THE STUDY OF PARAMETRIC AND NONPARAMETRIC MIXTURE DENSITY ESTIMATOR FOR FLOOD FREQUENCY ANALYSIS

Young-Il Moon
Department of Civil Engineering, University of Seoul, Seoul, Korea

Abstract: Magnitude-frequency relationships are used in the design of dams, highway bridges, culverts, water supply systems, and flood control structures. In this paper, possible techniques for analyzing flood frequency at a site are presented. A currently used approach to flood frequency analysis is based on the concept of parametric statistical inference In this analysis, the assumption is made that the distribution function describing flood data is known. However, such an assumption is not always justified. Even though many people have shown that the nonparametric method provides a better fit to the data than the parametric method and gives more reliable flood estimates, the nonparametric method implies a small probability in extrapolation beyond the highest observed data in the sample. Therefore, a remedy is presented in this paper by introducing an estimator which mixes parametric and nonparametric density estimates.

Key Words: flood frequency analysis, mixture, parametric, nonparametric

1. INTRODUCTION

A currently used approach to flood frequency analysis is based on the concept of parametric statistical inference. In these analyses, the assumption is made that the distribution function describing flood data is known. Distributions that are often used are Normal, Log-Normal, Pearson Type III, Log-pearson Type III, Wakeby, Gumbel, Gamma, and others. Some difficulties associated with parametric estimation are (i) the objective selection of a distribution (ii) the reliability of distributional parameters (especially for skewed data with a short record length) (iii) the inability to analyze multimodal distributions that may arise from a mixture of causative factors, and (iv) the treatment of outliers. Therefore, classical density estimation techniques may be inadequate for modelling such an annual maximum process or the extreme tail of the distribution. However, nonparametric methods do not require assumptions about the underlying populations from which the data are obtained. Therefore, they are better suited for multimodal distributions. In recent years, several nonparametric approaches that have promise for estimating the probability density function of annual floods have been introduced by Adamowski (1985, 1989, and 1996), Adamowski and Feluch (1990), Adamowski and Labatiuk (1987), Lall et al. (1993), Moon et al. (1993), Moon and Lall (1994 and 1995), and Moon (1996 and 1999).

Even though many people have shown that the nonparametric method provides a better fit to the data than the parametric method and gives more reliable flood estimates, the nonparametric method implies a very small probability in extrapolation beyond the highest observed data in the sample. Schuster and Yakowitz (1985) offered a remedy for these inadequacies by introducing an estimator which mixes parametric and nonparametric density estimates. They proposed a plan to implement the nonparametric estimate feature of sidestepping the model-choice dilemma while overcoming the drawback of the lack of an extended tail and also provided an avenue for incorporating prior information. In brief, the plan calls for choosing a parametric family and then approximating the unknown PDF(probability density function) $h_n(x)$ for the x(i)'s by a mixture

$$h_n(x) = A(n)f_n(x) + (1 - A(n))g_n(x) \qquad (1)$$

where $f_n(x)$ is a nonparametric density estimator of f(x), $g_n(x)$ is a parametric density estimator of g(x), and A(n) is a number in the unit interval [0,1]. Schuster and Yakowitz (1985) showed that as the number n of samples increases, the estimator $h_n(x)$ is strongly consistent. The mixture or *semiparametric estimator* is developed by constructing nonparametric and parametric estimators $f_n(x)$ and $g_n(x)$ and mixing them according to the rule given the equation (2) which maximizes the log likelihood function (Schuster and Yakowitz, 1985).

$$\ln(A) = \sum_{i=1}^{n} \ln[A(n)f_n(x(i)) + (1 - A(n))g_n(x(i))] \qquad (2)$$

In summary, if a correct parametric distribution is available, the mixture estimator may have value A(n) near 1. On the other hand, if the parametric distribution is not correct, then A(n) is expected to be less than 1 according to the equation (1).

2. NONPARAMETRIC KERNEL DENSITY FUNCTION

Rosenblatt (1956) introduced the kernel estimator, defined for all real x by

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \qquad (3)$$

where $x_1, \ldots, x_n$ are independent identically distributed real observations, $K(\cdot)$ is a kernel function, and h is a positive smoothing factor assumed to tend to zero as n tends to infinity. Parzen (1962) generalized and investigated the consistency properties of the Rosenblatt kernel estimator. He popularized it to such an extent that the estimator is also called the Rosenblatt-Parzen estimator. Silverman (1986) explained the basic concept of the kernel estimator. From the definition of a probability density, if the random variable x has density f(x), then

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h) \qquad (4)$$

For any given h, $P(x - h < X < x + h)$ can be estimated by the proportion of the sample falling in the interval (x-h, x+h). Thus, a natural estimator is given by choosing a small number h and setting

$$\hat{f}(x) = \frac{1}{2hn}[\#\ \text{of}\ X_1, \ldots, X_n\ \text{falling in}(x - h,\ x + h)] \qquad (5)$$

We shall call this the naive estimator. To express the estimator more transparently, define the weight function w(x) by

$$w(x)=1/2 \text{ if } abs(x) < 1 \qquad (6)$$
$$0 \quad \text{if otherwise}$$

Then it is easy to see that the estimator can be written as

$$f(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}W\left(\frac{x-X_i}{h}\right) \qquad (7)$$

It follows from equation (6) that the estimator is constructed by placing a box of width 2h and height $(2nh)^{-1}$ on each observation and then summing to obtain the estimator. This weight function is the kernel function which satisfies the condition

$$\int K(t)\,dt = 1, \quad \text{where } t=\frac{x-X_i}{h} \qquad (8)$$

The kernel function is usually required to be unimodal with peak at $x=0$, smoothness, and a symmetric function, that is, a density ($\int K(t)dt=1$) with expectation 0 ($\int tK(t)dt=0$) and finite variance ($\int t^2K(t)dt=\text{constant}$).

In addition, for the estimate to converge, as n approaches infinity, nh should tend to approach zero. When applying the method in practice it is necessary to choose a kernel function and a smoothing parameter. Some useful kernel functions are given in Table 1 and Fig. 1. Usually, different kernels should be examined depending the objective. For example, if continuity and differentiability of the density is needed, one may choose a kernel with infinite support rather than one with finite support.

While the choice of kernel does not seem to be critical, the choice of smoothing factor is quite a different matter. The value of h is critical and, in practice, not obvious. Too large an h implies large bias, an oversmoothed estimate, and consequent loss of information. Too small an h implies large variance and too rough an estimate (Adamowski and Labatiuk, 1987). Since all error measures depend on the unknown density, generally they cannot be used in deriving analytical expressions for selecting the smoothing factor h. Several measures of performance for using the data to produce suitable values for the smoothing parameter h have been proposed. The smoothing parameter h can be obtained by maximum likelihood cross-validation, least squares cross-validation, Breiman et al. method, and

Table 1. Typical Kernel Functions

Kernel	K(t)
Rectangular	$\frac{1}{2}$ for $\lvert t\rvert<1$; 0 if otherwise
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)$
Epanechnikov	$\frac{3}{4}\left(1-\frac{1}{5}t^2\right)/\sqrt{5}$ for $\lvert t\rvert<\sqrt{5}$
Cauchy	$\frac{1}{\pi(1+t^2)}$

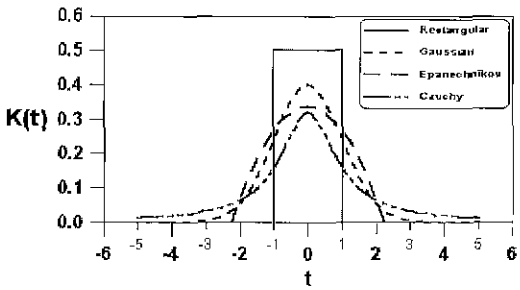

Fig. 1. The Shape of Kernel Functions.

Adamowski cross-validations.

2.1 Maximum Likelihood Cross-validation

Habbema et al. (1974) and Duin (1976) used the concept of likelihood to justify choosing h to maximize

$$\prod_{i=1}^{n} f_{-i}(x_i) \tag{9}$$

where $f_{-i}(x_i)$ is the density estimator based on the data with observation x_i omitted. The method of likelihood cross-validation is the natural development of the idea of using likelihood to judge the adequacy of fit of a statistical model. It can be applied generally, not just in density estimation (Stone, 1974; and Geisser, 1975). Suppose that, in addition to the original data set, an independent observation Y from f were available. Then, the likelihood of f as the density underlying the observation Y would be log(f(Y)), with f regarded as a parametric family of densities depending on the window width h but with the data $x_1, x_2, \ldots x_n$ fixed. This would give log(f(Y)), regarded as a function of h, as the log likelihood of the smoothing parameter h. Since an independent observation, Y, is not available, one of the original observations x_i from the sample used to construct the density estimate could be omitted, and X_i could be used as the observation Y. This would define log likelihood log f_{-i} to be the density estimate constructed from all the data points except X_i, that is,

$$\hat{f}_{-i}(x) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x - X_i}{h}\right) \tag{10}$$

Since there is nothing special about the choice of which observation to leave out, the log likelihood is averaged over each choice of omitted X_i to give the score function

$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_{-i}(X_i) \tag{11}$$

The likelihood cross-validation choice of h is then the value of h which maximizes the function CV(h) for the given data. It is known that, under lenient circumstances, this procedure is consistent if f(x) is 0 outside a finite interval. However, for long-tailed distributions, Schuster and Gregory (1981) have shown that the maximum likelihood method is not even consistent, and Marron (1985) demonstrated that when the kernel and the density are compactly supported, even though the resulting density estimator will be consistent, the consistency can be very slow and the selected bandwidth very poor.

2.2 Least Squares Cross-validation

Least squares cross-validation is a completely automatic method for choosing the smoothing parameter. It has only been formulated in recent years, but it is based on an extremely simple idea. The method was suggested by Rudemo (1982) and Bowman (1984).

Given any estimator $\hat{f}$ of a density f, the integrated squared error can be written

$$\int (\hat{f} - f)^2 = \int \hat{f}^2 - 2\int \hat{f}f + \int f^2 \tag{12}$$

Now the last term of above equation does not depend on estimator $\hat{f}$, and so the ideal choice of window width will correspond to the choice which minimizes the quantity R defined by

$$R(\hat{f}) = \int \hat{f}^2 - 2\int \hat{f}f \tag{13}$$

The basic principle of least square cross-validation is to construct an estimate of $R(\hat{f})$ from the data themselves and then to minimize this estimate over h to give the choice of window width. The term $\int \hat{f}^2$ can be found from the estimate $\hat{f}$. Define

$$\hat{f}_{-i}(x) = \frac{1}{(n-1)h} \sum_{j \neq i} k\left(\frac{x - X_j}{h}\right) \tag{14}$$

Therefore, define

$$M_0(h) = \int \hat{f}^2 - 2n^{-1} \sum_i \hat{f}_{-i}(X_i) \tag{15}$$

The score M_0 depends only on the data. The idea of least squares cross-validation is to minimize the score M_0 over h. The fact that the smoothing factor h chosen in this procedure is asymptotically correct for IMSE has been shown under various assumptions by Hall (1983), Hall and Marron (1987), and Stone (1984). The main strength of this smooth ing factor is that it is asymptotically correct under very weak smoothness assumptions on the underlying density. However, a drawback to least square cross-validation is that the score function $M_0(h)$ has a tendency towards having several local minimum values with some spurious ones often quite far over on the side of undersmoothing. For this reason, it is recommended that minimization be done by a grid search through a range of h's instead of by some sort of computationally more efficient step-wise minimization algorithm which will converge quite rapidly but only to a local minimum. An unexplored possibility for approaching the local minimum problem is to select the bandwidth which is the largest value of h for which a local minimum occurs.

2.3 Breiman et al. Method

Breiman et al. (1977) proposed a variable kernel estimator given by

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_k d_{j,k}} K\left(\frac{x - X_j}{h_k d_{j,k}}\right) \tag{16}$$

where $d_{j,k}$ is the distance from the point x_j to its k^{th} nearest neighbor, and h_k is a constant smoothing factor. Thus, the constant smoothing factor h in the fixed kernel estimate is replaced by a variable smoothing parameter $h_k d_{j,k}$ which is adjusted depending on the value of the local density of data. Breiman et al. (1977) proposed a numerical minimization procedure such that the variance of $\sum_{i=1}^{k} h_i d_{j,i}$ is a minimum. In an extensive numerical experiment, Breiman et al. (1977) made the empirical discovery that

$$h_k \frac{(\overline{d_k})^2}{\sigma(d_k)} = \text{constant} \tag{17}$$

for an optimal value of k where d_k is the mean of the k^{th} nearest neighbor distances, and $s(d_k)$ is their standard deviation. Based on a goodness-of-fit test, they developed a criterion to select optimal k and h. For the case when f is a p-variate density and the sample points $\{x_i\}_{i=1, \ldots, n}$ are random p-vectors, they use the argument that the random variable given by

$$w_j = \exp(-f(x_j)\, nV_j) \tag{18}$$

where V_j is the volume of the p-dimensional sphere with radius equal to the distance from x_j to its nearest neighbor, is uniformly distributed. To select optimizing values of k and h, in case of univariate, minimize the following

$$S=\sum_{j=1}^{n}\left(W_j-\frac{i}{n}\right)^2 \tag{19}$$

where $W_j=\exp(-\hat{f}(x_j)nd_{j,i})$ and $j=1,2,\ldots,n$

Breiman et al. (1977) showed that the variable kernel results were extremely sensitive to the choice of k and that the variable kernel had better results when k was large enough.

2.4 Adamowski Cross-validation

The value of h can be estimated, but the numerical approximation of the smoothing factor h can be determined by minimizing the following MSE expression (Adamowski, 1985).

$$\sum_{j=1}^{n}(\hat{F}_j(x)-\hat{F}(x_j))^2 \tag{20}$$

where the unknown probability $\hat{F}_j(x)$ is estimated by the probability plotting formula $\hat{F}(x_j)$ (the empirical distribution function).

The existing practice in selection of a particular formula is rather arbitrary, and Weibull's formula (i / (n+1)), which provides biased and conservative results, is often recommended. Based on the MSE criterion, Adamowski (1981) developed a new plotting formula

$$\hat{F}(x)=\frac{j-0.25}{n+0.5} \quad \text{where } j=1,2,\ldots,n \tag{21}$$

Adamowski showed that when the flood frequency analysis is performed using Gumbel Type I distribution then the new plotting formula can provide a good approximation to true exceedance probability at high values and that for Pearson Type III distribution no single plotting formula is strictly correct. However, if one formula which can be used with all distributions was required as a compromise, then the new plotting formula was suitable.

The Adamowski criteria is related to the Cramer-von Mises criteria given by

$$D(F,F_n)=n\int(F-F_n)^2\,dF(x) \tag{22}$$

where F is the true distribution function, and F_n is the estimated distribution function. Good and Gaskins (1980) proposed a method of choice of smoothing factor based on a goodness-of-fit statistic which measures the distance between the empirical distribution function F_n and $\hat{F}_n$, constructed by integrating the kernel density estimate, that is,

$$D(\hat{F}_n, F_n)=0.1188 \tag{23}$$

The constant, 0.1188, is the asymptotic median of $D(F, \hat{F}_n)$. Adamowski's MSE expression is similar to Good and Gaskins' goodness-of-fit statistic as follows:

$$\begin{aligned} D(F,F_n) &= n\int(F-F_n)^2 dF(x) \\ &= n\sum_{i=1}^{n}(F_i-F_{n,i})^2(F_i-F_{i-1}) \\ &= n\sum_{i=1}^{n}(\hat{F}_i-\hat{F}_i)^2(\hat{F}_i-\hat{F}_{i-1}) \\ &= n\sum_{i=1}^{n}(\hat{F}_i-\hat{F}_i)^2 w_i \end{aligned} \tag{24}$$

where

$$w_1=\frac{1}{2}\left(\frac{3-0.5}{n+0.5}\right)$$

$$w_i=\frac{1}{2}\left(\frac{i+1-0.25}{n+0.5}+\frac{i-0.25}{n+0.5}\right)-\frac{1}{2}\left(\frac{i-0.25}{n+0.5}+\frac{i-1-0.25}{n+0.5}\right)=\frac{1}{(n+0.5)},$$

as $i = 2, \ldots, n-1$

$$w_n=1-\frac{1}{2}\left(\frac{2n-1.5}{n+0.5}\right) \tag{25}$$

3. EXPERIMENTS WITH MIXTURE DENSITY FUNCTION

Nonparametric methods tend to place negligible probability on the interval of points larger than the maximum of the observed x(i)'s. For example, often the 100-year flood level needs to be estimated based on 20 or 50 annual data. The nonparametric estimators do not use hydrological experience or related data such as records at similar gauging stations. Schuster and Yakowitz (1985) proposed a remedy for these difficulties by introducing a mixture density estimator which mixes parametric and nonparametric density estimators. They used a maximum likelihood cross-validation to choose the bandwidth h. By this procedure, the mixture estimator could solve the model choice problem and overcome the drawback of lack of tail; in addition, an avenue could be provided for incorporating prior information. An approximate parametric estimator, such as normal, gamma, and Gumbel, may be chosen. In this experiment, the parametric family was chosen to be the normal density function. To construct the nonparametric PDF, the normal, Epanechnikov, or cauchy kernel functions were used. The nonparametric kernel estimators require the smoothing factor h, which is obtained by maximum likelihood cross-validation, least squares cross-validation, Breiman et al. approach, and Adamowski cross-validation methods.

The steps of the experimental procedure were:

1) Generate samples of size 20, and 100 from N(0,1), Pearson Type III (0,1,1), and mixture = {0.5 N(0,1) , 0.5 N(3,1)}.

2) Consider the normal density estimator.

3) Consider the combination of the nonparametric kernel estimators and parametric estimators.

4) Consider the normal, Epanechnikov, and cauchy kernel functions.

5) Consider the maximum likelihood method to select weighting factor A(n).

6) Consider tail probability estimates (cumulative density function (CDF) values) at P = 0.9, 0.95, 0.98, 0.99, and 0.995 (i.e., return period 10, 20, 50, 100, and 200 years). Specify the true x values for each parent corresponding to these probabilities, use each method to compute the predicted probability in each case, and compute the errors between estimators obtained by different methods and the population values using the bias and RMSE in prediction for each event, that is,

$$\text{bias} = E[\hat{F}(x) - F(x)] \quad (26)$$

$$\text{MSE} = E[\hat{F}(x) - F(x)]^2 = \text{bias}^2 + \text{variance} \quad (27)$$

$$\text{RMSE} = (\text{MSE})^{1/2} \quad (28)$$

or

$$\frac{\text{bias}}{1-P} \text{ and } \frac{\text{RMSE}}{1-P} \quad (29)$$

7) Repeat steps 1 - 5 1000 times.

4. RESULTS AND CONCLUSIONS

The average of the bias and RMSE for each selection procedure was then computed. The average of the bias and RMSE for the normal, Epanechnikov, and cauchy kernel functions with maximum likelihood cross-validation, least squares cross-validation, Breiman et al. approach, and Adamowski cross-validation methods. The results give a good indication which kernel function and bandwidth selection is a good combination for the normal data, gamma data, and mixture data of sample size 20 and 100. The results show that the Epanechnikov kernel function combined with maximum likelihood method usually produced the best result. Therefore, the variable kernel estimator combined the Epanechnikov kernel function, with maximum likelihood method, was used as a nonparametric family.

4.1 Case 1-Estimation when the Parametric Family is Correct

In the first case, 20 and 100 observations of the normal variable were simulated with mean=0 and variance=1, and the parametric, nonparametric, and the mixture estimators shown in Figs. 2 through 5 were applied. In this case, the parametric estimator can be assumed to be correct because observations are from parametric family. The mixture estimator provided better results than the nonparametric estimator.

4.2 Case 2-Estimation outside the Range of Data when the Parametric Family is Incorrect

In the second case, 20 observations were generated according to the gamma (0,1,1) and mixture {0.5N(0,1)+0.5N(3,1)} families, respectively. The results are shown in Figs. 6 through 9. Even though the parametric estimator is incorrect, the results are better than those of the nonparametric estimator for return period 100 and 200 years since the nonparametric estimators do not have tail on the interval of points larger than the maximum of the observed x(i)'s. That is, only 20-year maximum flood data is available, but the 100-year and 200-year maximum flood probability needs to be estimated. In this case, the mixture estimator also had better results than the nonparametric

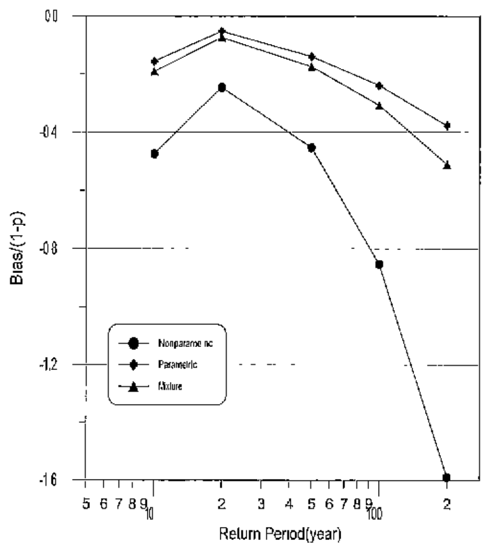

Fig. 2. Bias/(1-P) of Mixture Distribution for Normal Data N(0,1), n = 20

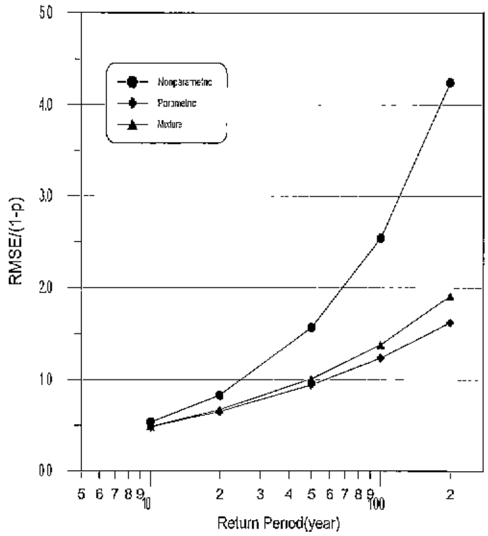

Fig. 3. RMSE/(1-P) of Mixture Distribution for Normal Data N(0,1), n = 20

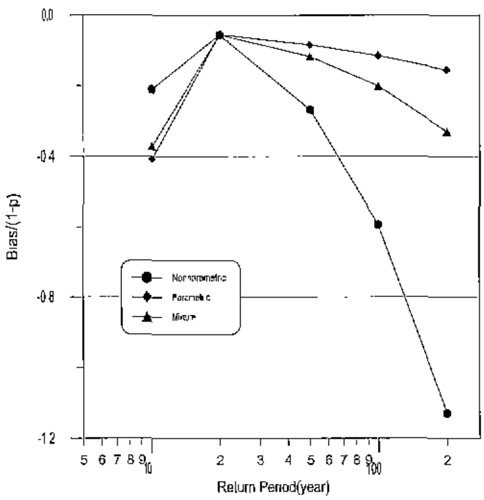

Fig. 4. Bias/(1-P) of Mixture Distribution for Normal Data N(0,1), n = 100

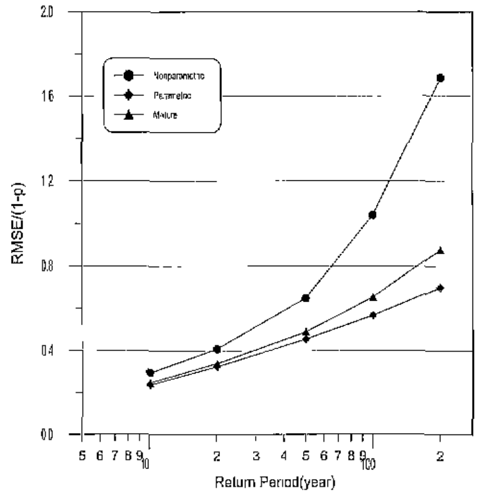

Fig. 5. RMSE/(1-P) of Mixture Distribution for Normal Data N(0,1), n =100

estimator.

4.3 Case 3-Estimation within the Range of Data when the Parametric Family is Incorrect

In the third case, 100 observations were generated according to the gamma (0,1,1) and mixture {0.5N(0,1)+0.5N(3,1)} families,

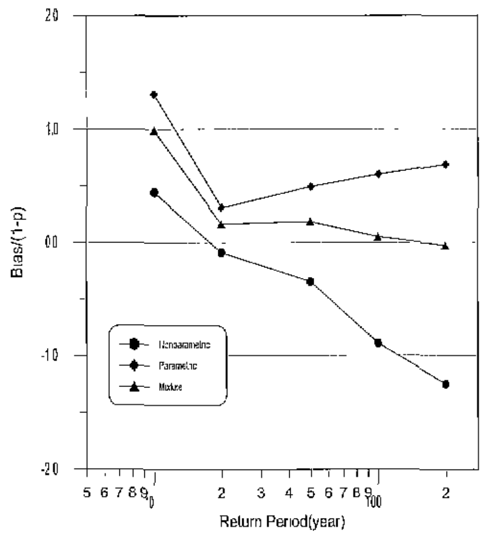

Fig. 6. Bias/(1-P) of Mixture Distribution for Gamma Data (0,1,1), n = 20

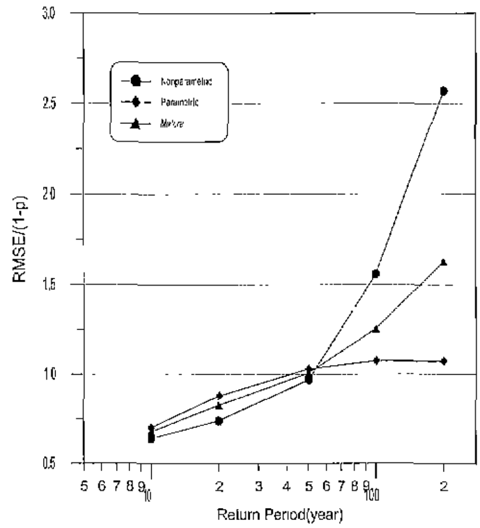

Fig. 7. RMSE/(1-P) of Mixture Distribution for Gamma Data (0,1,1), n = 20

respectively. As shown in Figs. 10 through 13, the nonparametric estimator usually provided better results than the parametric estimator. The nonparametric and mixture estimators had the same behavior.

The results for mixing weight A(n) values in each case are shown in Table 2.

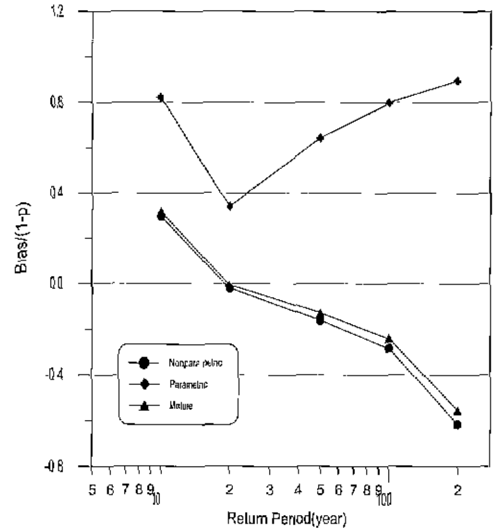

Fig. 8. Bias/(1-P) of Mixture Distribution for Gamma Data (0,1,1), n = 100

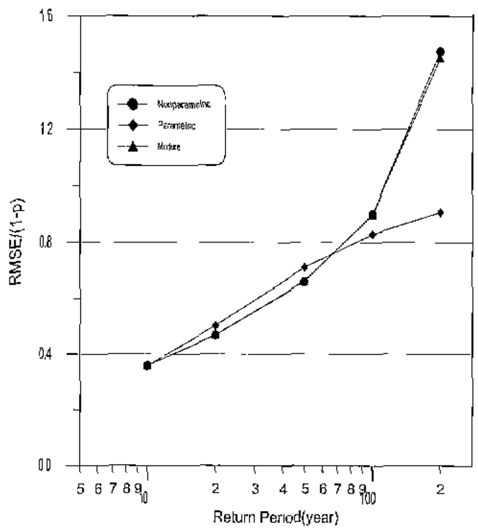

Fig. 9. RMSE/(1-P) of Mixture Distribution for Gamma Data (0,1,1), n =100

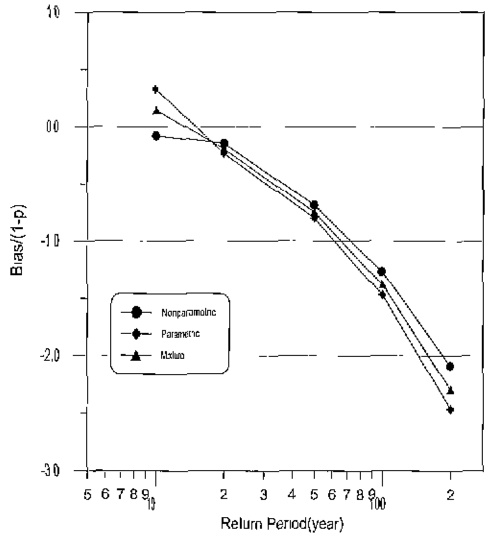

Fig. 10. Bias/(1-P) of Mixture Distribution for Mixture Data{0.5N(0,1) + 0.5N(3,1)}, n = 20

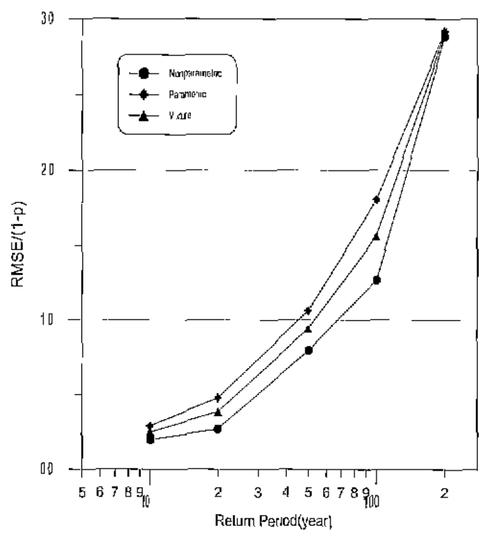

Fig. 11. RMSE/(1-P) of Mixture Distribution for Mixture Data {0.5N(0,1) + 0.5N(3,1)}, n = 20

If the right parametric estimator can be chosen, then there is no problem in estimating annual maximum flood. However, the true form of the distribution function is never exactly known in the hydrological context and must be assumed in practical situations. When the parametric estimator was not correct, only a limited data set was available, and it was necessary to estimate outside the range of data, the nonparametric estimator could produce good upper-tail results since the nonparametric estimator does

Table 2. Mixing Weight A(n) for Each Case

Case	Observations	Size	Mixing Weight A(n)
1	Normal	20	0.37
	Normal	100	0.45
2	Gamma	20	0.47
	Mixture	20	0.96
3	Gamma	100	0.92
	Mixture	100	0.98

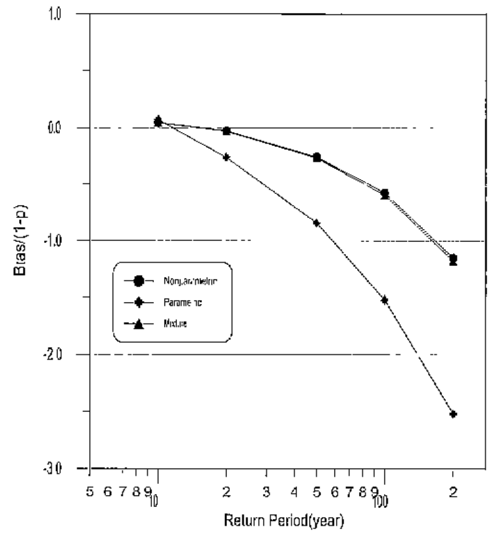

Fig. 12. Bias/(1-P) of Mixture Distribution for Mixture Data{0.5N(0,1) + 0.5N(3,1)}, n = 100

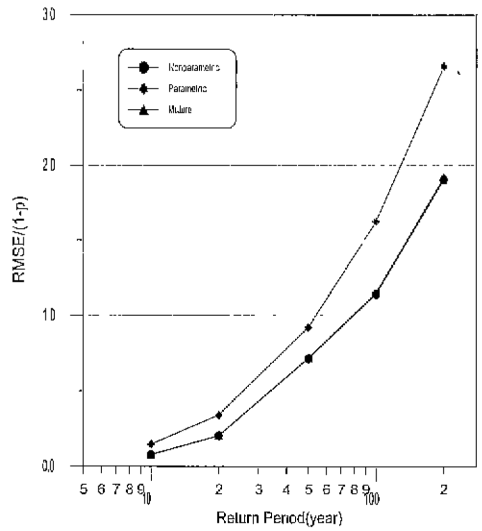

Fig. 13. RMSE/(1-P) of Mixture Distribution for Mixture Data{0.5N(0,1) + 0.5N(3,1)}, n = 100

not have tail on intervals greater than the largest of the observed data x(i)'s. In this case, the mixture estimator enhances the tail parts by mixing a parametric estimator with a nonparametric estimator. Improved methods for kernel function and bandwidth selection will help to analyze flood frequency. By this procedure, incorporation of prior information, experience, and regional data information is allowed.

ACKNOWLEDGEMENTS

Partial support of this work by the university of seoul is acknowledged.

REFERENCES

Adamowski, K. (1981). "Plotting formula for flood frequency." *Water Resources Bulletin*, Vol. 17, No. 2, pp. 197-202.

Adamowski, K. (1985). "Nonparametric kernel estimation of flood frequency." *Water Resources Research*, Vol. 21, No. 11, pp. 1585-1590.

Adamowski, K. (1989). "A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies." *J. of Hydrology*, Vol. 108, pp. 295-308.

Adamowski, K., and Feluch, W. (1990). "Nonparametric flood-frequency analysis with historical information." *J. of Hydraulic Engineering*, Vol. 116, No. 8, pp. 1035-1047.

Adamowski, K., and Labatiuk, C. (1987). "Estimation of flood frequencies by a nonparametric density procedure." *Hydrologic Frequency Modeling*, pp. 97-106.

Adamowski, K. (1996). "Nonparametric Estimation of Low-Flow Frequencies." *Journal of Hydraulic Engineering*, pp. 46-49.

Bowman, A. W. (1984). "An alternative method of cross-validation for the smoothing of density estimates." *Biometrika*, Vol. 71, pp. 353-360.

Breiman, L., Meisel, W. and Purcell, E. (1977). "Variable kernel estimates of multivariate densities." *Technometrics*, Vol. 19, No. 2, pp. 135-144.

Duin, R.P.W. (1976). "On the choice of smoothing parameters for parzen estimators of probability density functions." *IEEE Trans. Comput.*, C-25, pp. 1175-1179.

Geisser, S. (1975). "The predictive sample reuse method with a applications." *J. Amer. Statist Assoc.*, Vol. 70, pp. 320-328.

Good, I.J., and Gaskins, R.A. (1980). "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data." *Journal of the American Statistical Association*, Vol. 75, pp. 42-56.

Habbema, J.D.F., Hermans, J. and Broek, V. D. (1974). *A stepwise discrimination program using density estimation,* In G. Bruckman (Ed.). Physical verlag. Compstat Vienna, pp. 100-110.

Hall, P. (1983). "Large sample optimality of least squares cross-validation in density estimation." *Ann. Statist.*, Vol. 11, pp. 1156-1174.

Hall, P., and Marron, J.S. (1987). "Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation." Probability *Theory Rel. Fields*, Vol. 74, pp. 567-581.

Lall, U., Moon, Y.I., and Bosworth, K. (1993). "Kernel flood frequency estimators :bandwidth selection and kernel choice." *Water Resources Research*, Vol. 29, No. 4, pp. 1003-1015.

Moon, Young-Il, Lall, U., and Bosworth, K. (1993). "A comparison of tail probability estimators." *Journal of Hydrology*, Vol. 151, pp. 343-363.

Moon, Young-Il and Lall, U. (1994). "Kernel quantile function estimator for flood frequency analysis." *Water Resources Research*, Vol. 30, No. 11, pp. 3095-3103.

Moon, Young-Il and Lall, U. (1995). "Nonparametric flood frequency analysis by a kernel quantile function estimator." *European Geology Society*.

Moon, Young-Il. (1996). "Nonparametric

flood frequency analysis." *Journal of the Institute of Metropolitan Studies*, Vol. 22, No. 1, pp. 231-248.

Moon, Young-Il and Cha, Y.I., and Chun, S.Y. (1999). "The Estimate of Probability Precipitation by Frequency Analysis: Parametric and Nonparametric Methods." *Conference Proceedings of Korea Water Resources Association*, pp. 117-122.

Parzen, E. (1962). "On estimation of a probability density function and mode." *Ann. Math. Statist.*, Vol. 33, pp. 1065-1076.

Rosenblatt, M. (1956). "Remarks on some nonparametric estimates of a density function." *Ann. Math. Statist.*, Vol. 27, pp. 832-837.

Rudemo, M. (1982). "Empirical choice of histograms and kernel density estimators." *Scand J. Statist.*, Vol. 9, pp. 65-78.

Schuster, E.F., and Yakowitz, S. (1985). "Parametric/nonparametric mixture density estimation with application to flood-frequencyanalysis." *Water Resources Bulletin*, Vol. 21, No. 5, pp. 797-804.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, New York.

Stone, M. (1974). "Cross-validatory choice and assessment of statistical predictions." *J. Roy. Statist. Soc.*, B 36, pp. 111-147.

Stone, M. (1984). "An asymptotically optimal window selection rule for kernel density estimates." *Ann. Statistics*, Vol. 12, pp. 1285-1297.

Young-Il Moon, Department of Civil Engineering, University of Seoul, Chonnong -Dong, Tongdaemun-Gu, Seoul, 130-743, Korea.
(e-mail: ymoon@uoscc.uos.ac.kr)

(Received September 1, 1999; revised October 3, 1999; accepted November 4, 1999.)