L moment diagrams for censored observations

Antigoni Zafirakou-Koulouris and Richard M. Vogel
Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts

Scott M. Craig
Metcalf & Eddy Inc., Wakefield, Massachusetts

Joerg Habermeier
Winchester, Massachusetts

Abstract. Observed data sets containing values above or below the analytical threshold of measuring equipment are referred to as censored. Such data are frequently encountered in quality and quantity monitoring applications of water, soil, and air samples. Most of the previous literature on the statistical analysis of censored data relates to the problems of moment, parameter, and quantile estimation methods. Such estimation methods usually assume an underlying probability distribution. Few goodness-of-fit methods exist for censored data. We introduce L moment diagrams for the evaluation of the goodness of fit of alternative distributional hypotheses for left-censored data. Experiments with artificial censored data sets document the conditions under which L moment diagrams should be useful. Our approach, like *Hosking*'s [1995] approach for right censoring, derived L moment diagrams for left-censored observations from partial probability-weighted moments.

1. Introduction

In the area of environmental quality and quantity monitoring it is not uncommon to obtain censored sample readings such as data points that are below the detection limit of the available sensing equipment. In the fields of air, soil, and water quality analyses, left-censored data sets arise when contaminant concentrations less than the limit of measurement detection are reported as censored. In the field of hydrology, left-censored data sets arise because river discharges below some measurement threshold are often reported as zero. Such river discharges may have actually been zero or they may have been between zero and the measurement threshold, yet reported as zero [*Kroll and Stedinger*, 1996]. Sometimes it is actually advantageous to intentionally censor (or eliminate) observations in order to better understand the frequency and magnitude of flood [*Wang*, 1990a, b, 1996a] and drought events [*Durrans*, 1996]. Right-censored data sets can arise in hydrology when flood observations are reported as having occurred above some threshold [*Stedinger and Cohn*, 1986].

Censored data are categorized as either type I censoring, where the measurement threshold is fixed and the number of censored data points varies, or as type II censoring, where the number of censored data points is fixed and the implicit threshold varies [*David*, 1981]. Left-censored environmental data typically follow type I censoring because the censoring threshold is fixed by the measurement technology. Values for data below the measurement threshold are generally reported as "less than the detection limit," and data sets containing such points are referred to as (type I) left singly censored data. This study concentrates on type I left censoring.

Paper number 97WR03712.
0043-1397/98/97WR-03712$09.00

Most of the previous research on censored environmental data has introduced and compared various estimators of the mean, standard deviation, median, interquartile range [*Gilliom and Helsel*, 1986; *Helsel and Gilliom*, 1986; *El-Shaarawi*, 1989], and other quantiles [*Kroll and Stedinger*, 1996] of such data sets. *Helsel and Hirsch* [1992] and *Berthouex and Brown* [1994] summarize numerous estimation methods for use with censored data. While such estimation methods provide practical estimators for use with censored data, most methods are parametric, requiring an assumption of a specific probability distribution. Typically, the normal and lognormal distributions have been used to approximate the probability distribution of environmental data sets. *Berthouex and Brown* [1994] argue that traditional faith in these two models and the lack of a readily available goodness-of-fit test for the most appropriate underlying distribution are among the primary reasons for this choice.

The purpose of this study is to develop and test a methodology that enables the identification of a suitable probability distribution for use with censored data. Our approach is analogous to that of *Hosking* [1995], who introduced the application of L moment diagrams for use with right singly censored data, which often arise in life-testing and reliability applications in the field of manufacturing. Here we extend *Hosking*'s [1995] methodology for use with left singly censored data and perform experiments to evaluate the effectiveness of L moment diagrams.

2. L Moments and L Moment Diagrams

Several recent studies document that L moment diagrams are very useful for evaluating the goodness of fit of alternative probability distributions to complete (uncensored) data sets (*Hosking* [1990], *Chowdhury et al*. [1991], *Vogel and Fennessey* [1993], Hosking and Wallis [1997], and many others). Since their introduction by *Hosking* [1990], L moment diagrams have been used in nationwide assessments of the goodness of fit of various probability distributions to uncensored temperature [*Guttman*, 1996], precipitation [*Guttman et al*., 1993], and streamflow [*Vogel and Wilson* 1995] data sets.

Like ordinary product moments, L moments summarize the characteristics or shape of theoretical probability distributions and observed samples. Both moment types offer measures of distributional location (mean), scale (variance), skewness (shape), and kurtosis. L moments offer significant advantages over ordinary product moments, especially for environmental data sets, because of the following:

1. L moment ratio estimators of location, scale, and shape are nearly unbiased, regardless of the probability distribution from which the observations arise [*Hosking*, 1990].

2. L moment ratio estimators such as L-C_v, L-skewness, and L-kurtosis can exhibit lower bias than conventional product moment ratios, especially for highly skewed samples.

3. The L moment ratio estimators of L-C_v and L-skew do not have bounds which depend on sample size as do the ordinary product moment ratio estimators of C_v and skewness [*Kirby*, 1974].

4. L moment estimators are linear combinations of the observations and thus are less sensitive to the largest observations in a sample than product moment estimators, which square or cube the observations.

5. L moment ratio diagrams are particularly good at identifying the distributional properties of highly skewed data, whereas ordinary product moment diagrams are almost useless for this task [*Vogel and Fennessey*, 1993].

Recent references summarize the theory of L moments and the use of L moment diagrams for complete data sets, and the reader is referred to those studies (*Hosking* [1990], *Stedinger et al*. [1993], *Hosking and Wallis* [1997], and many others). The theoretical development that follows assumes a familiarity with the theory of L moments.

3. L Moments, Probability-Weighted Moments, and L Moment Diagrams for Complete Samples

Probability-weighted moments (PWMs) introduced by *Greenwood et al*. [1979] are linear functions of L moments. PWMs are defined as

$$\beta_r = E[x\{F(x)\}^r] \tag{1a}$$

which can be rewritten as

$$\beta_r = \int_0^1 x(F) F^r \, dF \tag{1b}$$

where $F = F(x)$ is the cumulative distribution function (CDF) for x, $x(F)$ is the inverse CDF of x evaluated at the probability F, and $r = 0, 1, 2, \ldots$, is a nonnegative integer. When $r = 0$, β_0 is equal to the mean of the distribution $\mu = E[x]$.

For any distribution the rth L moment λ_r is related to the rth PWM [*Hosking*, 1990] via

$$\lambda_{r+1} = \sum_{k=0}^{r} \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \tag{2}$$

For example, the first four L moments are related to the PWMs using

$$\lambda_1 = \beta_0 \tag{3a}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{3b}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{3c}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{3d}$$

Hosking [1990] defined the L moment ratios as follows:

$$\text{L-}C_v = \tau_2 = \lambda_2/\lambda_1 \tag{4a}$$

$$\text{L-skew} = \tau_3 = \lambda_3/\lambda_2 \tag{4b}$$

$$\text{L-kurtosis} = \tau_4 = \lambda_4/\lambda_2 \tag{4c}$$

L moment ratio diagrams compare sample estimates of the dimensionless L moment ratios in (4) with their theoretical counterparts. In the following sections we derive L moment diagrams for censored samples using the theory of PWMs.

4. Partial Probability-Weighted Moments for Censored Samples

Wang [1990a, b, 1996a] introduced the concept of partial probability-weighted moments (PPWMs) for the purpose of estimating the upper quantiles of flood flows when one's interest is in the right tail of the distribution and there is some benefit to censoring some of the smaller observations in the left tail. The initial idea of PPWMs was to remove the smaller observations from the process of fitting the distribution because such observations are of little interest in flood frequency analysis and such observations can be a nuisance to the fitting process. *Wang* [1990a, b] defined a left censoring PPWM as

$$\beta_r' = \int_c^1 X(F) F^r \, dF \tag{5}$$

where the lower limit $c = F(T)$ is the fraction of observations which are censored and T is the censoring threshold. The PPWM

$$\beta_r'' = \frac{1}{1 - c^{r+1}} \int_c^1 X(F) F^r \, dF \tag{6}$$

was introduced by *Wang* [1996a] to remain consistent with PWMs for complete samples and to simplify the mathematical derivations associated with PPWM sample estimators of distributional parameters. *Wang* [1990a,b] and *Kroll and Stedinger* [1996] used the PPWM in (5) and *Wang* [1996a] used the PPWM in (6) to derive censored quantile estimators for the generalized extreme value (GEV) and lognormal distributions. The procedure used by *Wang* [1990a, b, 1996a] to derive the PPWMs in (5) and (6) differs from the approach taken by *Hosking* [1995]. The PPWMs introduced here and by *Hosking* [1995] directly link the entire theory of L moments with PPWMs. The PPWMs introduced by *Wang* [1990a, b, 1996a] and *Kroll and Stedinger* [1996] are also easily related to L moments; however, we chose to use *Hosking*'s [1995] definition of PPWMs because he had already related them to L moments for the analogous problem of right censoring. One could relate the definition of PPWMs introduced here and by *Hosking* [1995] with the definitions introduced by *Wang* [1990a, b, 1996a] in (5) and (6), so in some sense all definitions of PPWMs are equivalent, except that the definition in (6) often leads to simplified mathematical derivations. Interestingly, both the work of *Hosking* [1995] and this study could have avoided the use of PPWMs by directly adapting L moment estimators for censored observations, similar to the approach taken by *Wang* [1996b].

4.1. PPWMs for Right Censoring

Hosking [1995] introduced two different PPWMs for use in constructing L moment diagrams for right-censored observations, though he did not use the terminology PPWM. In right censoring, the censored observations are greater than the measurement threshold. *Hosking*'s [1995] type A PPWM, β_r^A, is equivalent to the PWM of the uncensored observations. His type B PPWM, β_r^B, is equal to the PWM of the "completed sample," where the censored observations above the censoring threshold T are set equal to the censoring threshold. It is a well established fact that replacing the censored observations with a fixed value such as the measurement threshold leads to a significant bias in the resulting statistics such as the mean, the median, or a quantile. However, in this instance we are defining a theoretical statistic not a sample statistic; hence bias is not an issue.

The type A PPWM of a right-censored distribution is the ordinary PWM of a (complete) distribution with quantile function

$$y^A(u) = x(uc) \qquad 0 < u < 1 \tag{7}$$

where $y^A(u)$ denotes the quantile function of both the censored and uncensored observations and $c = F(T)$ is the (random) fraction of observations that are uncensored. Defining $\omega = uc$ leads to

$$y^A(u) = x(\omega) \qquad 0 < \omega < c \tag{8}$$

For right censoring, the censoring threshold $T = x(c)$ is greater than the uncensored observations. Using $\omega = cu$ and $d\omega = c(du)$, *Hosking* [1995] combines the definition of a PWM in (1b) with (8) to obtain

$$\beta_r^A = \int_0^1 u^r y^A(u)\,du = \int_0^c \left(\frac{\omega}{c}\right)^r x(\omega)\,\frac{d\omega}{c} = \frac{1}{c^{r+1}}\int_0^c u^r x(u)\,du \tag{9}$$

Hosking [1995] combines (9) with (2) to derive L moment diagrams for various distributions based on type A PPWMs for right censoring. Using a similar approach, *Hosking* [1995] derived type B PPWMs for right censoring by replacing the censored observations with $T = x(c)$, the measurement threshold. In that case the quantile function is given by

$$y^B(u) = \begin{cases} x(u) & 0 < u < c \\ x(c) & c \le u < 1 \end{cases} \tag{10}$$

and the type B PPWMs are given by

$$\beta_r^B = \int_0^c u^r x(u)\,du + \frac{1 - c^{r+1}}{r+1}\,x(c) \tag{11}$$

Hosking [1995] also derived a useful expression which related β_r^A and β_r^B:

$$\beta_r^B = c^{r+1}\beta_r^A + \frac{1 - c^{r+1}}{r+1}\,x(c) \tag{12}$$

4.2. PPWMs for Left Censoring

In this section we derive PPWMs for type I left censoring, following the same approach taken by *Hosking* [1995] for right censoring. We derive type A′ and type B′ PPWMs for left censoring, analogous to *Hosking*'s [1995] type A and type B PPWMs for right censoring.

The type A′ PPWM of a left-censored distribution is the ordinary PWM of a (complete) distribution with quantile function

$$y^{A'}(u) = x(\varphi) \qquad 0 < u < 1 \tag{13}$$

Multiplication of the inequality $0 < u < 1$ by $(1 - c)$ and adding c to each term leads to $c < [(1 - c)u + c] < 1$. Defining $\varphi = (1 - c)u + c$ leads to (10) with $c < \varphi < 1$, $u = (\varphi - c)/(1 - c)$, and $d\varphi/du = 1 - c$. Substitution into (1b) leads to the PPWM

$$\beta_r^{A'} = \int_0^1 u^r y^{A'}(u)\,du = \int_c^1 \left(\frac{\varphi - c}{1 - c}\right)^r x(\varphi)\,\frac{d\varphi}{1 - c} \tag{14a}$$

$$= \frac{1}{(1 - c)^{r+1}}\int_c^1 (u - c)^r x(u)\,du \tag{14b}$$

which is different from either of the PPWMs, β_r' or β_r'', in (5) and (6) introduced by *Wang* [1990a, b, 1996a] for left censoring.

Recall that the type B PPWM in (11) replaced the censored observations with the fixed threshold $x(c)$, above which measurements are unavailable. For left censoring, type B′ PPWMs may be derived by replacing the censored observations with the fixed threshold $x(c)$, below which measurements are unavailable. The corresponding quantile function is as follows:

$$y^{B'}(u) = \begin{cases} x(c) & 0 < u \le c \\ x(u) & c < u < 1 \end{cases} \tag{15}$$

The type B′ PPWM is obtained from substitution of (15) into (1b) leading to

$$\beta_r^{B'} = \int_0^c u^r x(c)\,du + \int_c^1 u^r x(u)\,du \tag{16a}$$

$$= x(c)\int_0^c u^r\,du + \int_c^1 u^r x(u)\,du \tag{16b}$$

$$= x(c)\left[\frac{u^{r+1}}{r+1}\right]_0^c + \int_c^1 u^r x(u)\,du \tag{16c}$$

$$= x(c)\,\frac{c^{r+1}}{r+1} + \int_c^1 u^r x(u)\,du \tag{16d}$$

Note that the integral in (16d) is identical to *Wang*'s [1990a, b] PPWM given in (5). Therefore our type B′ PPWM in (16d) only differs from Wang's PPWM by a constant term, so the two PPWM definitions are easily related to one another.

5. L Moment Ratio Diagrams for Censored Observations

Hosking [1990] and *Hosking and Wallis* [1997] introduce L moment ratio diagrams for the graphical evaluation of the goodness of fit of complete samples to various probability density functions (pdf's). *Hosking* [1995] extends the applica-

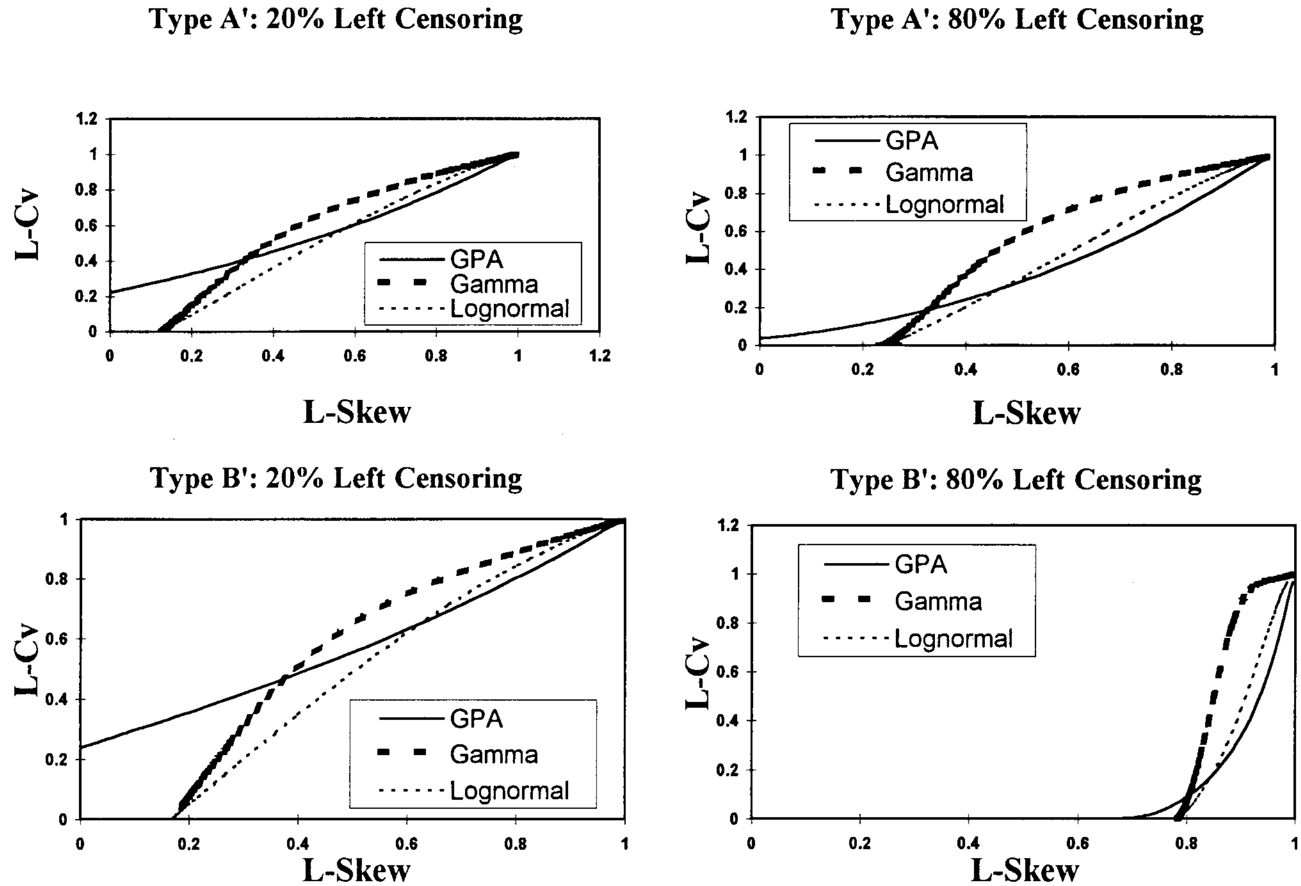

Figure 1. Theoretical relationships between L-C_v and L-skew for the two-parameter generalized Pareto (GPA), lognormal, and gamma distributions using type A′ and B′ partial probability-weighted moments (PPWMs) for left censoring for censoring levels of $c = 0.2$ and 0.8.

bility of L moment ratio diagrams to right-censored samples. In this section we summarize the utility of L moment ratio diagrams for summarizing left-censored observations.

5.1. Theoretical Relationships

In the background, an L moment ratio diagram contains the theoretical relationships between L-C_v and L-skew for one- and two-parameter pdf's or between L-kurtosis and L-skew for three-parameter pdf's. This section describes relations among theoretical L moment ratios for left-censored observations.

The theoretical relationships among L-C_v, L-skew, L-kurtosis, and censoring percentage c, using type A′ and B′ PPWMs for left censoring, can now be obtained from (3), (4), (14), and (16). Similarly, one could use (3), (4), (9), and (11) to obtain such relationships among L moment ratios and censoring levels using type A and B PPWMs for right censoring, as shown by *Hosking* [1995]. One need's only to substitute into those expressions the quantile function $x(p)$, corresponding to a particular theoretical pdf, to obtain such relationships. Figures 1 and 2 illustrate the theoretical relationships among L-C_v, L-skew, L-kurtosis, and censoring percentage c, using type A′ and B′ PPWMs for left censoring for the lognormal, gamma, and generalized Pareto (GPA) distributions. Figures 1 and 2 were obtained using numerical integration in (14) and (16). Quantile functions for these three distributions are given by *Stedinger et al.* [1993] and *Hosking and Wallis* [1997]. The two-parameter versions of these three pdf's were used to construct Figure 1 (lower bounds set to zero), and the three-parameter version of these pdf's was used to construct Figure 2.

Figure 1 illustrates that for the L-C_v versus L-skew relationships, regardless of the level of censoring, one is able to discriminate quite readily among the three populations considered. As expected for left censoring, increasing the level of censoring tends to increase the L-skewness. This effect is particularly pronounced using the type B′ PPWMs because the censored observations are replaced by the censoring threshold, tending to increase the leverage associated with the few remaining large observations. Still, for either type A′ or B′ PPWMs the various theoretical L moment ratio relationships tend to retain their relative positions among each other. Therefore our ability to discriminate among the two-parameter versions of these three pdf's using sample L-C_v versus L-skew diagrams is not expected to differ significantly between type A′ and B′ PPWMs, unless the sampling variability associated with L moment estimates differs significantly between A′ and B′ PPWMs.

Habermeier [1996] provides polynomial numerical approximations to the relationships among L-C_v, L-skew, and censoring level c, based on type A′ and B′ PPWMs for left censoring for the GPA distribution. Generalized mathematical software

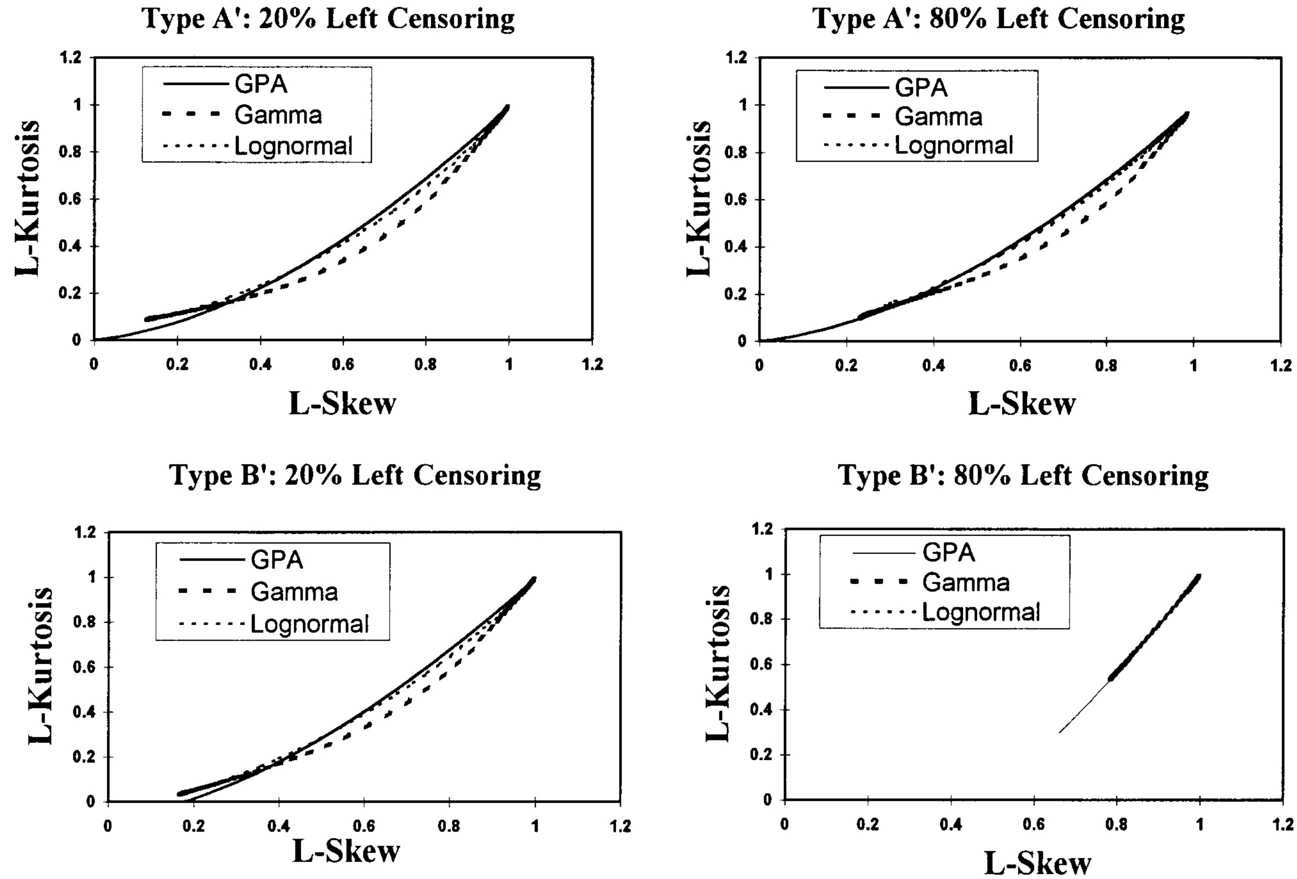

Figure 2. Theoretical relationships between L-kurtosis and L-skew for the three-parameter GPA, lognormal, and gamma distributions using type A′ and B′ PPWMs for left censoring for censoring levels of $c = 0.2$ and 0.8.

is now readily available to make implementation of numerical integration methods simple enough to dispense with the need to create such polynomial numerical approximations. In some instances, such as for unbounded distributions, it is possible that numerical integration will not yield theoretical relationships among the L moment ratios for censored data. In such cases one could compute theoretical L moments by generating a very long series of data and subsequently computing L moments of the generated data. As long as the data series is long enough, this approach may be more accurate and efficient than numerical integration for some unusual distributions. Experiments were also performed to confirm that all the theoretical relationships illustrated in Figures 1 and 2 for both A′ and B′ L moments are invariant to the choice of a location parameter.

5.2. Impact of the Level of Censoring on the Tail Behavior of a pdf

Figure 2 illustrates theoretical relationships between L-skew and L-kurtosis for the three-parameter versions of the gamma, GPA, and lognormal distributions. A three-parameter gamma distribution is usually referred to as a Pearson type III distribution. Figure 2 reveals that type A′ PPWMs lead to better discrimination among these three pdfs than type B′ PPWMs for high censoring levels. For high censoring levels, Figure 2 illustrates that the tail behavior of type B′ L moments for these three pdf's is roughly equivalent. Thus type B′ L moment ratio diagrams are relatively useless for discriminating among these pdf's at high censoring levels. *Hosking* [1995] obtained similar conclusions for type B L moment ratio diagrams for right censoring; however, he only examined a censoring threshold of $c = 0.5$.

Figures 3 and 4 illustrate the effect of the level of censoring on the tail behavior of a lognormal random variable. Figures 3 and 4 illustrate the relationships L-C_v versus L-skew and L-kurtosis versus L-skew, respectively, based on both type A′ and B′ PPWMs for left censoring, corresponding to the lognormal distribution over a wide range of censoring levels. Figures 3 and 4 show clearly that the tail behavior corresponding to type B′ L moments is much more sensitive to the level of censoring than the tail behavior for type A′ L moments. We conclude that type A and A′ L moment ratio diagrams are often equivalent and sometimes preferred to type B and B′ diagrams for the purpose of discriminating among the tail behavior of various pdf's.

6. Sample Estimators of Partial Probability-Weighted Moments and L Moments for Censored Observations

Hosking [1995] derived sample estimators of type A and B PPWMs for right-censored observations. Such estimators, termed b_r^A and b_r^B, are unbiased estimators of their theoretical counterparts, β_r^A and β_r^B, given in (9) and (11), respectively. In this section these results are reviewed and additional estima-

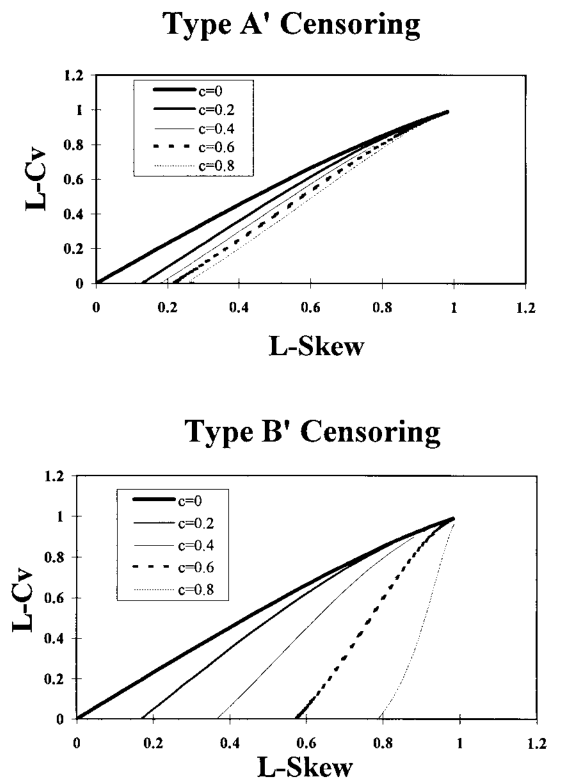

Figure 3. The effect of censoring level on the theoretical L-C_v versus L-skew relationship for a two-parameter lognormal variable.

tors for type A′ and B′ PPWMs and L moments for left censoring are derived. We introduce a standard notation for manipulation of the censored samples.

6.1. Ordered Observations

The order statistics of a complete sample of n observations are denoted by the following:

$$\underbrace{X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}}_{\substack{n \\ \text{observed}}} \tag{17}$$

Type I right censoring occurs when m of these values are observed ($m \le n$) and the remaining $n - m$ are censored above a known threshold T:

$$\underbrace{X_{1:n} \le X_{2:n} \le \cdots \le X_{m:n}}_{\substack{m \\ \text{censored}}}$$

$$\underbrace{\le T \le X_{m+1:n} \le \cdots \le X_{n-1:n} \le X_{n:n}}_{\substack{n-m \\ \text{censored}}} \tag{18}$$

Since the censoring threshold T is fixed in type I censoring, m is a random variable with a binomial distribution. Otherwise, type II censoring results, and T becomes the random variable, with m fixed.

Similarly, type I left censoring results when the observations below a fixed threshold T are censored:

$$\underbrace{X_{1:n} \le X_{2:n} \le \ \le X_{m-1:n}}_{\substack{n-k \\ \text{censored}}}$$

$$\underbrace{\le T \le X_{m:n} \le \cdots \le X_{n-1:n} \le X_{n:n}}_{\substack{k \\ \text{observed}}} \tag{19}$$

where the number of the censored values ($m - 1 = n - k$) is a random variable.

6.2. PPWM Estimators for Right Censoring

The following results are taken from *Hosking* [1995]. Type A PPWM and L moment estimators are simply PPWMs and L moments of the uncensored sample of m observations so that

$$b_r^A = \frac{1}{m} \sum_{j=1}^{m} \frac{(j-1)(j-2)\cdots(j-r)}{(m-1)(m-2)\cdots(m-r)} X_{j:n} \tag{20}$$

Landwehr et al. [1979] showed that (20) is an unbiased estimator of β_r^A. *Hosking and Wallis* [1995] show that unbiased estimators of PWMs are generally preferred over their biased alternatives. This is particularly true when constructing L moment diagrams.

Type B PPWM and L moment estimators are computed from the completed sample, where the $n - m$ censored values in (18) are replaced by the censoring threshold T, so

$$b_r^B = \frac{1}{n} \left\{ \sum_{j=1}^{m} \frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)} X_{j:n} \right.$$

$$\left. + \left(\sum_{j=m+1}^{n} \frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)} \right) T \right\} \tag{21}$$

Hosking [1995] used (12) to show that (20) and (21) are related by

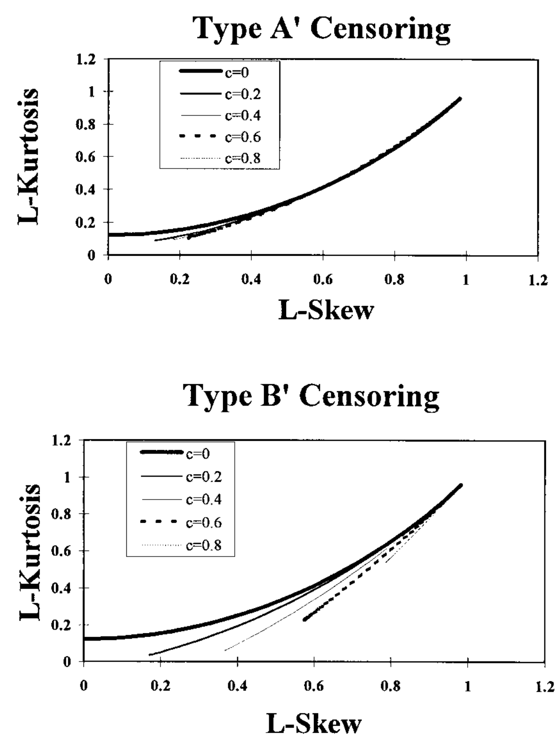

Figure 4. The effect of censoring level on the theoretical L-kurtosis versus L-skew relationship for a three-parameter lognormal variable.

Table 1. Information Content Associated With L Moment Diagrams in Figures 5 and 6

Experiment	Sample Size n	Number of Samples s	Censoring Level c	Information Content $I = ns\,(1 - c)$
1	15	100	0.8	300
2	15	100	0.2	1200
3	100	100	0.8	2000
4	100	100	0.2	8000

$$b_r^B = Zb_r^A + \frac{1 - Z}{r + 1} T \qquad Z = \frac{m(m - 1) \cdots (m - r)}{n(n - 1) \cdots (n - r)} \tag{22}$$

6.3. PPWM Estimators for Left Censoring

In this section we present unbiased estimators of type A′ and B′ PPWMs for left censoring. One can show that an unbiased estimator of type A′ PPWMs for left-censored observations is

$$b_r^{A'} = \frac{1}{k} \sum_{j=1}^{k} \frac{(j - 1)(j - 2) \cdots (j - r)}{(k - 1)(k - 2) \cdots (k - r)} X_{n-k+j:n} \qquad k = n - m + 1 \tag{23}$$

Similarly, an unbiased estimator of type B′ PPWMs for left-censored observations is

$$b_r^{B'} = \frac{1}{n} \left\{ \left(\sum_{j=1}^{n-k} \frac{(j - 1)(j - 2) \cdots (j - r)}{(n - 1)(n - 2) \cdots (n - r)} \right) T + \sum_{j=n-k+1}^{n} \frac{(j - 1)(j - 2) \cdots (j - r)}{(n - 1)(n - 2) \cdots (n - r)} X_{j:n} \right\} \tag{24}$$

Unbiased type A′ and B′ L moments for left censoring are obtained by substitution of $b_r^{A'}$ or $b_r^{B'}$ in place of β_r in (2) or (3).

6.4. Sampling Distribution of L Moment Estimators for Censored Data

The sampling distributions of L moments for complete samples apply to both right and left type I censored samples [*Hosking*, 1995]. The sample PPWM and L moment estimators introduced here are unbiased estimators of their corresponding population statistics. Asymptotically, they are normally distributed with asymptotic covariance that can be computed from theorem 3 introduced by *Hosking* [1990].

7. Experiments With L Moment Diagrams for Censored Observations

The real value of L moment diagrams is that sample L moment ratios can be directly compared with the theoretical curves depicted in Figures 1–4. L moment ratios are nearly unbiased; therefore ~50% of the sample L moment ratios are expected to lie above and 50% below the theoretical curves respectively.

7.1. Information Content of an L Moment Diagram

L moment diagrams compare the theoretical and observed tail behavior of pdf's. The larger the average sample size, the more samples, and the lower the censoring level, the more closely the sample L moments will follow the theoretical L moment relationship. We define the information content in the L moment diagram using

$$I = ns(1 - c) \tag{25}$$

where n is the average sample size, s is the number of samples, and c is the censoring level. The information content I is equal to the overall number of uncensored observations in the experiment. In the experiments in sections 7.2 and 7.3 we explore the information content of L moment diagrams for censored samples.

7.2. Experimental Design

Experiments were designed to evaluate the utility of L moment diagrams for assessing the goodness of fit of various distributions to artificial censored data sets. One hundred ($s = 100$) samples of synthetic gamma and generalized Pareto data sets of length $n = 15$ and 100 were generated with censoring levels of $c = 0.2$ and 0.8. Table 1 summarizes the information content associated with each of these experiments ranging from $I = 300$ to $I = 8000$. The conventional coefficient of variation for each sample was generated as a uniformly distributed random variable between 0.5 and 4.5. Since type A and A′ L moment relationships are usually preferred over type B and B′, we only evaluate type A′ L moment diagrams here. *Habermeier* [1996] and *Craig* [1997] summarize similar evaluations based on type B′ L moment diagrams for left censoring, and *Hosking* [1995] summarizes evaluations based on type A and B L moment diagrams for right censoring.

7.3. Results and Discussion

Figures 5 and 6 illustrate the results of four experiments, corresponding to information contents of $I = 300$, 1200, 2000, and 8000. In Figure 5 the observations were generated from a two-parameter generalized Pareto (GPA) distribution (lower bound set to zero), and in Figure 6 the observations were generated from a two-parameter gamma distribution. In both Figures 5 and 6 an information content of 300 is not nearly sufficient to distinguish the tail behavior of the observations. As expected, our ability to distinguish the tail behavior of the distributions improves as the information content increases. *Habermeier* [1996] and *Craig* [1997] performed numerous other experiments using data from other distributions (lognormal), type B′ PPWMs, and using L-kurtosis versus L-skew diagrams. Those evaluations led to the similar conclusions shown in Figures 5 and 6; hence they are not reported here. Synthetic samples with record lengths in excess of 100,000 were also generated to confirm that asymptotically, the L moment ratio estimators introduced here are unbiased.

8. Conclusions

Previous research (*Hosking* [1990], *Vogel and Fennessey* [1993], *Chowdhury et al*. [1991], and others) documents the value of L moment diagrams for evaluating the goodness of fit of alternative probability density functions to complete samples. Nationwide evaluations of temperature [*Guttman*, 1996], precipitation [*Guttman et al*., 1993], and streamflow [*Vogel and Wilson*, 1995] databases have revealed how powerful L moment

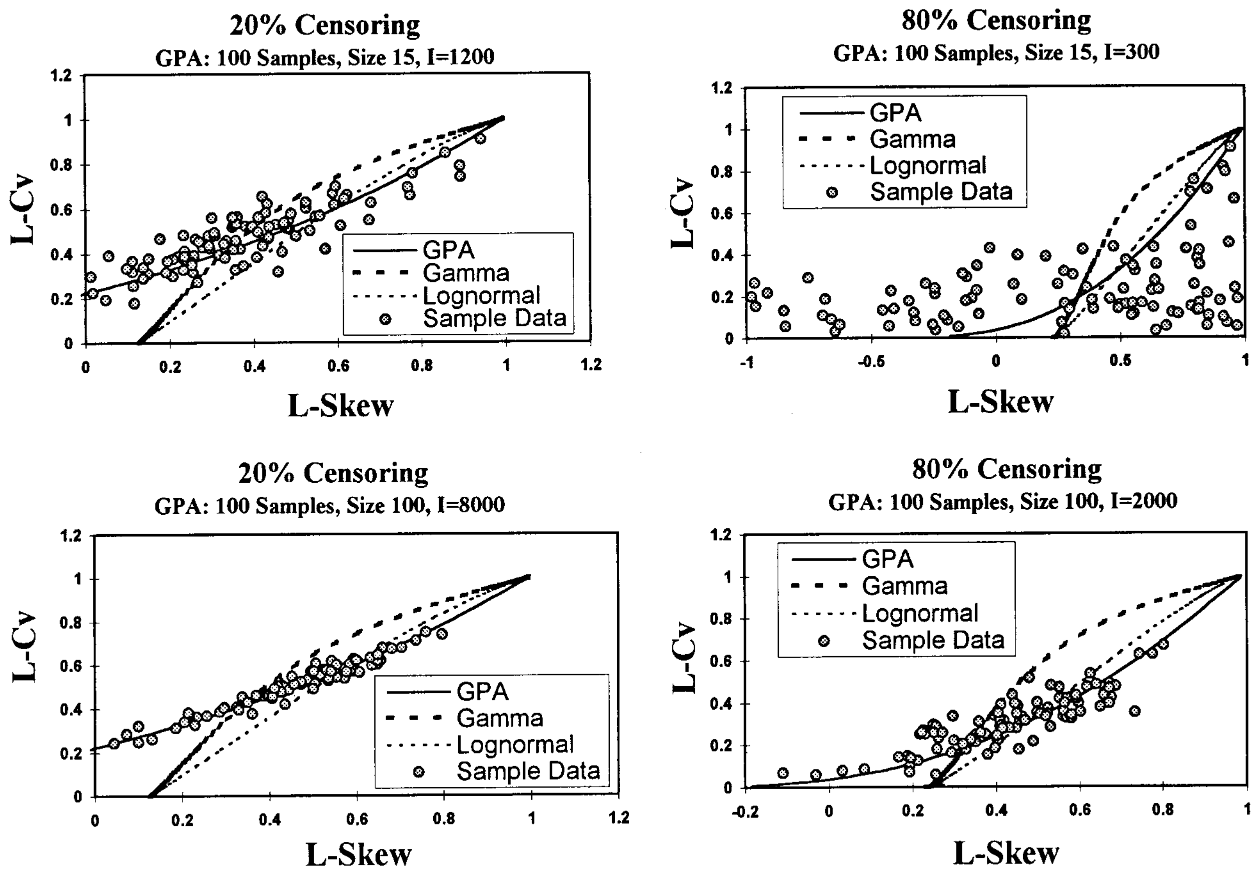

Figure 5. L moment diagrams based on 100 samples of synthetic (two-parameter) generalized Pareto data, using type A′PWMs for left censoring.

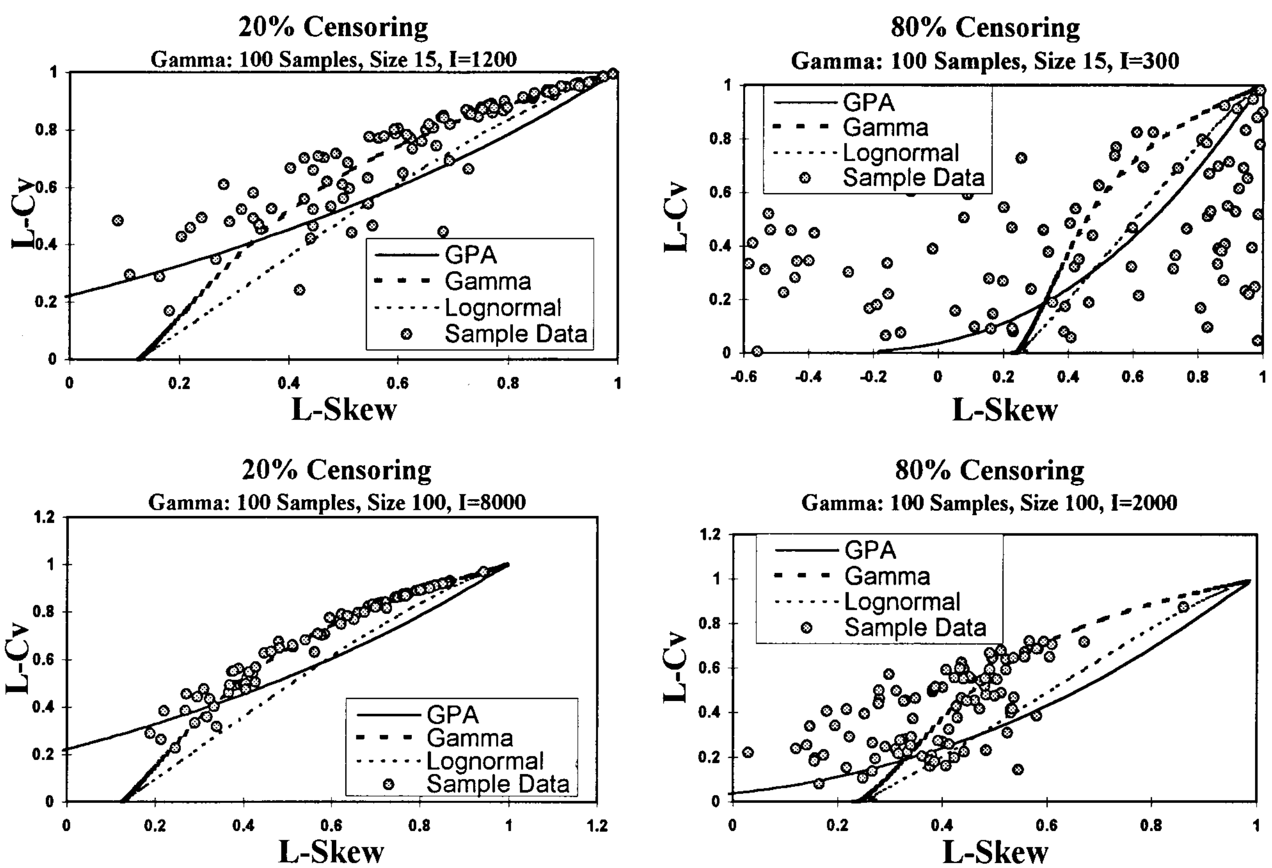

Figure 6. L moment diagrams based on 100 samples of synthetic gamma data, using type A′ PPWMs for left censoring.

diagrams can be for discriminating among various distributional alternatives. *Hosking* [1995] showed that complete-sample techniques based on L moments can be extended for use with right-censored observations. This study has extended those results to include left-censored observations. It is hoped that the methods introduced here provide a rigorous and useful approach for examining databases of left-censored observations such as water, air, and soil quality and quantity measurements.

This study has focused on the application of L moment diagrams as graphical aids in determining the visual goodness of fit of alternative pdf's to observed samples. We have shown that such evaluations depend significantly on the information content of the diagram $I = ns\,(1 - c)$, where n is the average sample size, s is the number of samples, and c is the fraction of the observations which are censored. Our evaluations support those of recent nationwide empirical evaluations which indicate that the information content of an L moment diagram needs to be in the thousands to enable one to (confidently) discern the tail behavior of similar distributional alternatives.

The equations introduced here for left censoring and by *Hosking* [1995] for right censoring may also be used to estimate parameters and quantiles of a distribution from observations. *Hosking* [1995] has shown that L moment estimators of a distribution's parameters under right censoring can be competitive with computationally more complex methods such as maximum likelihood. *Hosking* [1995] documents that a partial probability-weighted moment estimator of the parameters of a reverse Gumbel distribution is only slightly less efficient than maximum likelihood estimator (MLE) under type I right censoring and that these estimators are almost equivalent for censoring levels above $c = 0.5$ and for larger samples. We expect similar results under left censoring, using the PPWM and L moment estimators introduced here. *Kroll and Stedinger* [1996] document that a PPWM estimator based on (5) (in log space) is competitive with both a log-probability plot estimator [*Gilliom and Helsel*, 1986; *Helsel and Gilliom*, 1986] and an MLE for estimating moments and quantiles of left-censored lognormal observations, particularly for smaller sample sizes and censoring levels.

The evaluations of L moment diagrams introduced here assume that the censoring threshold c is known. In practice, the censoring threshold c will need to be estimated and may vary from sample to sample. Future research relating to the use of L moment diagrams for censored observations should address the fact that c is a random variable.

Acknowledgments. The authors gratefully acknowledge the constructive review comments and suggestions by Q. J. Wang, Ned Guttman, Jery R. Stedinger, and an anonymous reviewer.

References

Berthouex, P. M., and L. C. Brown, *Statistics for Environment Engineers*, Lewis, Boca Raton, Fla., 1994.

Chowdhury, J. U., J. R. Stedinger, and L.-H. Lu, Goodness-of-fit tests for regional generalized extreme value flood distributions, *Water Resour. Res.*, *27*(7), 1765–1776, 1991.

Craig, S. M., Application of L-moments in frequency analysis of censored data, M. S. thesis, Tufts Univ., Medford, Mass., Feb. 1997.

David, H. A., *Order Statistics*, John Wiley, New York, 1981.

Durrans, S. R., Low-flow analysis with a conditional Weibull tail model, *Water Resour. Res.*, *32*(6), 1749–1760, 1996.

El-Shaarawi, A. H., Inferences about the mean from censored water quality data, *Water Resour. Res.*, *25*(4), 685–690, 1989.

Gilliom, R. J., and D. R. Helsel, Estimation of distributional parameters for censored trace level water quality data, 1, Estimation techniques, *Water Resour. Res.*, *22*(2), 135–146, 1986.

Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis, Probability weighted moments: Definitions and relation to parameters of several distributions expressible in inverse form, *Water Resour. Res.*, *15*(6), 1049–1054, 1979.

Guttman, N. B., Statistical characteristics of US historical climatology network temperature distributions, *Clim. Res.*, *6*(1), 33–43, 1996.

Guttman, N. B., J. R. M. Hosking, and J. R. Wallis, Regional precipitation quantile values for the continental United States computed from L-moments, *J. Clim..*, *6*(12), 2326–2340, 1993.

Habermeier, J., *An evaluation of L-moment diagrams for left censored data*, M.S. thesis, Tufts Univ., Medford, Mass., Nov. 1996.

Helsel, D. R., and R. J. Gilliom, Estimation of distributional parameters for censored trace level water quality data, 2, Validation techniques, *Water Resour. Res.*, *22*(2), 147–155, 1986.

Helsel, D. R., and R. M. Hirsch, *Statistical Methods in Water Resources*, Elsevier, New York, 1992.

Hosking, J. R. M., L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc.*, *Ser. B*, *52*, 105–124, 1990.

Hosking, J. R. M., The use of L-moments in the analysis of censored data, in *Recent Advances in Life-Testing and Reliability*, edited by N. Balakrishnan, chap. 29, pp. 546–560, CRC Press, Boca Raton, Fla., 1995.

Hosking, J. R. M., and J. R. Wallis, A comparison of unbiased and plotting-position estimators of L moments, *Water Resour. Res.*, *31*(8), 2019–2025, 1995.

Hosking, J. R. M., and J. R. Wallis, *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge Univ. Press, New York, 1997.

Kirby, W., Algebraic boundedness of sample statistics, *Water Resour. Res.*, *10*(2), 220–222, 1974.

Kroll, C. N., and J. R. Stedinger, Estimation of moments and quantiles using censored data, *Water Resour. Res.*, *32*(4), 1005–1012, 1996.

Landwehr, J. M., N. C. Matalas, and J. R. Wallis, Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, *15*(5), 1055–1064, 1979.

Stedinger, J. R., and T. A. Cohn, Flood frequency analysis with historical and paleoflood information, *Water Resour. Res.*, *22*(5), 785–793, 1986.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Geourgiou, Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. A. Maidment, chap. 18, McGraw-Hill, New York, 1993.

Vogel, R. M., and N. M. Fennessey, L-moment diagrams should replace product moment diagrams, *Water Resour. Res.*, *29*(6), 1745–1752, 1993.

Vogel, R. M., and I. Wilson, Probability distribution of annual maximum, mean, and minimum streamflows in the United States, *J. Hydrol. Engin.*, *1*(2), 69–76, 1995.

Wang, Q. J., Estimation of the GEV distribution from censored samples by method of partial probability weighted moments, *J. Hydrol.*, *120*, 103–114, 1990a.

Wang, Q. J., Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution, *J. Hydrol.*, *120*, 115–124, 1990b.

Wang, Q. J., Using partial probability weighted moments to fit the extreme value distributions to censored samples, *Water Resour. Res.*, *32*(6), 1767–1771, 1996a.

Wang, Q. J., Direct sample estimators of L moments, *Water Resour. Res.*, *32*(12), 3617–3619, 1996b.

S. M. Craig, Metcalf and Eddy Inc., 30 Harvard Mill Square, P.O. Box 4071, Wakefield, MA 01880–5371.

J. Habermeier, 50 Dunster Lane, Number 2, Winchester, MA 01890.

R. M. Vogel and A. Zafirakou-Koulouris, Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155. (e-mail: rvogel@tufts.edu)

(Received November 17, 1997; revised December 12, 1997; accepted December 16, 1997.)