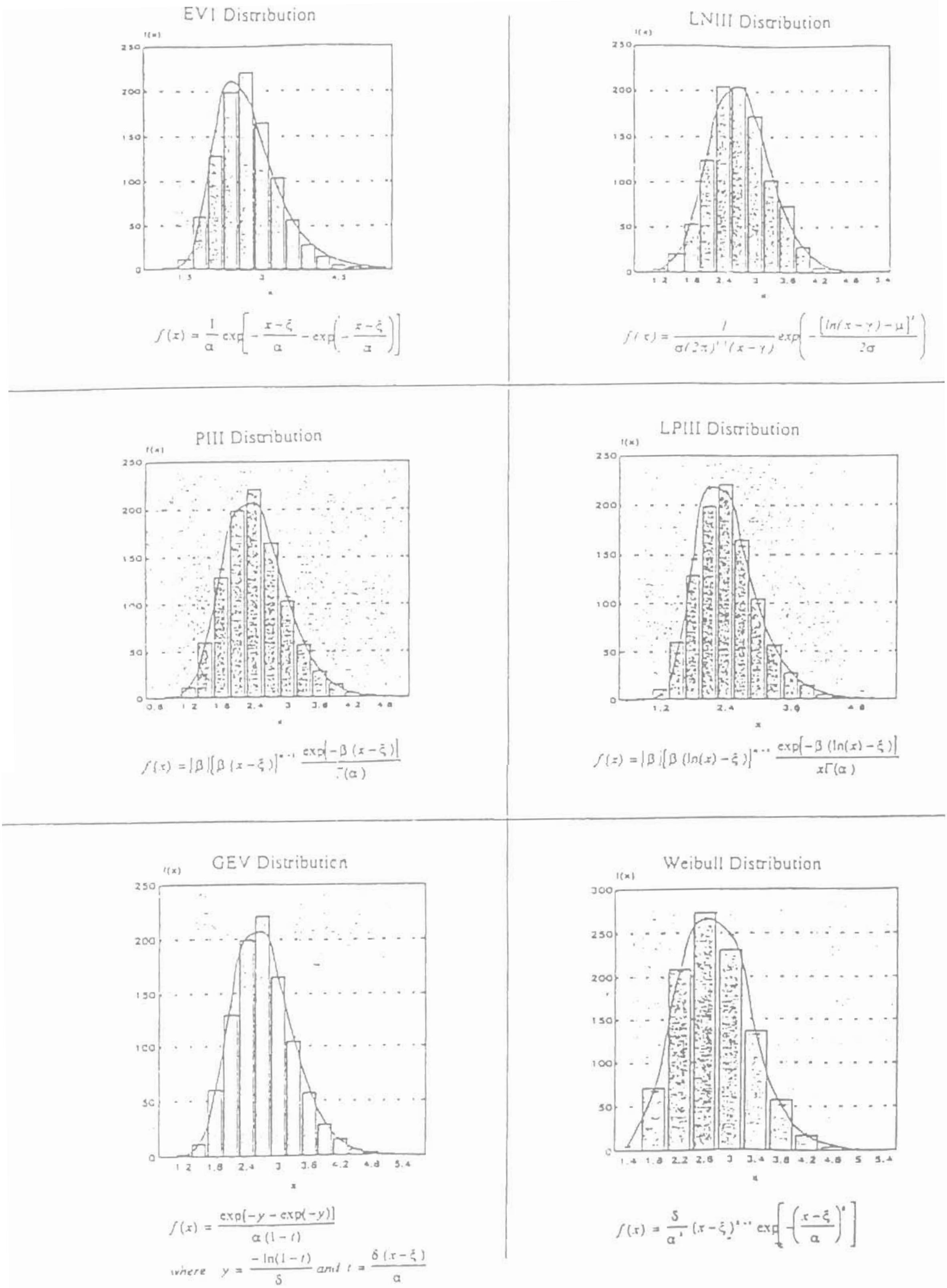



Figure 2 : Histogram And The EV1, LNIII, PIII, LPIII, GEV and Weibull Distributions of Maxima Peaks of AR(1) Process, $\alpha = 0.5$

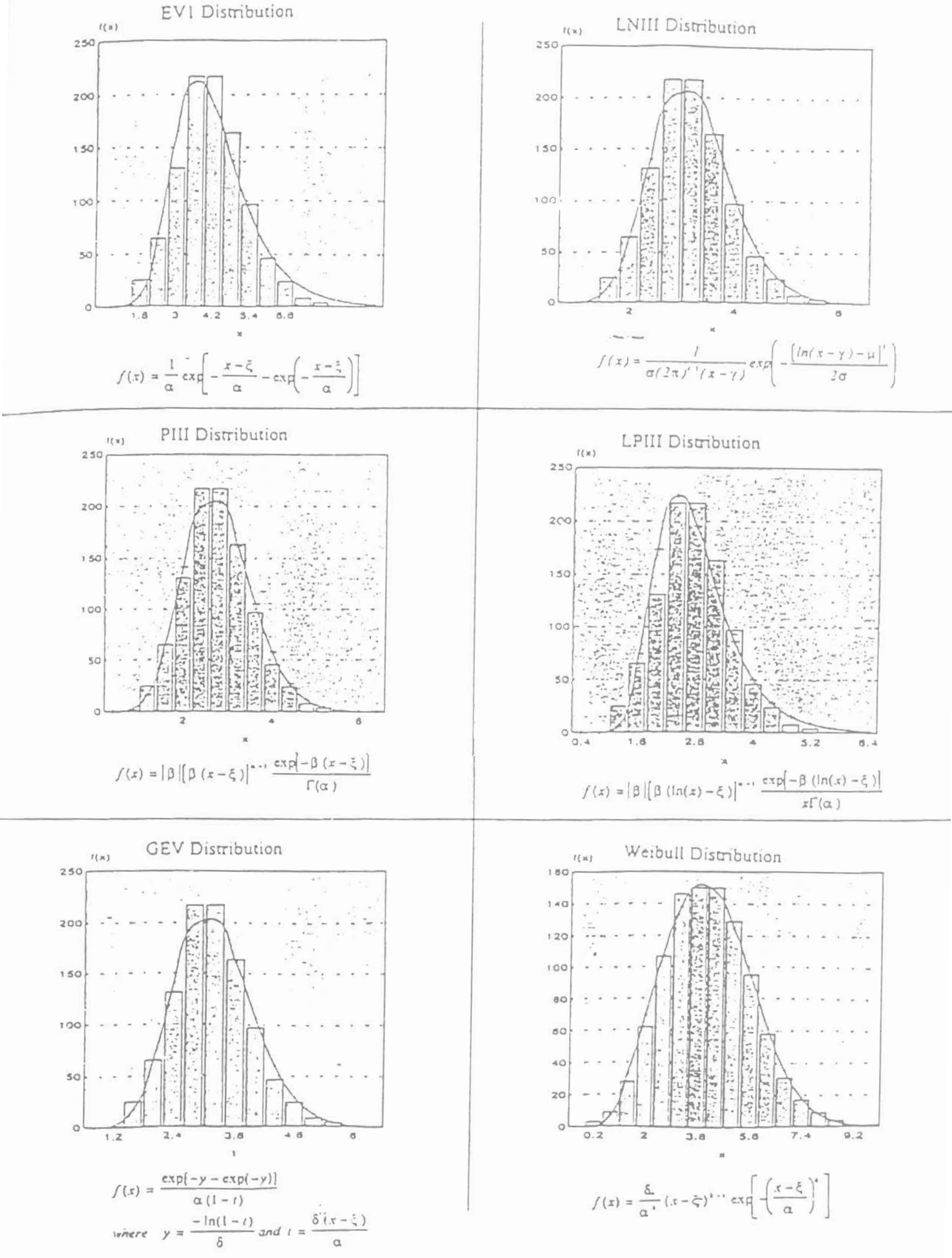



Figure 3 : Histogram And The EV1, LNIII, PIII, LPIII, GEV and Weibull Distributions of Maxima Peaks of AR(1) Process, $\alpha = 0.7$

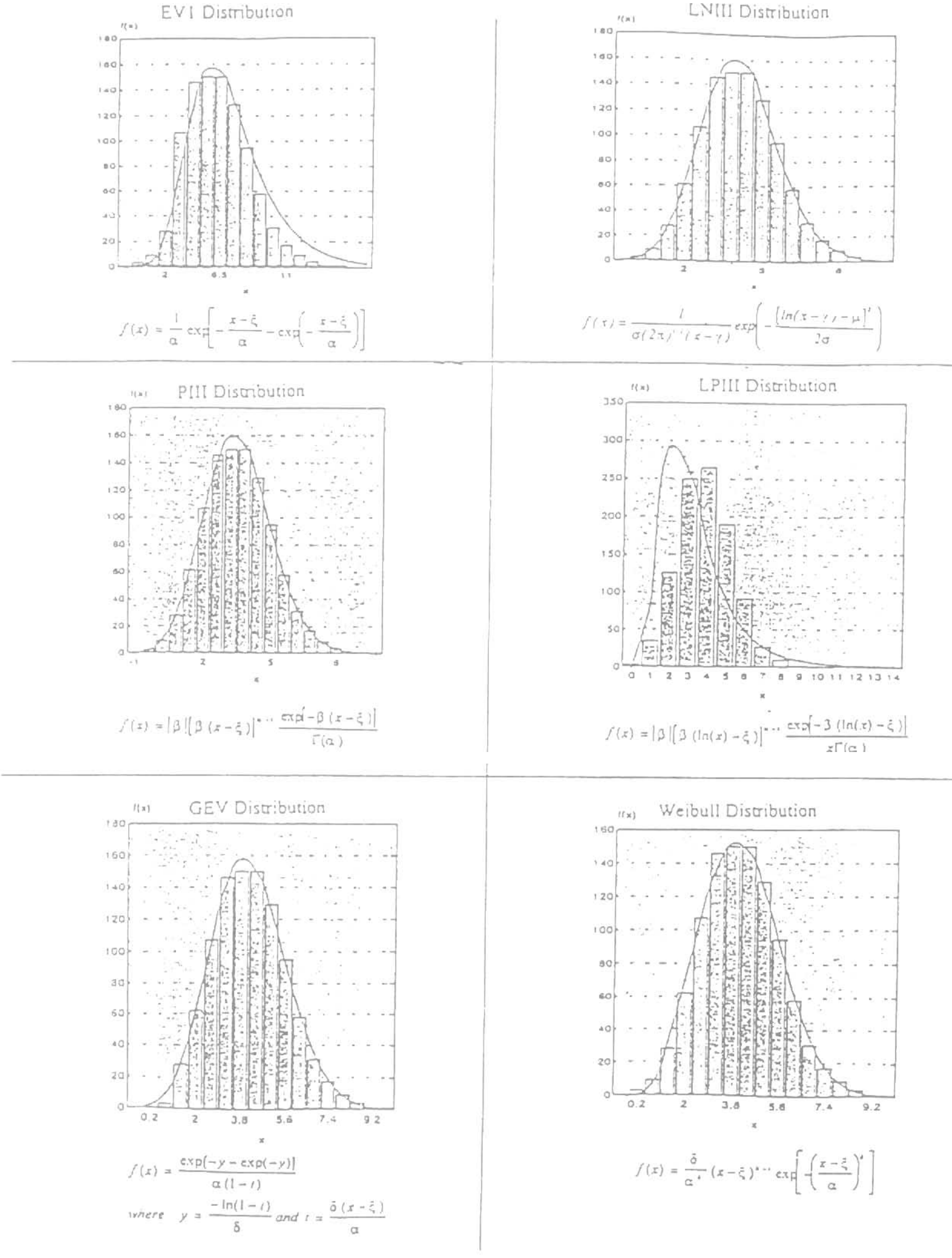
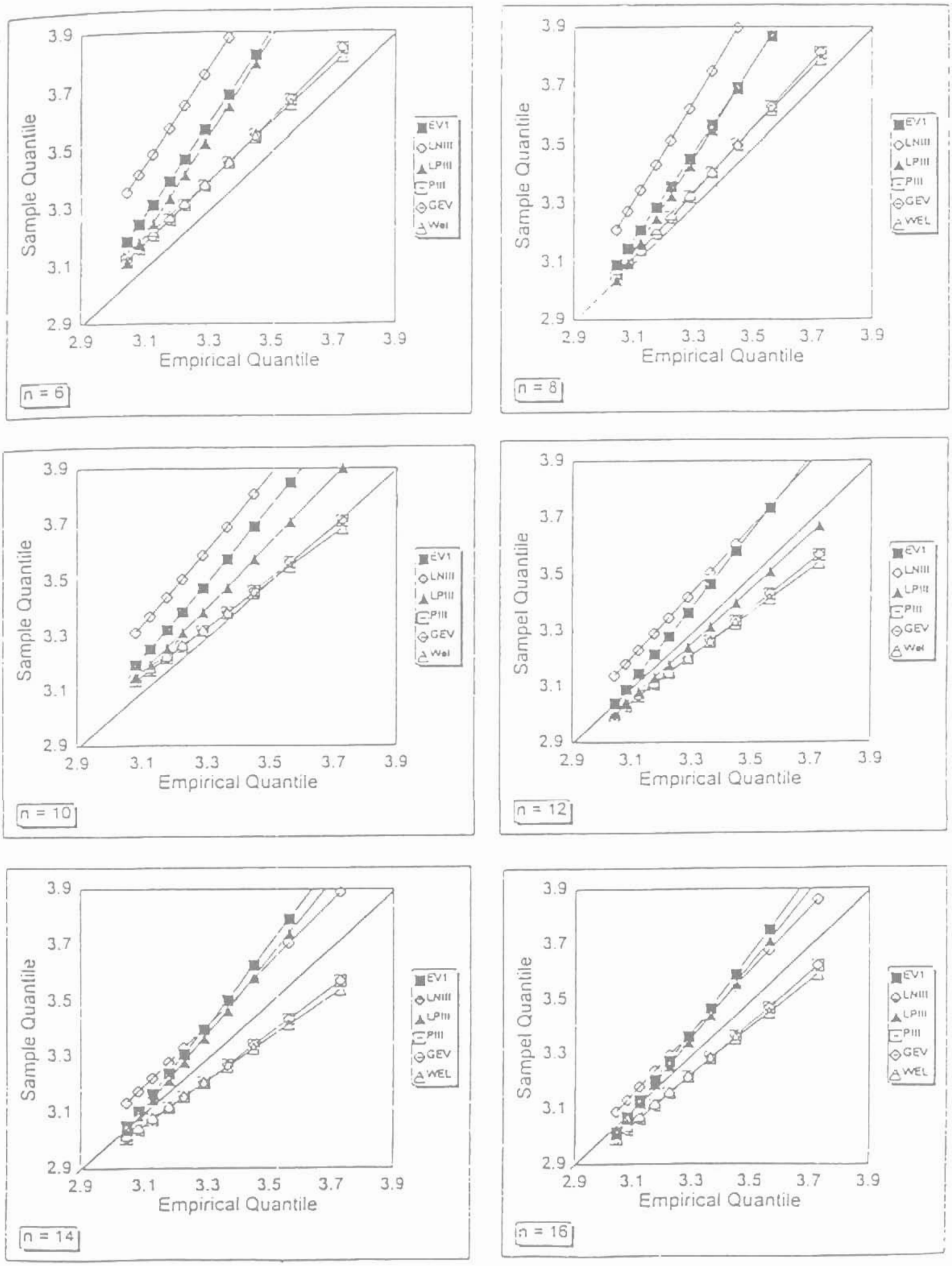



Figure 4 : Histogram And The EV1, LNIII, PIII, LPIII, GEV and Weibull Distributions of Maxima Peaks of AR(1) Process, $\alpha = 0.9$




(Continue)

Figure 5a and 5b : Graph of Sample Quantile Against Empirical Quantile for EV1, LNIII, PIII, LPIII, GEV and Weibull. $\alpha = 0.3$

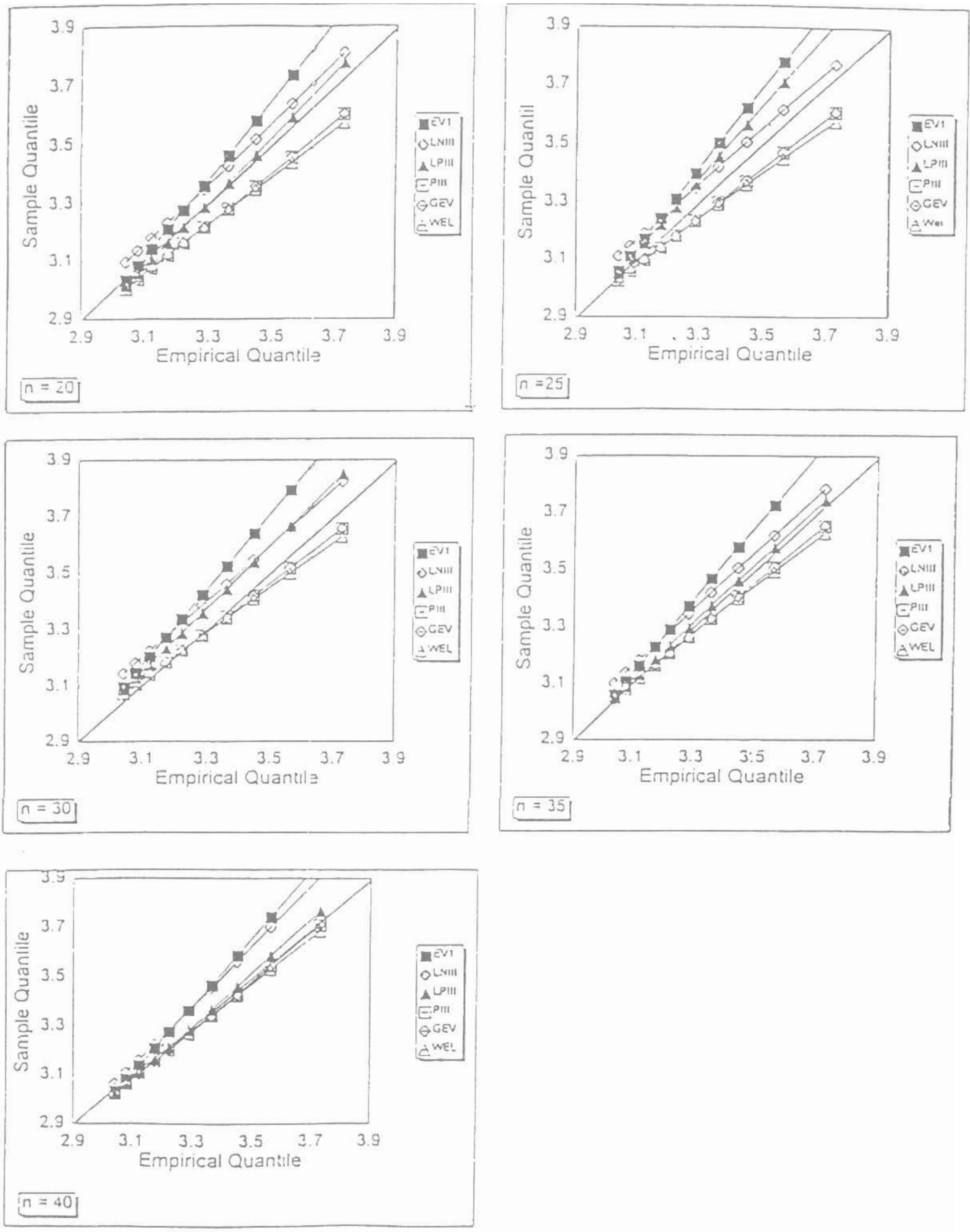
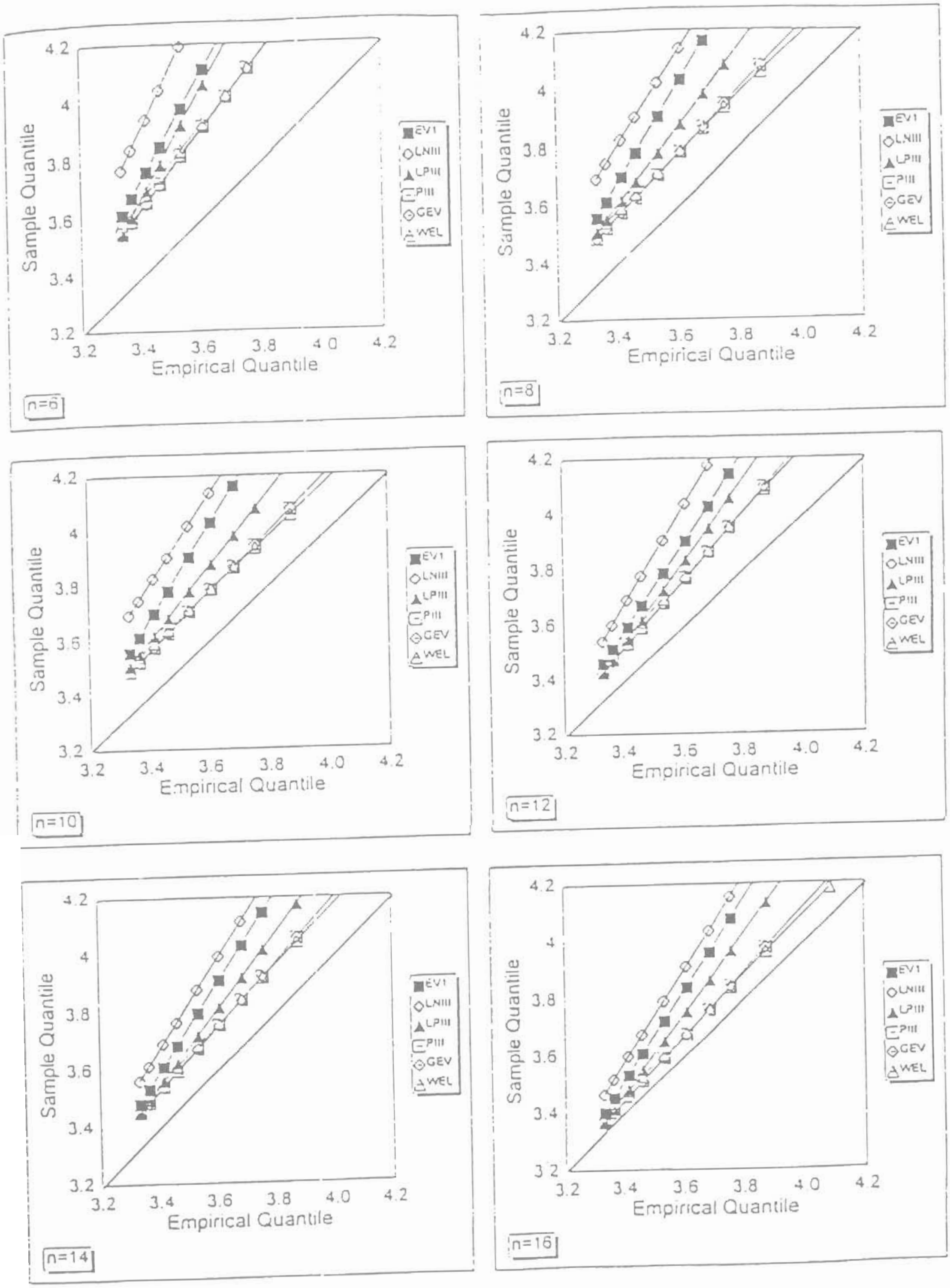

3.9
3.7
3.5
3.3
3.1
2.9
Sample Quantile
2.9 3.1 3.3 3.5 3.7 3.9
Empirical Quantile
EV1
LNIII
LPIII
PIII
GEV
WEL
n = 20
Sample Quantil
Empirical Quantile
Wei
n =25
Sample Quantile
Empirical Quantile
n = 30
Sample Quantile
Empirical Quantile
n = 35
Sample Quantile
Empirical Quantile
n = 40

Figure 5 b




(continue)

Figure 6a and 6b : Graph of Sample Quantile Against Empirical Quantile for EV1, LNIII, PIII, LPIII, GEV and Weibull. $\alpha = 0.5$

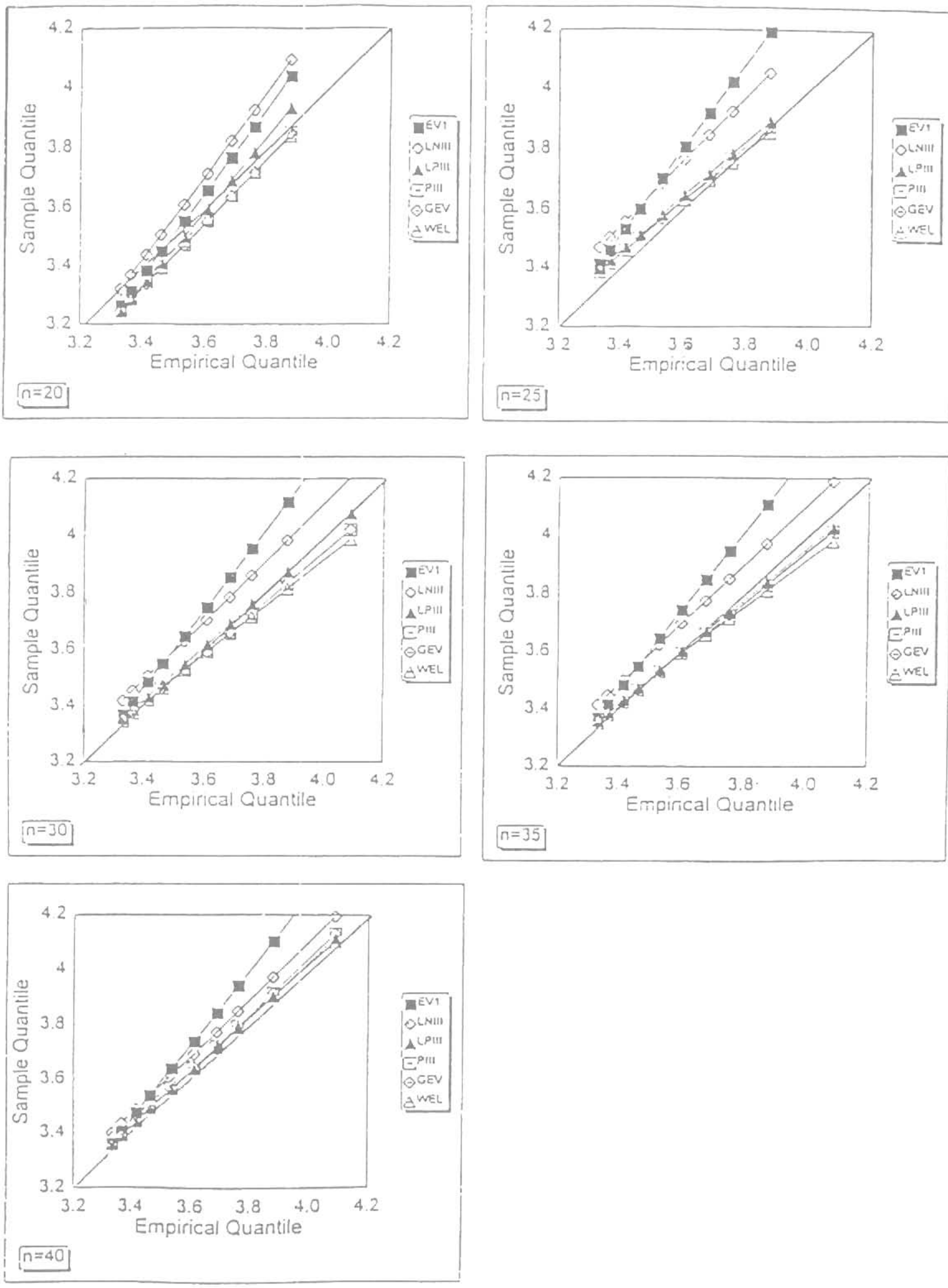
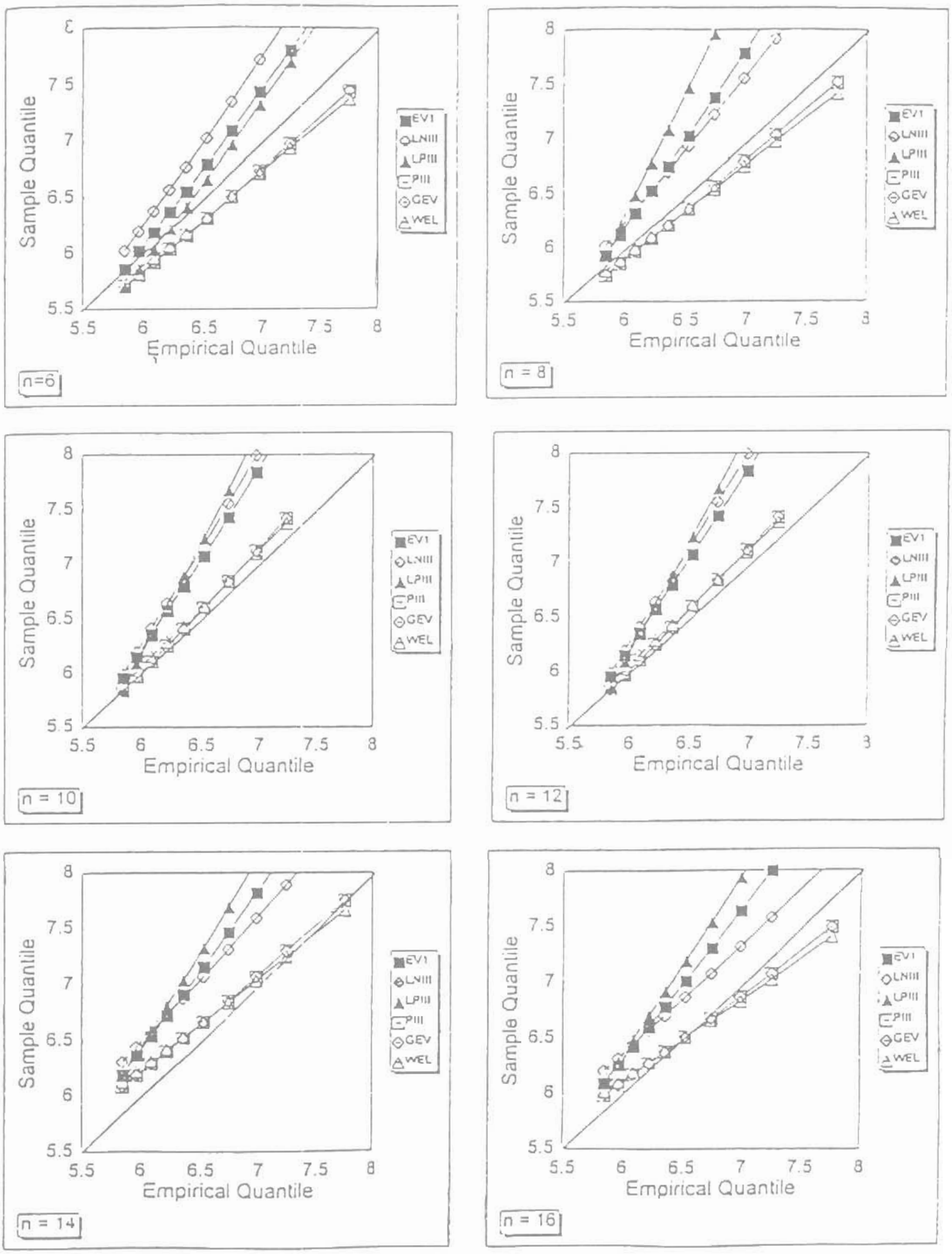

Sample Quantile
Empirical Quantile
EV1
LNIII
LPIII
PIII
GEV
WEL
n=20
n=25
n=30
n=35
n=40

Figure 6 b




(continue)

Figure 7a and 7b : Graph of Sample Quantile Against Empirical Quantile for EV1, LNIII, PIII, LPIII, GEV and Weibull. $\alpha = 0.7$

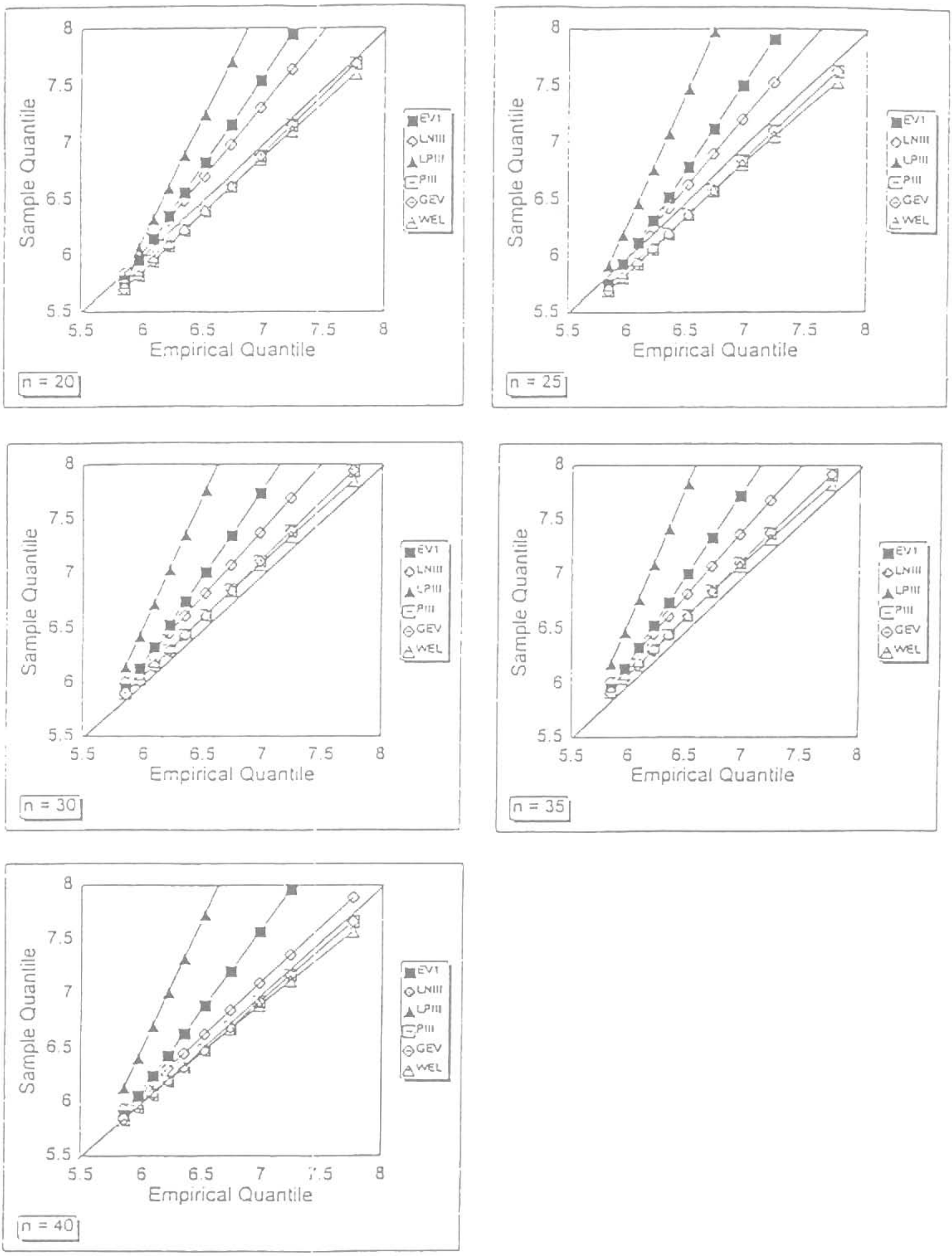
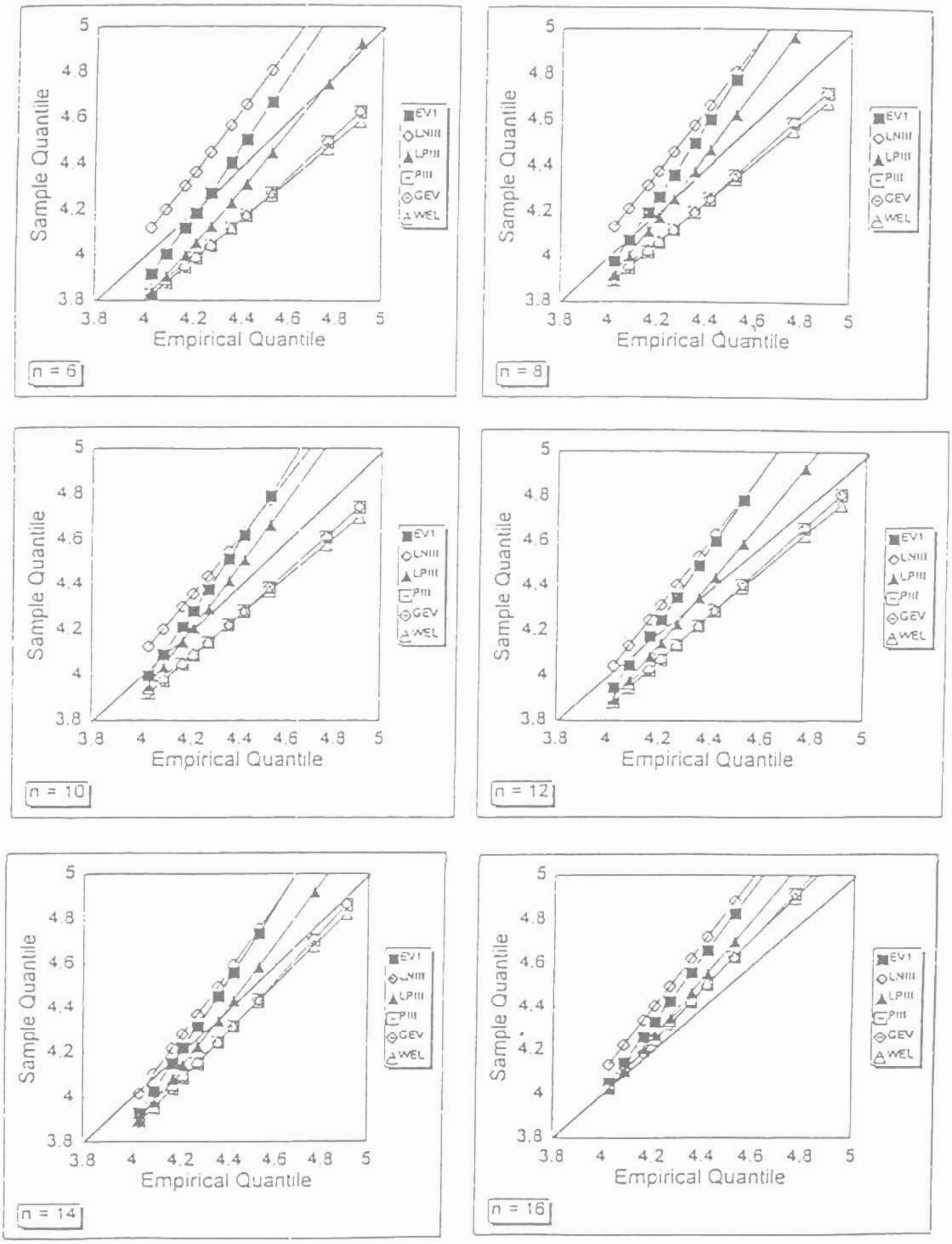

Sample Quantile
Empirical Quantile
5.5
6
6.5
7
7.5
8
EV1
LNIII
LPIII
PIII
GEV
WEL
n = 20
n = 25
n = 30
n = 35
n = 40

Figure 7 b




(Continue)

Figure 8a and 8b : Graph of Sample Quantile Against Empirical Quantile for EV1, LNIII, PIII, LPIII, GEV and Weibull. $\alpha = 0.9$

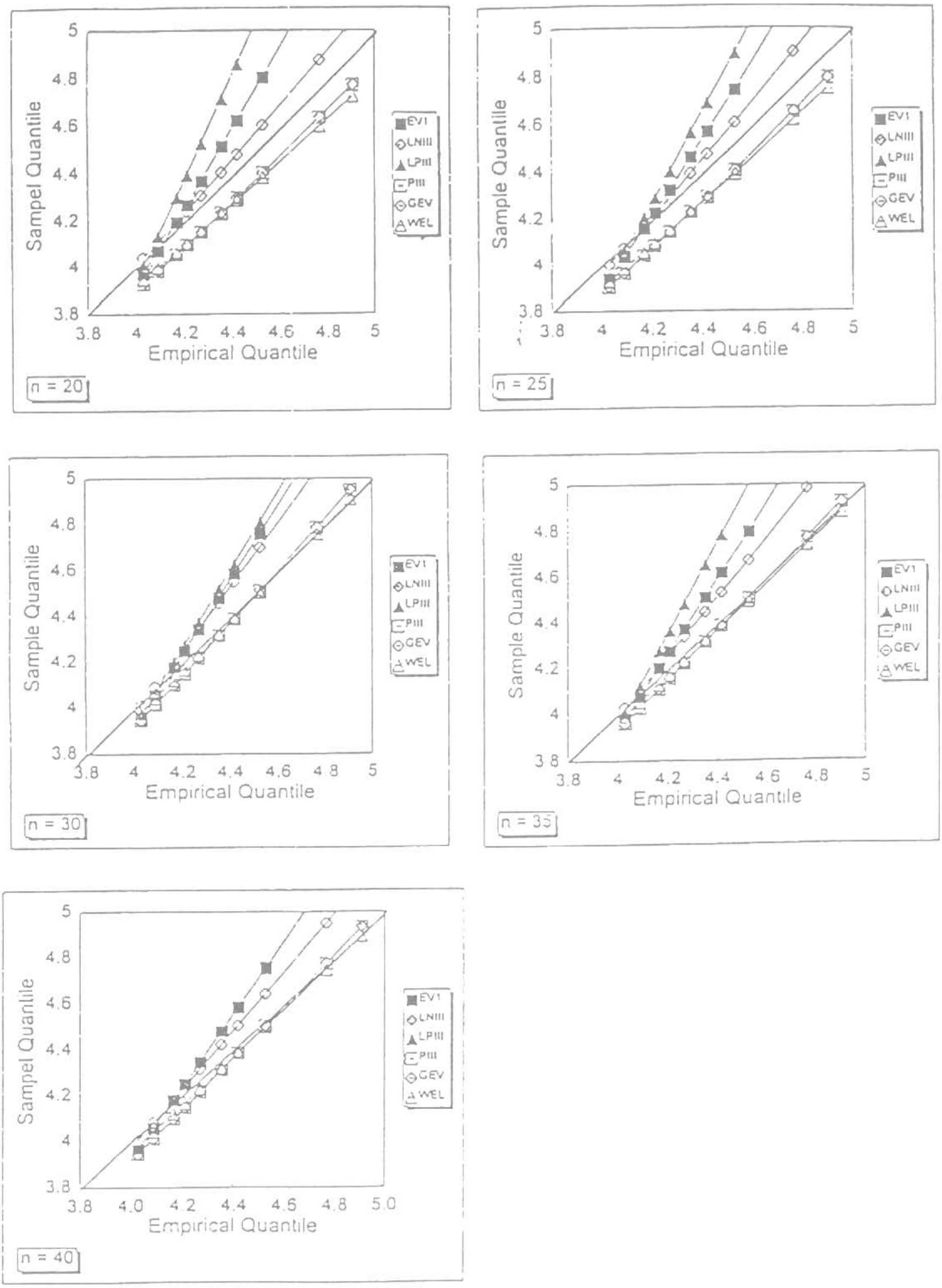

Sampel Quantile
Empirical Quantile
EV1
LNIII
LPIII
PIII
GEV
WEL
n = 20
Sample Quantile
Empirical Quantile
n = 25
Sample Quantile
Empirical Quantile
n = 30
Sample Quantile
Empirical Quantile
n = 35
Sampel Quantile
Empirical Quantile
n = 40

Figure 8 b

Table 1 : List of Stations Name

| Station Name | Station Number | Area | Recod Length (Years) |
|---|---|---|---|
| Sayong River | 1836402 | Johor Tenggara | 11 |
| Bekok River | 2130422 | Yong Peng | 18 |
| Linggui River | 1836401 | Tanah Jengli | 10 |
| Johor River | 1737451 | Rantau Panjang | 23 |
| Kahang River | 2235401 | Kluang | 11 |

Table 2 (a) : Quantilies values of LNIII, PIII, Weibull ,LPIII, EV1 and GEV Distributions
( T from 10 to 100 years)

a). Sayung River (Station Number. : 1836402)

| Station Name | T<br>p | 10<br>0.90 | 12.5<br>0.92 | 20<br>0.95 | 25<br>0.96 | 50<br>0.98 | 100<br>0.99 |
|---|---|---|---|---|---|---|---|
| Sayung River | LNIII | 207.933 | 221.053 | 248.472 | 261.432 | 301.593 | 341.660 |
| | PIII | 209.879 | 223.299 | 251.081 | 264.081 | 303.749 | 342.396 |
| | Weibull | 212.383 | 225.778 | 253.041 | 265.591 | 303.237 | 339.192 |
| | LPIII | 211.533 | 227.734 | 263.019 | 280.356 | 336.700 | 396.754 |
| | EV1 | 208.780 | 221.614 | 248.265 | 260.790 | 299.374 | 337.673 |
| | GEV | 208.291 | 223.581 | 256.872 | 273.265 | 326.956 | 385.375 |

b). Bekok River (Station Number : 2130422)

| Station Name | T<br>p | 10<br>0.90 | 12.5<br>0.92 | 20<br>0.95 | 25<br>0.96 | 50<br>0.98 | 100<br>0.99 |
|---|---|---|---|---|---|---|---|
| Bekok River | LNIII | 61.005 | 63.148 | 67.332 | 69.186 | 74.504 | 79.270 |
| | PIII | 61.005 | 63.149 | 67.332 | 69.186 | 74.503 | 79.267 |
| | Weibull | 62.756 | 65.107 | 69.663 | 71.670 | 77.397 | 82.520 |
| | LPIII | 64.216 | 68.055 | 76.102 | 80.089 | 92.366 | 104.888 |
| | EV1 | 61.300 | 64.360 | 70.710 | 73.690 | 82.880 | 92.000 |
| | GEV | 62.316 | 64.502 | 68.647 | 70.425 | 75.299 | 79.347 |

c). Linggui River (Station Number : 1836401)

| Station Name | T<br>p | 10<br>0.90 | 12.5<br>0.92 | 20<br>0.95 | 25<br>0.96 | 50<br>0.98 | 100<br>0.99 |
|---|---|---|---|---|---|---|---|
| Linggui River | LNIII | 43.151 | 45.146 | 49.287 | 51.232 | 57.210 | 63.110 |
| | PIII | 43.409 | 45.435 | 49.606 | 51.546 | 57.436 | 63.130 |
| | Weibull | 43.819 | 45.834 | 49.905 | 51.787 | 57.313 | 62.559 |
| | LPIII | 42.870 | 45.857 | 52.748 | 56.332 | 68.905 | 83.932 |
| | EV1 | 42.915 | 44.885 | 48.977 | 50.900 | 56.824 | 62.703 |
| | GEV | 42.531 | 45.198 | 51.248 | 54.349 | 65.065 | 77.686 |

Table 2(b)

**(d). Johor River (Station Number : 1836401)**

| Station Name | T<br>P | 10<br>0.90 | 12.5<br>0.92 | 20<br>0.95 | 25<br>0.96 | 50<br>0.98 | 100<br>0.99 |
|---|---|---|---|---|---|---|---|
| Johor River | LNIII | 405.246 | 431.469 | 485.729 | 511.136 | 588.981 | 665.438 |
| | PIII | 408.393 | 434.958 | 489.493 | 514.812 | 591.449 | 665.287 |
| | Weibull | 413.927 | 440.282 | 493.380 | 517.602 | 589.507 | 657.253 |
| | LPIII | 413.125 | 447.117 | 520.358 | 555.928 | 669.561 | 787.397 |
| | EV1 | 405.730 | 432.547 | 488.238 | 514.411 | 595.036 | 675.066 |
| | GEV | 404.592 | 436.725 | 506.804 | 541.369 | 654.829 | 778.681 |

**(e). Kahang River (Station Number : 2235401)**

| Station Name | T<br>P | 10<br>0.90 | 12.5<br>0.92 | 20<br>0.95 | 25<br>0.96 | 50<br>0.98 | 100<br>0.99 |
|---|---|---|---|---|---|---|---|
| Kahang River | LNIII | 346.485 | 359.268 | 384.293 | 395.413 | 427.420 | 456.238 |
| | PIII | 346.503 | 359.283 | 384.300 | 395.415 | 427.401 | 456.190 |
| | Weibull | 348.125 | 360.220 | 383.469 | 393.632 | 422.389 | 447.818 |
| | LPIII | 352.035 | 365.647 | 391.685 | 402.945 | 434.066 | 460.197 |
| | EV1 | 347.773 | 365.625 | 402.698 | 420.121 | 473.792 | 527.068 |
| | GEV | 358.432 | 372.327 | 398.976 | 410.542 | 442.700 | 470.024 |