# Flood Frequency Analysis by the Box-Cox Transformation

**Lee, Soon Hyuk*, Jo, Seong Kab**, Park, Myeong Keun****

**Department of Agricultural Engineering, Chung Buk National University*
***Graduate School, Chung Buk National University*

**Abstract** ☐ This study was conducted to pursue the normalization of frequency distribution by making an approach to the coefficient of skewness to nearly zero through the Box-Cox transformation, to get probable flood flows can be calculated by means of the transformation equation which has been derivated by Box-Cox transformation in the annual maximum series of the applied watersheds. It has been concluded that Box-Cox transfromation is proved to be more efficient than logarithmic, square root and SMEMAX transformation which is based on the trigonometric solution of a right triangle whose three verteces repesent the smallest, median and largest observed values of a population in making the coefficient of skewness nearer to zero. Consequently, it is shown that probable flood flows according to the return period based on Box-Cox transformation are closer to the observed data as compared to other methods including SMEMAX transformation and fitted probability distributions such as the three parameter lognormal and the type 1 extremal distribution for the applied watersheds.

**Keywords** ☐ Skew coefficient, Box-Cox transformation, SMEMAX transformation, Goodness of fit test. Three parameter lognormal and Type 1 extremal distribution, Frequency factor, Flood flows of desired frequency.

## I. INTRODUCTION

Estimation of design flow must be suggested on a preferential basis for the rational design and management of hydraulic strcutures such as dams, spillways and bridges.

Especially, extreme events including the annual maximum series can rarely be normal, but is usually distributed asymmetrically. In many cases, neither the logarithmic nor the square-root transformation will normalize the skewed distribution. In consequence, data from extreme value series form their own distribution. Thus, it is better to find the best fitted distribution for the given data instead of fitting a known distribution to the data.

Alternatively, the given data could be reconstituted by some transformations so that the transformed series follow a particular distribution.

Bethalahmy[2] and Chander[13] et al have suggested using the SMEMAX and the Box-Cox transformation to normalize skewed data, respectively.

Consequently, this study is mainly conducted to get probable flood flows by the normalization of frequency distribution through the Box-Cox transformation and to compare with the results obtained by the best fitted distributions and by the SMEMAX transformation which was presented by Lee[9] et al with annual maximum series for the six watersheds of five major river basins in Korea.[5,6,7,8,9]

## II. BOX-COX TRANSFORMATION AND DATA USED FOR APPLICATION

1. Box-Cox transformation

Box-Cox transformation is that Box and Cox[3] have suggested the following transformation for normality of the skewed distribution.

$$Z_i = \frac{y_i^{\lambda} - 1}{\lambda} \quad \text{in which} \quad \lambda \neq 0$$

$$Z_i = \log y_i \quad \text{in which} \quad \lambda = 0 \qquad (1)$$

where $y_i$ = the variates of a given series i.e. original skewed flow, $Z_i$ = transformed flows, and $\lambda$ = a constant of transformation such that $Z_i$ have zero skew. Transformation equation, (1) hold for $y_i > 0$.

The constant, $\lambda$ may be estimated by the maxinum likelihood method. Alternatively, $\lambda$ may be estimated by a trial and error method such that the coefficient of skewness of the transformed flows is zero. The value of $\lambda$ generally ranges from -1.0 to 1.0. It is very helpful for the estimation of $\lambda$ that an increase or decrease in $\lambda$ follows an increase or decrease in the coefficient of skewness. Trial and error method was used for getting $\lambda$, a constant of transformation with flow chart as shown in Fig. 1 in this study.

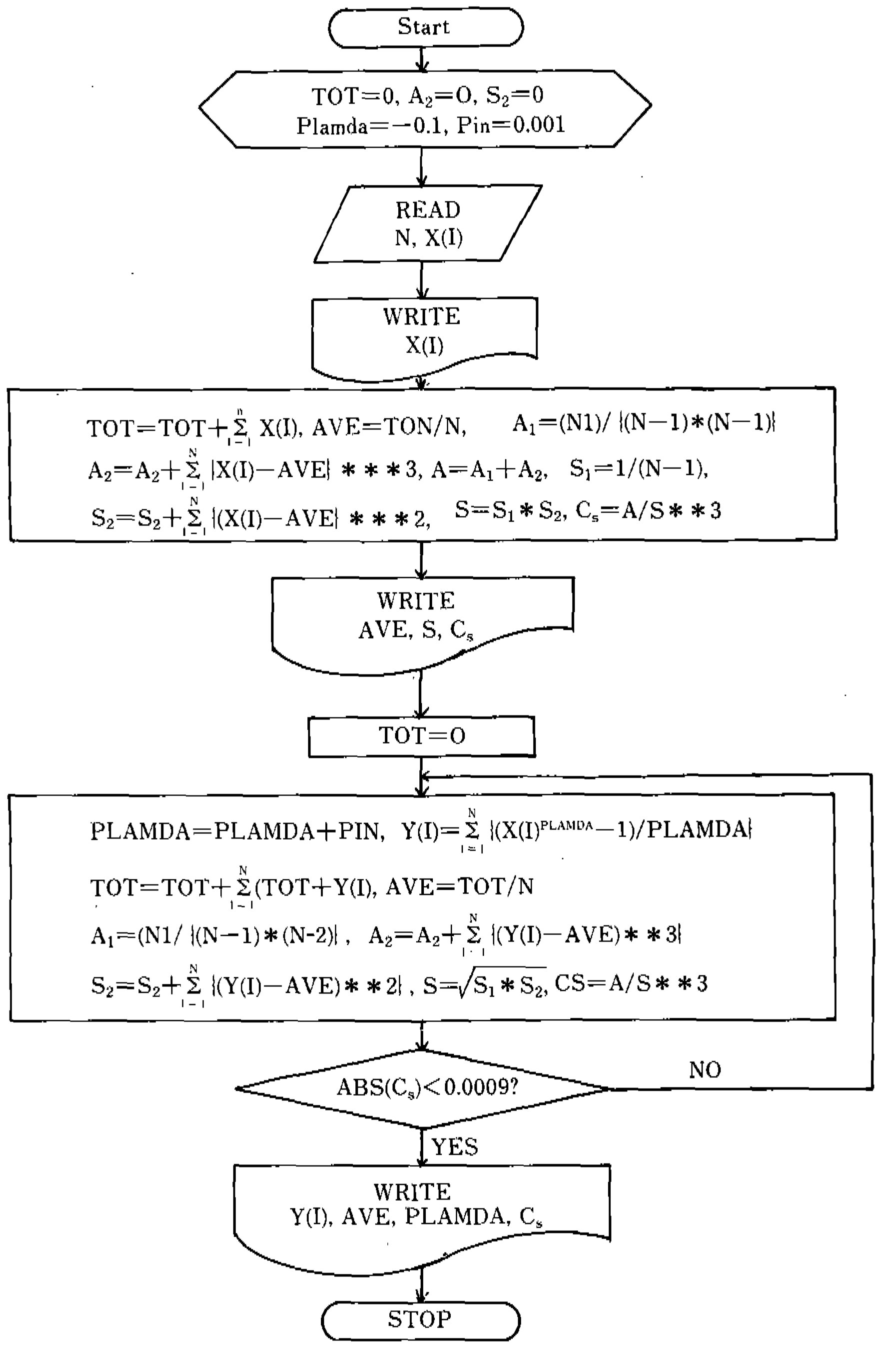


**Fig. 1. Flow chart for the Box-Cox transformation.**

2. Data used for application

Six watersheds selected as research basins are Jeong Sun, Gyu Am, Seog Hwa, Im Ha, Ab Nog and Ma Reug watersheds along Han, Geum, Nag Dong, Seom Jin and Yeong San river system which may be considered as main river systems in Korea, respectively. The used data were the annual maximum series at 6 selected stations above mentioned. Physical characteristics for the research basins are shown as Table 1.[6,7,8]

**Table 1. Gauging stations and watershed physical characteristics**

| River | Station | Area ($km^2$) | Length of Main Stream(km) | Average basin width(km) | Shape factor | Observed duration ($y_{rs}$) |
|---|---|---|---|---|---|---|
| Han River | Jeong Sun | 1709.7 | 106.8 | 16.0 | 0.15 | 24 |
| Geum River | Gyu Am | 8273.0 | 338.0 | 24.5 | 0.07 | 29 |
| | Seog Hwa | 1834.7 | 85.0 | 21.58 | 0.25 | 29 |
| Nag Dong River | Im Ha | 1360.5 | 97.2 | 14.0 | 0.14 | 20 |
| Yeong San River | Ma Reug | 685.0 | 56.0 | 12.23 | 0.22 | 27 |
| Seom Jin River | Ab Nog | 2448.0 | 162.3 | 15.08 | 0.09 | 26 |

## III. RESULTS AND DISCUSSION

1. Analysis for probability distribution functions

a. Probability distributions

Two porbability distributions are used for this analysis.

1) Three parameter lognomal distribution

2) Type 1 extremal distribution

b. Basic statistics

Basic statistics obtained from applied watersheds are shown in Table 2. Those were within the range of 250.39 to 1612.69, 0.334 to 1.496 and from 0.52 to 0.68 for standard deviation, coefficient of skewness and coefficient of variation, respectively.

The efficiency of transformation can be judged by checking whether the coefficient of skewness tend to zero in the transformed series. It can be seen from Table 3 that the Box-Cox transformation is more efficient than the logarithmic, square root and SMEMAX transformation [2,12] in making the coefficient of skewness nearer to zero.

**Table 2. Basic statistics**

| River | Station | Observed Years (N) | Mean (X) | Variance ($S^2$) | Standard deviation (S) | Coefficient of variation ($C_v$) | Cofficient of Skewness ($C_s$) |
|---|---|---|---|---|---|---|---|
| Han River | Jeong Sun | 24 | 500.50 | 79,436.9 | 281.85 | 0.56 | 1.496 |
| Geum River | Gyu Am | 29 | 2388.62 | 2,600,770 | 1612.69 | 0.68 | 1.207 |
| | Seog Hwa | 29 | 1035.73 | 465,811 | 682.50 | 0.66 | 0.541 |
| Nag Dong River | Im Ha | 20 | 587.07 | 115,474 | 339.82 | 0.58 | 0.337 |
| Yeong San River | Ma Reug | 27 | 441.96 | 62,694.5 | 250.39 | 0.57 | 0.882 |
| Seom Jin River | Ab Nog | 26 | 2211.54 | 1,342,270 | 1158.56 | 0.52 | 0.334 |

**Table 3. Transformation effect for the coefficient of skewness**

| River | Station | Transformation | | | | |
|---|---|---|---|---|---|---|
| | | None | Ln | Square Root | SMEMAX | Box-Cox |
| Han River | Jeong Sun | 1.703 | −0.647 | 0.470 | −0.439 | −0.000808 |
| Geum River | Gyu Am | 1.343 | −0.070 | 0.652 | −0.031 | −0.000799 |
| | Seog Hwa | 0.602 | −0.314 | 0.130 | 0.186 | −0.0005319 |
| Nag Dong River | Im Ha | 0.394 | −1.268 | −0.270 | 0.148 | −0.000514 |
| Yeong San River | Ma Reug | 0.992 | −0.809 | 0.024 | −0.134 | −0.000388 |
| Seom Jin River | Ab Nog | 0.398 | −0.605 | −0.103 | −0.135 | −0.000179 |

c. Goodness of fit test for the probability distributions.

$\chi^2$ and Kolmogorov-Smirnov tests were carried out for getting the best fitted probability distribution by applying the three parameter lognormal and the type 1 extremal distribution for the applied watersheds and each of the suitable probability distributions in consequence was appeared as shown in Table 4 and Table 5.[10]

**Table 4. $\chi^2$ and K-S test for the three parameter lognormal distribution**

| River | Station | $\chi^2$ | Test | K-S Test | Test |
|---|---|---|---|---|---|
| Han River | Jeong Sun | 3.982 | 0 | 0.11 | 0 |
| Nag Dong River | Im Ha | 0.469 | 0 | 0.19 | 0 |
| Seom Jin River | Ab Nog | 5.280 | 0 | 0.20 | 0 |

0: Non Significant

**Table 5. $\chi^2$ and K-S test for the type 1 extremal distribution**

| River | Station | $\chi^2$ | Test | K-S Test | Test |
|---|---|---|---|---|---|
| Geum River | Seog Hwa | 8.489 | s | 0.08 | 0 |
| | Gyu Am | 7.241 | 0 | 0.08 | 0 |
| Yeong San River | Ma Reug | 0.965 | 0 | 0.08 | 0 |

o: Non significant s: significant

The three parameter lognormal distribution was confirmed as a suitable distribution at Jeong Sun, Im Ha and Ab Nog watersheds while type 1 extremal distribution was tested as a suitable one at Seog Hwa, Gyu Am and Ma Reug watersheds as shown in Table 4 and Table 5, respectively.

2. Derivation of probable flood flow

a. Three parameter lognormal distribution

1) Parameters

Evaluation of the parameters for the three parameter lognormal distribution was based on the method of moment by using an electronic computer as shown in Table 6.

**Table 6. Parameters for the three parameter lognormal distribution**

| River | Station | $\mu$ | $\sigma$ | $z_1$ | $z_2$ | a | $\gamma_1$ | $\omega$ | $\mu_y$ | $\sigma_y$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Han River | Jeong Sun | 500.5 | 281.85 | 0.5631 | 0.4926 | −72.5 | 1.5947 | 0.4811 | 6.2408 | 0.4661 |
| Nag Dong River | Im Ha | 587.1 | 339.81 | 0.5788 | 0.1208 | −2226.1 | 0.3632 | 0.8344 | 7.9349 | 0.1204 |
| Seom Jin River | Ab Nog | 2207.7 | 1,157.37 | 0.5242 | 0.1244 | −7095.8 | 0.3751 | 0.8299 | 9.1305 | 0.1239 |

$\mu$: Mean, $\sigma$: Standard deviation, $z_1$: Coefficient of variation of the distribution X. $z_2$: Coefficient of variation of the distribution (X–a), a: Lower boundary, $(\mu - \frac{\sigma}{z_2})$, $\gamma_1$: Coefficient of skew of the distribution X, $\omega$: Coefficient of replacement with $\gamma_1$, $\mu_y$: Mean of the logarithm of (X–a), $\sigma_y$: Standard deviation of logarithm of (X–a).

2) Derivation of probable flood flows according to the return period.

General frequency equation of the three parameter lognormal distribution can be estimated from :

$$y_T = \mu_y + t\sigma_y = \ln(X_T - a) \quad (2)$$

in which $y_T$ is calculated by the mean of the logarithm of (X−a), standard deviation of logarithm of (X−a) and frequency factor, t for the normal and the lognormal distribution.

Frequency factors, t and values of general frequency equation, $y_T$ according to the return period are shown as in Table 7 and Table 8, respectively.

Table 9 show formulas for the probable flood flows and probable flood flows according to the return period of the three parameter lognormal distribution for the selected watersheds. The results are plotted on a extremal probability paper as an example of Jeong Sun watershed in the Han river as shown in Fig. 2.

It was found that probable flood flows are generally increased in proportion to the size of watersheds and the return period.

**Table 7. Frequency factors for the three parameter lognormal distribution**

| Return Period($y_{rs}$) | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
|---|---|---|---|---|---|---|---|
| Frequency factor | 0 | 0.8416 | 1.2816 | 1.6449 | 2.0538 | 2.3264 | 2.580 |

**Table 8. $y_T$ values according to the return period ($y_{rs}$)**

| River | Station | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
|---|---|---|---|---|---|---|---|---|
| Han River | Jeong Sun | 6.2408 | 6.6331 | 6.8382 | 7.0075 | 7.1981 | 7.3251 | 7.4433 |
| Nag Dong River | Im Ha | 7.9349 | 8.0362 | 8.0892 | 8.1329 | 8.1822 | 8.2150 | 8.2455 |
| Seom Jin River | Ab Nog | 9.1305 | 9.2348 | 9.2893 | 9.3343 | 9.3850 | 9.4187 | 9.4501 |

**Table 9. Formulas for the probable flood flows and probable flood flows according to the return period for the selected watersheds (3 P.L.N) unit:cms**

| River | Station | Formula($X_T$) | Return Period ($y_{rs}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
| Han River | Jeong Sun | $-72.5 + e^{YT}$ | 441 | 687 | 860 | 1032 | 1264 | 1445 | 1636 |
| Nag Dong River | Im Ha | $-2226.1 + e^{YT}$ | 567 | 865 | 1033 | 1179 | 1351 | 1470 | 1584 |
| Seom Jin River | Ab Nog | $-7095.8 + e^{YT}$ | 2137 | 3152 | 3726 | 4224 | 4813 | 5221 | 5614 |

3 P.L.N : 3 Parameter lognormal

b. Type 1 extremal distribution

1) Parameters

Parameters for the type 1 extremal distribution were calculated by the method of moment using an electronic computer as shown in Table 10. Cumulative probability of the type 1 extremal distribution can be expressed as follows.

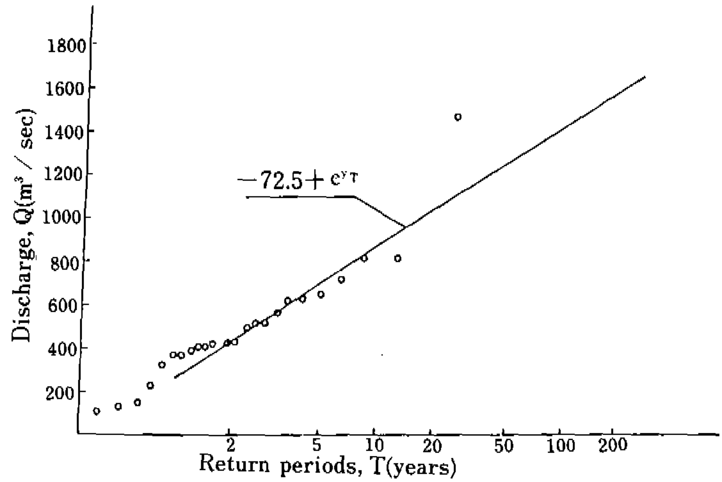


**Fig. 2. Probable flood flows according to the return period at Jeong Sun watershed in the Han river system**

$$P(X) = e^{-e^{-\alpha(X-\beta)}} \tag{3}$$

Where $\alpha$ is a concentration parameter and $\beta$ is a measure of central tendency, substiuting the expression $\alpha(X-\beta) = \bar{y}$ from Equation (3)

$$P(X) = e^{-e^{-\bar{y}}} \tag{4}$$

Consequently, return period, T becomes

$$T = \frac{1}{1-P(X)} \tag{5}$$

From the cumulative probability distribution, Equation(4), the expression relating the reduced variable, $\bar{y}$, to the return period, T, is

$$\bar{y}_T = -\ln(-\ln(T-1)/T) \tag{6}$$

If the n recorded events are placed in order of magnitude so that m=1 for the largest event and m=n for the smallest event then T=n+1/m and Equation (6) can be written as

$$\bar{y}_m = -\ln[-\ln\{(n+1-m)/(n+1)\}] \tag{7}$$

If the mean, $\mu_y$, and the variance, $\sigma^2_y$, of the series $\bar{y}_m$, m=1, 2, ...., n, are computed from the reduced sample as:

$$\mu_y = \sum_{m=1}^{n} \bar{y}_m / n \tag{8}$$

$$\sigma_y^2 = \sum_{m=1}^{n} (\bar{y}_m - \mu_y)^2 / n \tag{9}$$

and if $\mu$ and $\sigma^2$ are the mean and variance of the recorded events, then the parameters $\alpha$ and $\beta$ can be defined as

$$\alpha = \sigma_y / \sigma \tag{10}$$

$$\beta = \mu - \mu_y / \sigma \tag{11}$$

Introducing these relationships into the equation for the reduced variate, $\bar{y}$

$$\bar{y}_m = \alpha(X-\beta) \tag{12}$$

and rearrnaging for X, gives

$$X = \mu + (\bar{y}_m - \mu_y)\,\sigma / \sigma_y \tag{13}$$

Parameters of the type 1 extremal distribution for the applied watersheds are shown in Table 11. For convenience, Table 12 gives values of the reduced variable, $\bar{y}_T$ for some commonly used return periods.

**Table 10. Evaluation of parameters for the type 1 extremal distribution**

| River | Station | $\alpha$ | $\beta$ |
|---|---|---|---|
| Geum River | Seog Hwa | 0.001879 | 728.60 |
| | Gyu Am | 0.000795 | 1662.90 |
| Yeong San River | Ma Reug | 0.005122 | 329.29 |

**Table 11. Parameters for the type 1 extremal distribution**

| River | Station | Sample Size (n) | Discharge(cms) | | Mean and standard deviation of order statistics for various sample size | |
|---|---|---|---|---|---|---|
| | | | $\mu$ | $\sigma$ | $\mu_y$ | $\sigma_y$ |
| Geum River | Seog Hwa | 29 | 1035.73 | 682.50 | 0.5349 | 1.1086 |
| | Gyu Am | 29 | 2388.62 | 1612.69 | 0.5349 | 1.1086 |
| Yeong San River | Ma Reug | 27 | 441.96 | 250.39 | 0.5331 | 1.1006 |

**Table 12. Values of the reduced variable, $y_T$ according to the return period**

| Return Period($y_{rs}$) | Reduced Variable, $\bar{y}_T$ | Return Period($y_{rs}$) | Reduced Variable, $\bar{y}_T$ |
|---|---|---|---|
| 2 | 0.3665 | 50 | 3.9019 |
| 5 | 1.4999 | 100 | 4.6001 |
| 10 | 2.2504 | 200 | 5.2958 |
| 20 | 2.9702 | | |

**Table 13. Frequency factors for the type 1 extremal distribution according to the return period**

| Sample Size | Return Period($y_{rs}$) | | | | | | |
|---|---|---|---|---|---|---|---|
| (n) | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
| 27 | −0.1514 | 0.8784 | 1.5603 | 2.2143 | 3.0609 | 3.6953 | 4.3274 |
| 29 | −0.1519 | 0.8705 | 1.5474 | 2.1967 | 3.0372 | 3.6670 | 4.2945 |

comparing Equation (13) with the general frequency equation, it is apparent that for the type 1 extremal distribution the frequency factor, K, is defind as :

$$K=\frac{\bar{y}_m-\mu_y}{\sigma_y} \qquad (14)$$

Frequency factors, K can be tabulated as in Table 13 according to the sample sizes and the return period.

2) Derivation of probable flood flows according to the return period.

Table 14 gives formulas for the probable flood flows and probable flood flows according to the return period of the type 1 extremal distribution for the selected watersheds.

**Table 14. Formulas for the probable flood flows and probable flood flows according to the return period (Type 1 extremal distribution)**

Unit: Cms

| River | Station | Formula($X_T$) | Return Period($y_{rs}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
| Geum River | Seog Hwa | 1035.7+ 682.5K | 932 | 1630 | 2092 | 2535 | 3109 | 3538 | 3967 |
| | Gyu Am | 2388.6+1612.7K | 2144 | 3793 | 4884 | 5931 | 7287 | 8302 | 9314 |
| Yeong San River | Ma Reug | 441.96+ 250.4K | 404 | 662 | 833 | 996 | 1208 | 1367 | 1526 |

The results above mentioned are also plotted on a extremal probability paper as examples of Seog Hwa and Gyu Am watersheds in the Geum river basin as shown in Fig. 3.

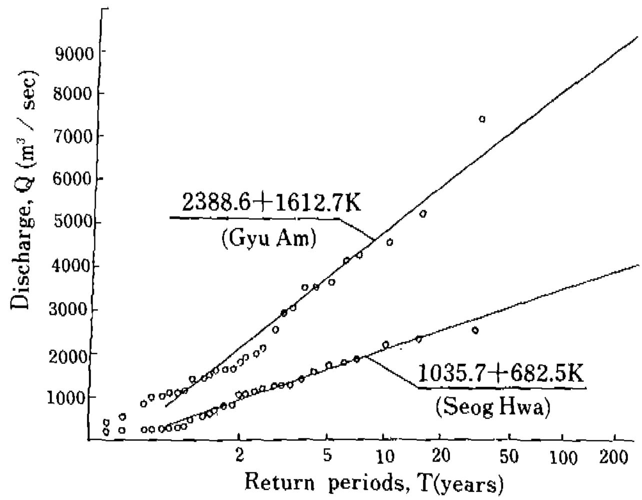


**Fig. 3. Probable flood flows according to the return period at Gyu Am and Seog Hwa watersheds in the Geum river system**

3. Probable flood flow by the Box-Cox transformation

a. Basic statistics

Basic statistics calculated by the Box-Cox transformation are within the range of 8.96 to 163.61, 0.93 to 55.40, −0.000808 to −0.000179 and 0.043 to 0.682 for the mean, standard deviation, coefficient of skewness and coefficient of transformation, respectively as shown in Table 15.

**Table 15. Basic statistics calculated by the Box-Cox transformation**

| River | Station | Mean ($\bar{z}$) | Standard deviation($\sigma_z$) | Coefficient of skewness($C_s$) | Coefficient of transformation($\lambda$) |
|---|---|---|---|---|---|
| Han River | Jeong Sun | 17.67 | 3.59 | −0.000808 | 0.302 |
| Nag Dong River | Im Ha | 108.01 | 46.40 | −0.000514 | 0.683 |
| Seom Jin River | Ab Nog | 163.61 | 55.40 | −0.000179 | 0.602 |
| Geum River | Seog Hwa | 29.09 | 8.38 | −0.0005319 | 0.359 |
| | Gyu Am | 8.96 | 0.93 | −0.000799 | 0.043 |
| Yeong San River | Ma Reug | 39.20 | 11.235 | −0.000388 | 0.487 |

b. Frequency factor

The frequency factors for the normal distribution are given in Table 16 for the corrseponding return period.

c. Probable flood flows by the Box-Cox transformation

Probable flood flows, y(T) of return period, T can be estimated from :

$$y(T)=(\lambda Z(T)+1)^{1/\lambda} \quad (15)$$

in which $Z(T)=\overline{Z}+K\sigma_z$, $\overline{Z}$ and $\sigma_z$ are the mean and the standard deviation of the transformed z series, respectively.

Probable flood flows were estimated by the Box-Cox transformation for applied watersheds as shown in Table 17.

**Table 16. Frequency factors according to the return period ($y_{rs}$)**

| Return Period | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
|---|---|---|---|---|---|---|---|
| Frequency Factor | 0 | 0.8416 | 1.2816 | 1.6449 | 2.0538 | 2.3264 | 2.5800 |

**Table 17. Flood flow prediction by the Box-Cox transformation for each watershed(cms)**

| River | Station | Return Period ($y_{rs}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
| Han River | Jeong Sun | 461 | 734 | 914 | 1082 | 1286 | 1444 | 1597 |
| Nag Dong River | Im Ha | 564 | 884 | 1073 | 1231 | 1417 | 1549 | 1675 |
| Seom Jin River | Ab Nog | 2157 | 3267 | 3919 | 4492 | 5174 | 5650 | 6108 |
| Geum River | Seog Hwa | 993 | 1619 | 2114 | 2564 | 3121 | 3550 | 3983 |
| | Gyu Am | 2120 | 3534 | 4663 | 5846 | 7405 | 8672 | 10304 |
| Yeong San River | Ma Reug | 394 | 640 | 829 | 1037 | 1221 | 1392 | 1476 |

d. Comparison of probable flood flows by the suitable frequency distributions, Box-Cox and SMEMAX transformation.

Comparing the relative suitabilities of the frequency distributions and transformations, the methods studied are used to calculate probable flood flows for various frequencies and these results are plotted along with the observed data on a normal probability paper as shown in Fig. 4 to Fig. 9.

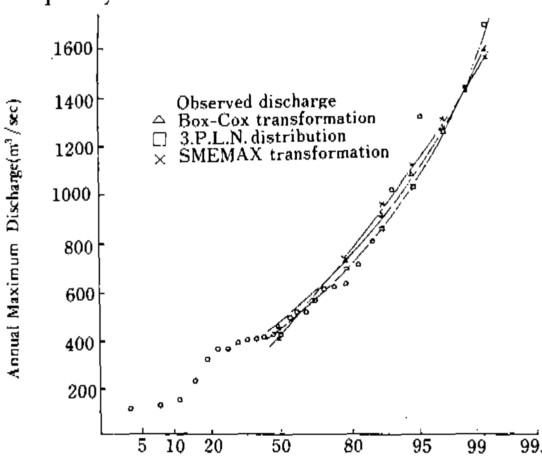


**Fig. 4. Comparison of probable flood flows by the transformations and three parameter lognormal distribution at Jeong Sun watershed in the Han river**

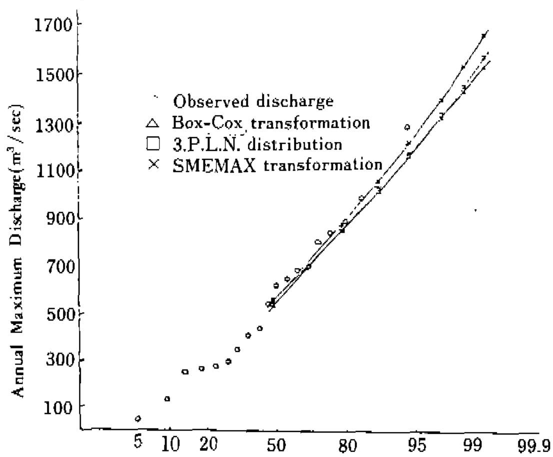


**Fig. 5. Comparison of probable flood flows by the transformations and three parameter lognormal distribution at Im Ha watershed in the Nag Dong River**

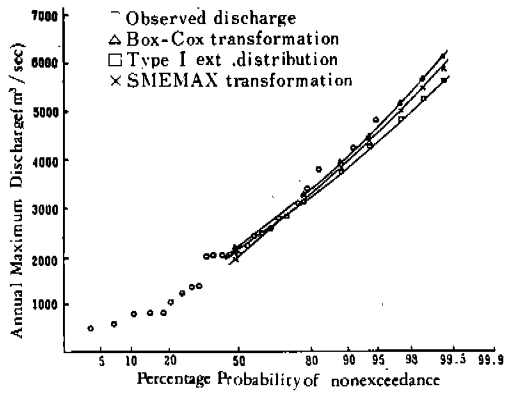


**Fig. 6. Comparison of probable flood flows by the transformations and three parameter lognormal distribution at Ab Nog watershed in the Seom Jin River**

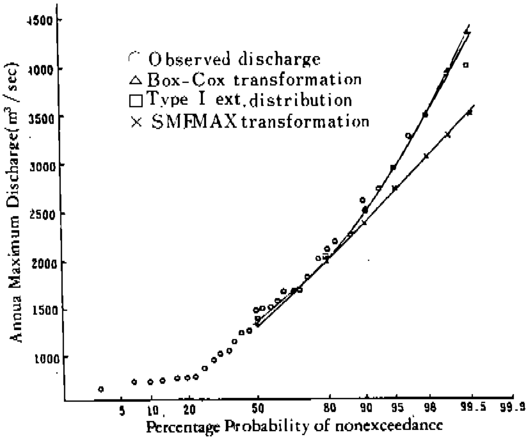


**Fig. 7. Comparison of probable flood flows by the transformations and type 1 extremal distribution at Seog Hwa watershed in the Geum River**

The plotting position is based on the weibull formula in which the probability of nonexceedance is calculated as $P=1-m/n+1$, in which n is the sample size and m is the rank commencing with the largest value. It can be seen from Fig. 4 and Fig. 9 that the results computed by the methods above mentioned are generally much closer to the observed data at the return period of less than ten years. Especially, it was confirmed that computed values based on Box-Cox transformation were shown to be closer to the observed data in comparison with the other methods even at the return period of more than ten years.

Consequently, it was proved that probable flood flows can be calculated by the Box-Cox transformation, the best method used in this study.

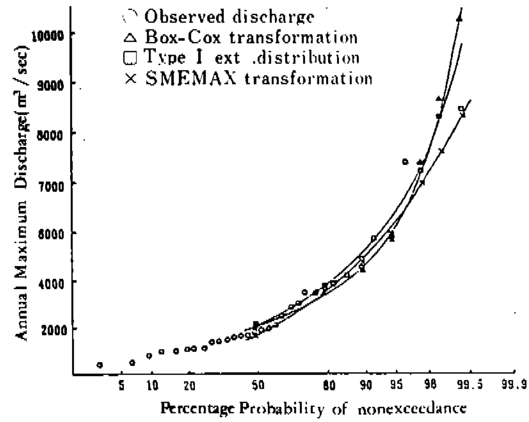


**Fig. 8. Comparison of probable flood flows by the transformations and type 1 extremal distribution at Gyu Am watershed in the Geum River**

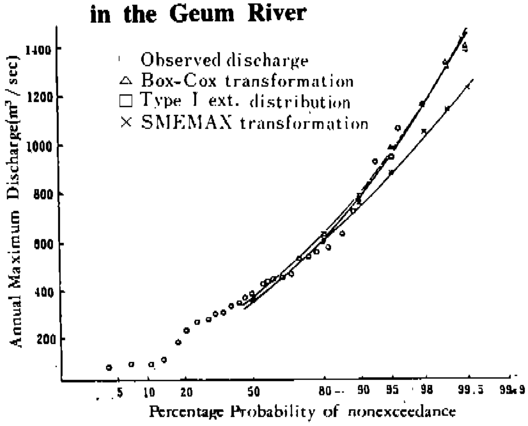


**Fig.9. Comparison of probable flood flows by the transformations and type 1 extremal distribution at Ma Reug watershed in the Yeong San River**

1) Comparison of probable flood flows between the Box-Cox and the SMEMAX transformation.

The proable flood flows estimated using the SMEMAX transformation presented by Lee[9] et al were compared with ones obtained from the Box-Cox transformation which is acknowledged as a suitable method for the applied watersheds as shown in Table 18.

Relative errors in the probable flood flows by the SMEMAX to those by the Box-Cox transformation were shown to be 0.5 to 8 percent in all return periods at Jeong Sun, Im Ha and Ab Nog watersheds and to be 7 to 21 percent in the range of fifty to two hundred years of the return period, while they were within 10 percent from two to twenty years of the return period, at Seog Hwa, Gyu Am and Ma Reug watersheds.

**Table 18. Comparison of probable flood flows calculated by the Box-Cox and SMEMAX transformation**

| River | Station | Distribution & Relative Error | Return Period ($y_{rs}$) 2 | 5 | 10 | 20 | 50 | 100 | 200 |
|---|---|---|---|---|---|---|---|---|---|
| Han River | Jeong Sun | Box-Cox | 461 | 734 | 914 | 1082 | 1286 | 1444 | 1597 |
| | | SMEMAX | 409 | 758 | 963 | 1133 | 1323 | 1451 | 1568 |
| | | R.E. | 0.113 | 0.033 | 0.054 | 0.047 | 0.029 | 0.005 | 0.018 |
| Nag Dong River | Im Ha | Box-Cox | 564 | 884 | 1073 | 1231 | 1417 | 1549 | 1675 |
| | | SMEMAX | 553 | 869 | 1042 | 1184 | 1343 | 1450 | 1549 |
| | | R.E. | 0.020 | 0.017 | 0.029 | 0.038 | 0.052 | 0.064 | 0.075 |
| Seom Jin river | Ab Nog | Box-Cox | 2157 | 3267 | 3919 | 4492 | 5174 | 5650 | 6108 |
| | | SMEMAX | 1988 | 3217 | 3891 | 4446 | 5073 | 5490 | 5879 |
| | | R.E. | 0.078 | 0.015 | 0.007 | 0.010 | 0.020 | 0.028 | 0.037 |
| Geum River | Seog Hwa | Box-Cox | 933 | 1619 | 2114 | 2564 | 3121 | 3550 | 3983 |
| | | SMEMAX | 926 | 1583 | 1978 | 2355 | 2671 | 2916 | 3144 |
| | | R.E. | 0.008 | 0.022 | 0.064 | 0.082 | 0.144 | 0.179 | 0.211 |
| | Gyu Am | Box-Cox | 2120 | 3534 | 4663 | 5846 | 7405 | 8672 | 10304 |
| | | SMEMAX | 1716 | 3672 | 4851 | 5822 | 6915 | 7643 | 8320 |
| | | R.E. | 0.191 | 0.039 | 0.040 | 0.004 | 0.066 | 0.119 | 0.193 |
| Yeong San River | Ma Reug | Box-Cox | 394 | 640 | 829 | 1037 | 1221 | 1392 | 1476 |
| | | SMEMAX | 384 | 655 | 815 | 948 | 1097 | 1197 | 1289 |
| | | R.E. | 0.025 | 0.023 | 0.017 | 0.085 | 0.102 | 0.140 | 0.127 |

2) Comparison of probable flood flows between Box-Cox transformation and suitable probability distributions.

Probable flood flows calculated by the Box-Cox transformation were compared with ones obtained from the three parameter longnormal and the type 1 extremal distribution which are judged by suitable distributions as in Table 19 and Table 20. In relative errors of the probable flood flows calculated by the three parameter lognormal and the type 1 extremal distribution compared with the ones obtained from the Box-Cox transformation, both of them were found to be within 10 percent in all return periods in those two groups of watersheds.

This can clearly be seen in Table 20 that relative errors of the type 1 extremal distribution to the Box-Cox transformation appeared as lower values than 1.1 percent in all return periods at Seog Hwa watershed.

**Table 19. Comparison of probable flood flows calculated by Box-Cox transformation and the three parameter lognormal distribution**

| River | Station | Distribution & Relative Error | Return Period ($y_{rs}$) 2 | 5 | 10 | 20 | 50 | 100 | 200 |
|---|---|---|---|---|---|---|---|---|---|
| Han River | Jeong Sun | Box-Cox | 461 | 734 | 914 | 1082 | 1286 | 1444 | 1597 |
| | | 3.P.L.N | 441 | 687 | 860 | 1032 | 1264 | 1445 | 1636 |
| | | R.E. | 0.043 | 0.064 | 0.059 | 0.046 | 0.017 | 0.001 | 0.024 |
| Nag Dong River | Im Ha | Box-Cox | 564 | 884 | 1073 | 1231 | 1417 | 1549 | 1675 |
| | | 3.P.L.N | 567 | 865 | 1033 | 1179 | 1351 | 1470 | 1584 |
| | | R.E. | 0.005 | 0.021 | 0.037 | 0.042 | 0.047 | 0.051 | 0.054 |
| Seom Jin River | Ab Nog | Box-Cox | 2157 | 3267 | 3919 | 4492 | 5174 | 5650 | 6108 |
| | | 3.P.L.N | 2137 | 3152 | 3726 | 4224 | 4813 | 5221 | 5615 |
| | | R.E. | 0.009 | 0.035 | 0.049 | 0.060 | 0.070 | 0.076 | 0.081 |

**Table 20. Comparison of probable flood flows calculated by Box-Cox transformation and type 1 extremal distribution**

| River | Station | Distribution & Relative Error | Return Period ($y_{rs}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | 2 | 5 | 10 | 20 | 50 | 100 | 200 |
| Geum River | Seog Hwa | Box-Cox | 933 | 1619 | 2114 | 2564 | 3121 | 3550 | 3983 |
| | | Type 1 ext. | 932 | 1630 | 2092 | 2535 | 3109 | 3538 | 3967 |
| | | R.E.(%) | 0.001 | 0.007 | 0.011 | 0.011 | 0.004 | 0.003 | 0.004 |
| | Gyu Am | Box-Cox | 2120 | 3534 | 4663 | 5846 | 7405 | 8672 | 10304 |
| | | Type 1 ext. | 2144 | 3793 | 4884 | 5931 | 7287 | 8302 | 9314 |
| | | R.E.(%) | 0.011 | 0.073 | 0.047 | 0.015 | 0.016 | 0.043 | 0.096 |
| Yeong San River | Ma Reug | Box-Cox | 394 | 640 | 829 | 1037 | 1221 | 1392 | 1476 |
| | | Type 1 ext. | 404 | 662 | 833 | 996 | 1208 | 1367 | 1525 |
| | | R.E.(%) | 0.025 | 0.034 | 0.005 | 0.040 | 0.011 | 0.018 | 0.033 |

## IV. SUMMARY AND CONCLUSIONS

This paper has attempted to show that probable flood flows can be estimated by means of the Box-Cox transformation which is more efficient for the normalization of frequency distribution than any other transformations in making the coefficient of skewness nearer to zero and to compare with the results computed by the SMEMAX transformation and fitted probability distributions to the annual maximum series of six watersheds in the Han, Geum, Nag Dong, Seom Jin and Yeong San river basins. The results were analyzed and summarized as follows.

1. The Box-Cox transformation has been found to be the best in comparison to the SMEMAX, longarithmic and squared root transformation for making the coefficient of skewness closer to zero as a means of getting the normalization of frequency distribution.

2. The Box-Cox transformation was proved to be more effective than the SMEMAX transformation in normalizing the skewed distribution.

3. Three parameter lognormal and type 1 extremal distributions were tested as suitable ones at six applied watersheds by the results of $\chi^2$ and the Kolmogorov-Smirnov test in the annual maximum series. The former was well fitted to Jeong Sun, Im Ha and Ab Nog watersheds in the Han, Nag Dong, and Seom Jin rivers, respectively while the latter was well fitted to the Seog Hwa and Gyu Am watersheds of the Geum river and Ma Reug of the Yeong San river.

4. Probable flood flows according to the return periods were derivated by the Box-Cox and SMEMAX transformations and by good fitted distributions for the applied watersheds.

5. Judging by the relative suitabilities of the various methods, it was confirmed that the values calculated using the Box-Cox transformation are nearer to the observed data as compared with other methods, especially at higher probability of nonexceedance.

## REFERENCES


1. Benjamin, J.R. and C.A. Cornell (1970), Probability, Statistics, and Decision for civil Engineers, McGraw-Hill, New York, pp. 370-500, English.
2. Bethlahmy, N.(1977), Flood Analysis by SMEMAX Transformation, Journal of the Hydraulic Division, ASCE, Vol. 103, No.HY1. Proc. Paper 12701 Jan., pp.69-80, English.
3. Box, G.E.P. and D.R. Cox (1964), An analysis of transformation, Journal of the Royal Statistical Society, Vol. B 26, pp.211-252, English.
4. Emir Zelenhasic (1970), Theoretical probability distribution for flood peaks, Hyd. Papers, No. 42, C.S.U., pp.1-21, English.
5. Ministry of Construction (1962-1978), Hydrological Investigation of Korean Watersheds (Rainfall and Water Stage), Korean.
6. Ministry of Construction (1964), An Annual Report of Hydrological Investigation in Korea, Korean.
7. Ministry of Construction(1974), Watershed Investigation Report in Korea, Korean.

8. Ministry of Construction(1979-1982), An Annual Report of Hydrological Investigation in Korea, Korean.
9. Lee, S.H. and M.K. Park(1985), Flood frequency analysis by SMEMAX transformation, Journal of Chung Buk National University, Vol.30, pp.165-175, Korean.
10. Lee, S.H. and M.K. Park(1985), Hydrological studies on the design flood and risk of failure of the Hydraulic structure (I) (on the annual maximum series), Journal of the Korean Society of Agricultural Engineers, Vol.27(2), pp.23-37, Korean.
11. Lee, S.H., S.P. Hong, M.K. Park(1988), A study on the flood frequency analyzed in consideration of low outliers, Journal of the Korean Society of Agricutural Engineers, Vol.30(4), pp.62-70, Korean.
12. Rasheed, H.R., M.V. Ramamoorthy and A.S. Aldabbagh(1982), Modified SMEMAX transformation for frequency analysis, W.R.P., Vol.18, No.3, pp.509-511, English.
13. Subbash Chander, S.K.Spolia, and Arun Kumar (1978), Flood frequency Analysis by power transformation, Journal of the Hydraulic Division, Vol.Hy 11, pp.1495-1503. English.
14. Yevjevich, V.(1972), Probability and Statistics in Hydrology, W.R.P. Colorado, pp.118-167, English.