ESTIMATION OF HARMONIC MEAN OF A LOGNORMAL VARIABLE

By James F. Limbrunner,[1] Richard M. Vogel,[2] and Linfield C. Brown,[3] Members, ASCE

ABSTRACT: The harmonic mean has numerous engineering applications including characterization of the large-scale effective permeability in layered porous media, characterization of petrochemical properties of heterogeneous media, and the design of declining rate filter beds. The U.S. EPA also recommends the use of the harmonic mean daily streamflow as a design streamflow for the protection of human health against lifetime exposure to suspected carcinogens. The sampling properties of various estimators of the harmonic mean are derived and compared for observations arising from a lognormal distribution. Previous applications have recommended the use of a moment estimator of the harmonic mean. We document that the moment estimator of the harmonic mean exhibits significant upward bias and large root-mean-square error, particularly for large skews. A maximum likelihood estimator of the harmonic mean is generally preferred because it is nearly unbiased and can provide dramatic reductions in the root-mean-square error, compared with the moment estimator. In addition, a maximum likelihood estimator of the generalized mean (or p-norm) of a lognormal distribution is introduced.

INTRODUCTION

There are numerous measures of central tendency for a random variable including the median, mode, and the harmonic, geometric, and arithmetic means. Each of these measures is appropriate for different situations. For example, when estimating the long-term expectation of a random variable, the arithmetic mean is a natural choice. This study concentrates on problems in which the harmonic mean is a natural choice as a measure of central tendency. We begin by describing the general properties of various measures of the mean, followed by an analysis of the sampling properties of the harmonic mean for a lognormal variable. We end by discussing the implications of our results in the context of the two general application areas: (1) design streamflow estimation for wasteload allocation; and (2) estimation of effective petrochemical and geophysical properties of a heterogeneous system of porous media.

Harmonic, Arithmetic, and Geometric Means

The harmonic mean is a measure of central tendency of a random variable X and is defined as the reciprocal of the expected value of the reciprocal of the random variable X

$$H = \frac{1}{E\left[\frac{1}{X}\right]} = \left[\int_{-\infty}^{\infty} \frac{1}{x} f(x)\, dx\right]^{-1} \tag{1}$$

In some cases a distribution's quantile function is easier to integrate than its probability density function, and Kendall et al. (1977, section 3.1) provide a useful alternate definition of H

$$H = \left[\int_0^1 (x_p)^{-1}\, dp\right]^{-1} \tag{2}$$

where x_p = quantile function of the random variable X. The harmonic mean does not exist if a random variable can take on zero values. Interestingly, Burk (1985) has shown the harmonic mean of a random variable is always less than its geometric mean G, which is always less than its arithmetic mean μ, regardless of the probability distribution from which the variable arises

$$H \le G \le \mu \tag{3}$$

Similarly, H is always greater than the smallest observation. The arithmetic mean is defined as the expected value of the random variable

$$\mu = E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx \tag{4}$$

and the geometric mean is given by

$$G = \left[\int_0^{\infty} x^{1/n} f(x)\, dx\right]^n \tag{5}$$

Landwehr (1978) describes the properties of the geometric mean. More generally, one can define the power mean, p-norm, or generalized mean

$$\mu_p = [E[x^p]]^{1/p} \tag{6}$$

which reduces to the harmonic, geometric, and arithmetic means for $p = -1$, $p \to 0$, and $p = 1$, respectively. Jensen (1998), and others cited therein, discuss the general properties of moment estimators of the generalized mean in (6) and provide citations to other literature dealing with applications in which p takes on values other than $p = -1$, $p \to 0$, and $p = 1$.

Properties of Harmonic Mean of Lognormal Distribution

This study concentrates on the properties of the harmonic mean H for a two-parameter lognormal (LN2) random variable. We concentrate on the LN2 distribution because this is the assumed distribution in the applications discussed later. A comprehensive treatment of the lognormal distribution is given by Aitchison and Brown (1969) and Crow and Shimizu (1988). A comprehensive comparison of quantile estimators for the two- and three-parameter lognormal distribution is given by Stedinger (1980). The probability density function (PDF) for the LN2 distribution is

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu_y}{\sigma_y}\right)^2\right] \tag{7a}$$

for $x > 0$ with $Y = \ln(X)$ and μ_y and σ_y^2 equal to the mean and variance of Y, respectively. The relations between the mean and variance in real and log space are

[1]Engr., Haley & Aldrich, Inc., 340 Granite St., 3rd Floor, Manchester, NH 03102.

[2]Prof., Dept. of Civ. and Envir. Engrg., Tufts Univ., Medford, MA 02155. E-mail: rvogel@tufts.edu

[3]Prof., Dept. of Civ. and Envir. Engrg., Tufts Univ., Medford, MA.

Note. Discussion open until June 1, 2000. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on February 1, 1998. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 5, No. 1, January, 2000. ©ASCE, ISSN 1084-0699/00/0001-0059–0066/$8.00 + $.50 per page. Paper No. 17577.

$$\mu = \exp\left(\mu_y + \frac{\sigma_Y^2}{2}\right) \tag{7b}$$

$$\sigma^2 = \mu^2[\exp(\sigma_Y^2) - 1] \tag{7c}$$

where μ and σ^2 denote the mean and variance of X. Aitchison and Brown (1969) show that the harmonic mean of a lognormal random variable is

$$H = \exp\left(\mu_y - \frac{1}{2}\sigma_y^2\right) \tag{8}$$

The harmonic mean of an LN2 population can differ dramatically from its arithmetic mean. A plot of the ratio of the harmonic mean H to the arithmetic mean μ, as a function of C_v, is illustrated in Fig. 1, which is based on the relation

$$\frac{H}{\mu} = \frac{1}{1 + C_v^2} \tag{9}$$

Note that the harmonic mean H is only a small fraction of the arithmetic mean μ for highly skewed samples. The harmonic mean of an LN2 population can also differ substantially from its median; this is also illustrated in Fig. 1. For an LN2 population, the mean and median are $\exp(\mu_y + \sigma_y^2/2)$ and $\exp(\mu_y)$, respectively. Gomez-Hernandez and Gorelick (1989) document that the ratio of the geometric mean to the harmonic mean of an LN2 variable is

$$\frac{G}{H} = \sqrt{1 + C_v^2} \tag{10}$$

Rossman (1990) combines (9) and (10) to show that the harmonic, geometric, and arithmetic mean of an LN2 variable are related by

$$H = \frac{G^2}{\mu} \tag{11}$$

ESTIMATION OF HARMONIC MEAN

The most common and natural approach to estimation of the harmonic mean is to use the moment estimator

$$\hat{H}_1 = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{12}$$

The estimator $\hat{H}_1$ is a nonparametric estimator because its derivation does not depend on the form of the PDF for x. Analogous to product moment estimators that are sensitive to very large observations, far away from the mean, the estimator $\hat{H}_1$ is highly sensitive to small observations. Rossman (1990), EPA (1991), and Martin and Ruhl (1993), have recommended the estimator $\hat{H}_1$ to compute the harmonic mean daily streamflow. Muskat (1937), Gomez-Hernandez and Gorelick (1989), Ababou and Wood (1990), Jensen et al. (1997), and Jensen (1998) suggested $\hat{H}_1$ to characterize the large-scale effective permeability in layered porous media. Because we are unaware of any literature that compares the sampling properties of alternate estimators of the harmonic mean, and previous experience suggests that estimators sensitive to only a few observations may perform poorly (Vogel and Fennessey 1993), we investigate alternate estimators.

Rossman (1990) introduced another estimator based on the relationship between H, G, and μ in (11). The sample estimator introduced by Rossman (1990) is obtained by substitution of sample moment estimates of G and μ into (11), leading to

$$\hat{H}_2 = \frac{\left[\left(\prod_{i=1}^{n} x_i\right)^{1/n}\right]^2}{\frac{1}{n}\sum_{i=1}^{n} x_i} \tag{13}$$

The estimate of G in the numerator of (13) is often numerically unstable, hence, a preferred and mathematically identical version for the LN2 distribution is

$$\hat{H}_2 = \frac{\left[\exp\left(\frac{1}{n}\left\{\sum_{i=1}^{n} \ln(x_i)\right\}\right)\right]^2}{\frac{1}{n}\sum_{i=1}^{n} x_i} \tag{14}$$

Using actual daily streamflow records for 60 gauging stations in the United States, Rossman (1990) found $\hat{H}_2$ results in estimates nearly identical to $\hat{H}_1$.

Maximum likelihood estimates (MLE) are often preferred to alternate estimators, particularly in large-sample problems arising from known distributions, because asymptotically, they have minimum-mean-square error (MMSE) among all com-

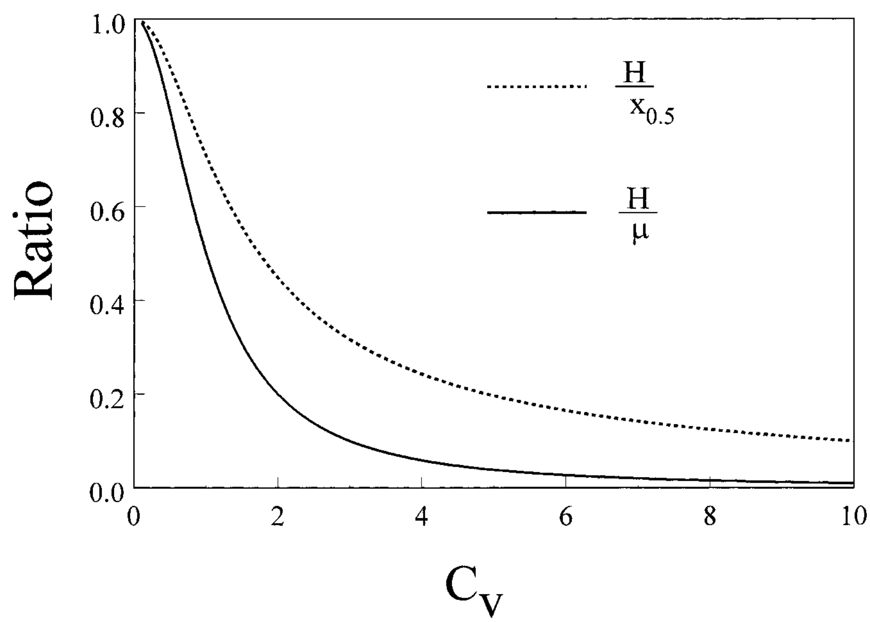

FIG. 1. Ratio of Harmonic Mean to Arithmetic Mean and Median as Function of C_v for LN2 Distribution

peting estimators. An MLE for H is obtained by substitution of the maximum likelihood estimates for μ_Y and σ_Y into (8), yielding

$$\hat{H}_3 = \exp\left(\hat{\mu}_y - \frac{1}{2}\hat{\sigma}_y^2\right) \qquad (15a)$$

where

$$\hat{\mu}_y = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \qquad (15b)$$

$$\hat{\sigma}_y^2 = \frac{1}{n}\sum_{i=1}^{n} (\ln x_i - \hat{\mu}_y)^2 \qquad (15c)$$

Aitchison and Brown (1969) and Crow and Shimizu (1988) summarize numerous other estimators of mean and variance of an LN2 distribution including the method of moments, a graphical method based on a probability plot, and Finney's (1941) uniformly minimum variance unbiased (UMVU) estimator. Stedinger (1980) also documents the sampling properties of numerous estimators of quantiles of both the two- and three-parameter lognormal distribution. He showed that the maximum likelihood method is generally best among other methods compared, for fitting the LN2 distribution to samples of 25 or more. Crow and Shimizu (1988, Chapter 2) compare various estimators of the general satistic $\theta = \exp(a\mu_y + b\sigma_y^2)$ for LN2 sampling. Note that $\theta = H$ when $a = 1$ and $b = -1/2$. Crow and Shimizu document the sampling properties of the UMVU estimator, the MLE, and an MMSE estimator. These three estimators have mean square errors that are equal to $O(1/n)$, so that for large n, there is very little difference between the estimators. Crow and Shimizu recommended the MLE because it performs similarly to the other estimators for large n, because it cannot generate negative values for θ and because it is much simpler to compute than either the UMVU or the MMSE. For these reasons, we do not consider either the UMVU or the MMSE here.

SAMPLING PROPERTIES OF HARMONIC MEAN ESTIMATORS

In this section, analytical and Monte Carlo comparisons are provided of the sampling properties of the various estimators of the harmonic mean introduced in the previous section. We concentrate on a comparison of the estimators $\hat{H}_1$ and $\hat{H}_3$ because it was found that the sampling properties of $\hat{H}_1$ and $\hat{H}_2$ are almost identical. Following this section, we discuss the implications of these results in the context of two application areas.

Vali (1943) derived the sampling distribution of $\hat{H}_1$ when samples are drawn from the normal and Pearson-type III distributions. More recently, Jensen et al. (1997) and Jensen (1998) approximated the bias and variance of $\hat{H}_1$ and an estimator of the generalized mean (p-norm), respectively, when samples are drawn from an LN2 distribution. Their derivations assume that the quantity $\sum_{i=1}^{n} x_i^p$ for $-1 \le p \le 1$ is lognormal.

In this section, we derive first-order estimates of the bias and root-mean-square error (RMSE) of estimates of $\hat{H}_1$ and $\hat{H}_3$ using first-order methods described elsewhere (Benjamin and Cornell 1970). The first-order approximations are derived by first approximating each estimator using a Taylor series and then deriving the expectation and variance of each estimator based on the initial terms in the Taylor series expansion. Experience has shown that such first-order approximations often provide good approximations to the sampling properties of complex estimators.

Bias and RMSE are defined as follows:

$$\%\ \text{bias}[\hat{H}] = \frac{E[\hat{H}] - H}{H}\cdot 100 \qquad (16a)$$

$$\%\ \text{RMSE}[\hat{H}] = \frac{[\text{bias}[\hat{H}]^2 + \text{var}[\hat{H}]]^{1/2}}{H}\cdot 100 \qquad (16b)$$

Because first-order methods are quite standard, we do not provide our derivations here; however, all analytical approximations are compared later with Monte Carlo experimental results.

First-Order Approximation to Sampling Properties of $\hat{H}_1$

A first-order approximation to the bias and variance of $\hat{H}_1$ are

$$\text{bias}[\hat{H}_1] = \frac{H\cdot C_v^2}{n} \qquad (17a)$$

$$\text{var}[\hat{H}_1] = \frac{H^2\cdot C_v^2}{n} \qquad (17b)$$

Jensen et al. (1997) obtain identical results to (17a); however, they obtain

$$\text{var}[\hat{H}_1] = H^2(1 + C_v^2/n)C_v^2/n \qquad (17c)$$

which is always greater than (17b).

First-Order Approximation to Sampling Properties of $\hat{H}_3$

A first-order approximation to the bias and variance of $\hat{H}_3$ are

$$\text{bias}[\hat{H}_3] = \frac{H\cdot\ln(1 + C_v^2)}{2n}\left[1 + \frac{\ln(1 + C_v^2)}{2}\right] \qquad (18a)$$

$$\text{var}[\hat{H}_3] = \frac{H^2\cdot\ln(1 + C_v^2)}{n}\left[1 + \frac{\ln(1 + C_v^2)}{4}\right] \qquad (18b)$$

Monte Carlo Experiments

To evaluate the first-order approximations in (17) and (18), and to investigate the sampling distributions of the three harmonic mean estimators $\hat{H}_1$, $\hat{H}_2$, and $\hat{H}_3$, 20,000 replicate LN2 samples each of size n = 100 and 1,000 were generated for various values of C_v. The three harmonic mean estimators were applied to each sample and compared with the true value of the harmonic mean.

Results

Our first results in Fig. 2 compare the analytic formulas for the standard deviation of $\hat{H}_1$ developed here [(17b)], and by Jensen et al. (1998) [(17c)], with the results of our Monte Carlo experiments. Fig. 2 reports % Stdev($\hat{H}_1$) defined as $100\sqrt{\text{var}(\hat{H}_1)}/H$. Fig. 2 documents that the analytical approximations break down for values of C_v in excess of about 5. Unfortunately, when values of C_v are that large, sample estimates of C_v are remarkably downward biased, even for very large samples; so in practice, one never knows if actual samples exhibit values of $C_v > 5$, unless one employs the theory of L-moments (Vogel and Fennessey 1993). It is clear from Fig. 2 that (17b) provides a better approximation than (17c) to the variance of $\hat{H}_1$, particularly for small samples; however, both expressions are nearly equivalent for $C_v < 5$.

Figs. 3 and 4 compare the bias and variance of the estimators $\hat{H}_1$ and $\hat{H}_3$, using solid lines to depict the analytic formulas [17(a,b) and 18(a,b)] and data points to illustrate the results of the Monte Carlo experiments for the cases C_v = 0.1, 1, 5, 10, and 20. Figs. 3 and 4 document that (18a) and (18b) provide an excellent representation of the bias and RMSE of $\hat{H}_3$ for all values of n and C_v considered. Figs. 3 and 4 also

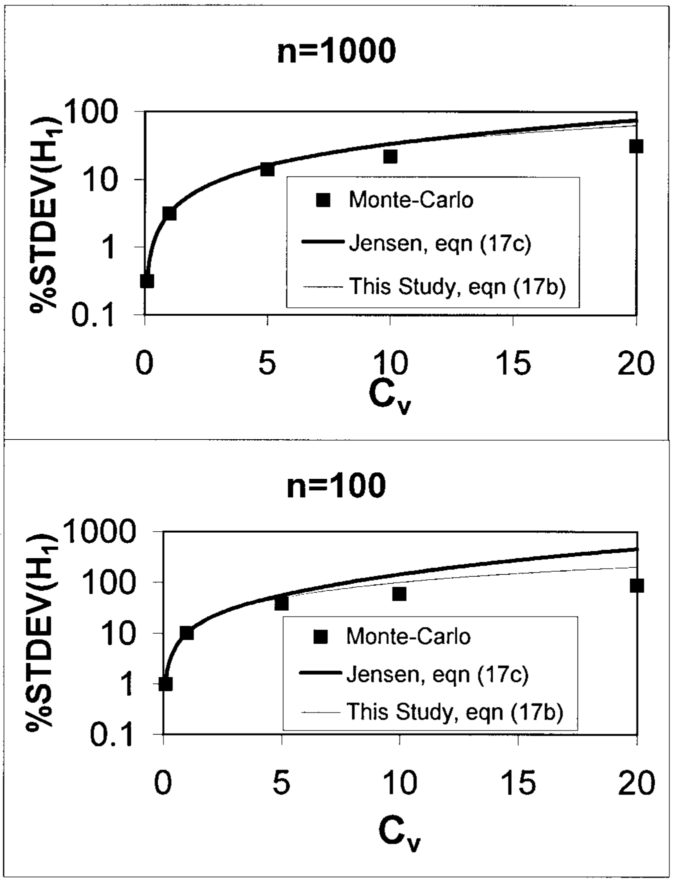

FIG. 2. Comparison of Standard Deviations of $\hat{H}_1$ Using Analytic Eqs. (17*b*) and (17*c*) and Monte Carlo Experiments

show that the analytic first-order approximations for the sampling properties of $\hat{H}_3$ are much more accurate than the analytic first-order approximation to the sampling properties of $\hat{H}_1$.

Most importantly, Figs. 3 and 4 document the clear advantage of $\hat{H}_3$ over $\hat{H}_1$ in terms of both bias and RMSE. Compared with $\hat{H}_1$, $\hat{H}_3$ is nearly unbiased, even for small samples and large values of C_v. Fig. 5 illustrates the efficiency gains associated with $\hat{H}_3$ over $\hat{H}_1$ computed as RMSE $(\hat{H}_1)$/RMSE$(\hat{H}_3)$ on the basis of the Monte Carlo experiments. Here, values of efficiency above unity are a measure of the improvement in the overall precision of $\hat{H}_3$ relative to $\hat{H}_1$. The relative improvement increases as n and C_v increase. It can be seen that for populations with large coefficients of variation, the MLE leads to dramatic improvements in terms of both bias and RMSE when compared with either $\hat{H}_1$ or $\hat{H}_2$. Because it was found that the sampling properties of $\hat{H}_2$ are nearly identical to those of $\hat{H}_1$, we elected to omit the results for $\hat{H}_2$ from our comparisons in Figs. 2–5.

APPLICATIONS OF HARMONIC MEAN TO DETERMINATION OF WASTELOAD ALLOCATIONS

The United States Federal Clean Water Act requires establishment of total maximum daily loading (TMDL) of contaminants for "water quality limited stream segments," such that loading in excess of this limit will risk violation of water quality standards (EPA 1991). The TMDL for suspected human carcinogens is established by first calculating the human health reference ambient concentration for a particular contaminant. Once this allowable concentration is established, it is combined with an estimate of streamflow available for contaminant dilution to determine the allowable TMDL for that stream segment. Because the risk of contracting cancer is estimated from exposure concentrations over an entire lifespan, the average

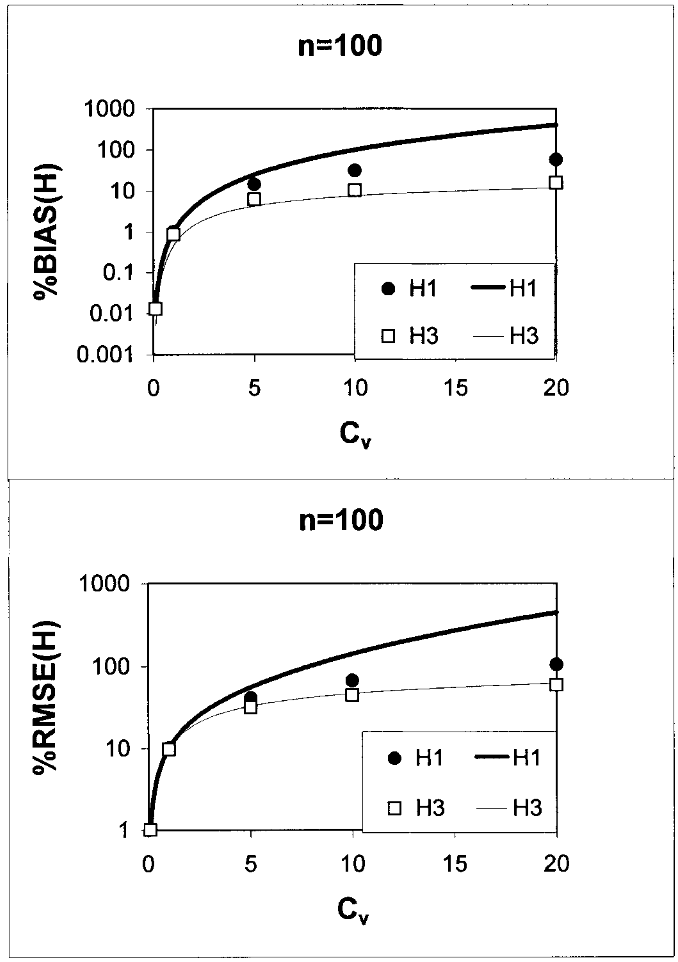

FIG. 3. Comparison of % Bias and % RMSE of Estimators $\hat{H}_1$ and $\hat{H}_3$ Using Monte Carlo and Analytical Approximation for Case *n* = 100

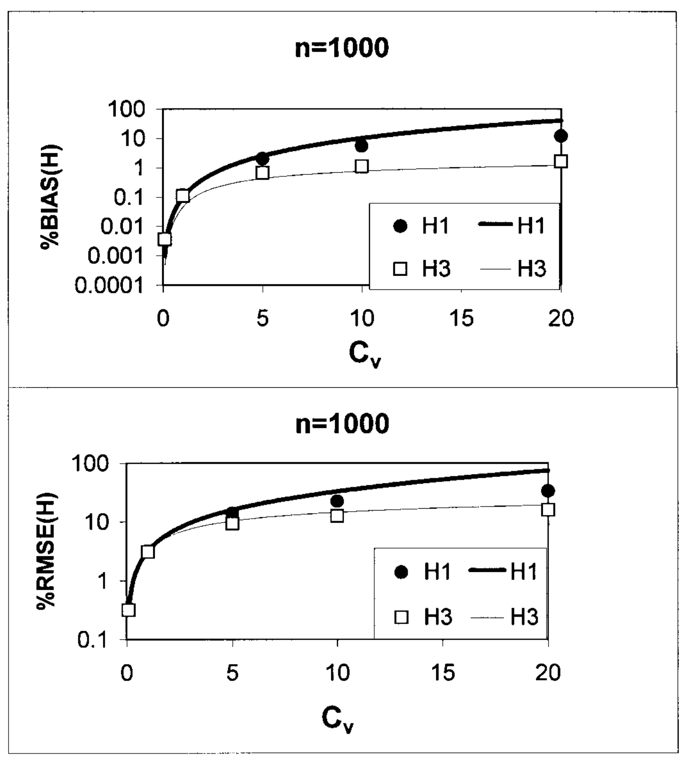

FIG. 4. Comparison of % Bias and % RMSE of Estimators $\hat{H}_1$ and $\hat{H}_3$ Using Monte Carlo and Analytical Approximation for Case *n* = 1,000

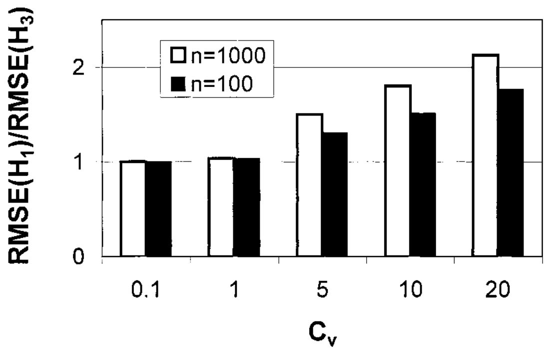

FIG. 5. Efficiency of $\hat{H}_3$ over $\hat{H}_1$

daily exposure concentration is used as an index to determine whether exposures are below acceptable risk thresholds.

Water quality standards are usually stated in terms of a maximum allowable x-day average concentration (EPA 1991). If a constant toxicant loading of L is introduced into a river, then the resulting in-stream concentration, accounting for dilution with streamflow, is

$$C_t = \frac{L}{X_t} \tag{19}$$

where C_t denotes concentration on day t; and X_t denotes streamflow on day t. For a fixed load L (fixed TMDL), the expectation of lifetime exposure concentration is

$$E[C] = L \cdot E\left[\frac{1}{Q}\right] = \frac{L}{H} \tag{20}$$

The harmonic mean streamflow H thus becomes the design streamflow on which the health risk assessment is based. This is the logic employed by Rossman (1990) and the EPA (1991) in recommending the harmonic mean flow for setting carcinogen TMDLs. The EPA (1991) currently recommends estimation of H from daily streamflow observations using $\hat{H}_1$. However, this research documents that $\hat{H}_1$ performs rather poorly when compared with the maximum likelihood estimator $\hat{H}_3$, particularly for the large sample sizes and large values of C_v exhibited by daily flow observations.

The U.S. Geological Survey (Martin and Ruhl 1993) have developed a regional model for estimating the harmonic mean of daily streamflow in Kentucky, based on $\hat{H}_1$. Given the results of Monte Carlo experiments reported here, future regionalization studies would be improved by using a maximum likelihood estimator $\hat{H}_3$, which has a higher precision than $\hat{H}_1$. To prove this point, one need only document that observed sequences of daily streamflow are well approximated by an LN2 distribution with values of $C_v > 5$.

Probability Distribution of Daily Streamflow

Previous literature on the statistical properties of daily streamflow is sparse. Rossman (1990) and the EPA (1991) recommend the estimators $\hat{H}_1$ or $\hat{H}_2$. The estimator $\hat{H}_2$ was derived assuming an LN2 distribution. To our knowledge there are no comprehensive studies that have evaluated the PDF of daily streamflow. Vogel and Fennessey (1994) discuss methods for fitting empirical nonparametric PDFs to samples of daily streamflow. Vogel and Fennessey (1993) use L-moment diagrams to document that samples of daily streamflow at 23 sites in Massachusetts are well approximated by a three-parameter lognormal (LN3) and a generalized Pareto (GP) distribution, with the GP distribution providing a slightly better fit overall. Interestingly, their evaluations also show that ordinary product moment ratios, such as the coefficient of variation and skewness, reveal almost no information about the distributional properties of daily streamflow. This remarkable phenomenon is due to the large skew associated with samples of daily streamflow, despite the fact that sequences of daily streamflows often contain tens of thousands of observations.

In this section we evaluate the PDF of daily streamflow in the United States by exploiting a national database of streamflow. The streamflow data set consists of records of average daily streamflow at 1,571 sites located throughout the United States. This data set termed the hydro-climatic data network (HCDN), is available on CD-ROM and the worldwide web (Slack et al. 1993). Because this database is described elsewhere, we do not provide details here.

To test the goodness-of-fit for various distributions to sequences of daily streamflow at these sites, we constructed L-moment diagrams. Hosking (1990) introduced L-moments and L-moment diagrams as an aid for obtaining an unbiased representation of the distributional properties of any random variable. L-moments are related to ordinary product moments, and the interpretation of L-skew and L-kurtosis is similar to that

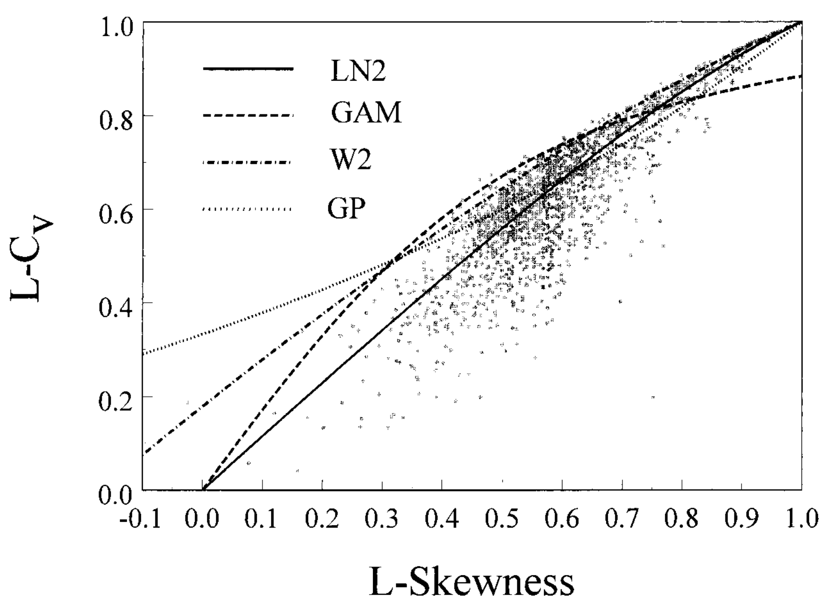

FIG. 6. *L*-Moment Diagram Illustrating Relationship between *L*-C_v and *L*-Skewness for Daily Streamflow at 1,571 Sites across United States

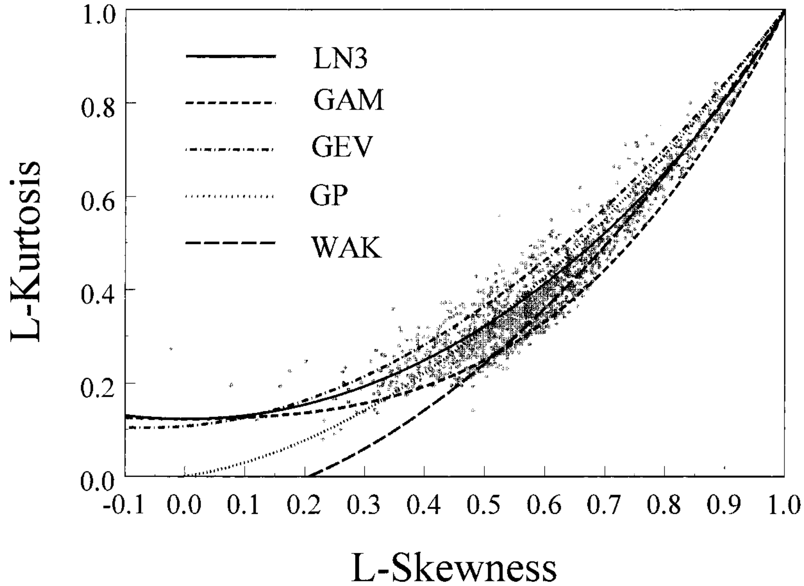

FIG. 7. *L*-Moment Diagram Illustrating Relationship between *L*-Kurtosis and *L*-Skewness for Daily Streamflow at 1,571 Sites across United States

of ordinary skew and kurtosis (Hosking 1990). Vogel and Fennessey (1993) found that *L*-moment diagrams are always preferred over ordinary product moment diagrams for assessing the goodness-of-fit of PDFs to observations, and are particularly useful for evaluating highly skewed observations.

Figs. 6 and 7 illustrate sample estimates of L-C_v, *L*-Skew, and *L-K*urtosis at the 1,571 HCDN sites. The streamflow data set contains records, which range from 9 to 115 years with an average record length of 44.8 years per site. Each point in Figs. 6 and 7 represents a single site with an average record length of (44.8 years) $\times$ (365 days/year) = 16,400 daily observations. In total, Figs. 6 and 7 are based on more than 25,000,000 daily streamflow observations. The theoretical relations between L-C_v and *L*-skewness are shown in Fig. 6 for LN2, gamma (GAM), two-parameter Weibull (W2), and generalized Pareto (GP) distributions. Fig. 7 shows the relation between *L-K*urtosis and *L*-Skewness for LN3, gamma (GAM), GP, generalized extreme value (GEV), and the lower bound of the five-parameter Wakeby (WAK) distributions. From Fig. 6, it can be seen that the distribution of daily streamflows is well approximated by the LN2 distribution. Fig. 7 illustrates that daily streamflows are also well approximated by a three-parameter lognormal distribution, as would be expected from the results in Fig. 6. In both Figs. 6 and 7, one expects significant sampling variability about the true LN2 relationship. All one can hope for is a theoretical relationship that passes through the center of mass of all the points in those figures.

Implications of Results on Wasteload Allocations

Fig. 6 illustrated that values of L-C_v and *L*-Skew are often well above 0.4 for observed sequences of daily streamflow in the United States. Fig. 8 illustrates the theoretical relationships between ordinary C_v and both L-C_v and *L*-Skew for an LN2 variable. Fig. 8 documents that values of L-C_v above 0.4 correspond to values of ordinary $C_v > 0.8$, and that values of *L*-skews larger than 0.4 correspond to ordinary $C_v > 1$ for an LN2 distribution. This implies that observations of daily streamflow usually exhibit values of C_v in excess of 1. According to Figs. 6–8, values of C_v of daily streamflow are often greater than 5. Because the estimator $\hat{H}_3$ was shown to be clearly superior to both $\hat{H}_1$ and $\hat{H}_2$ for large C_v and n, the estimator $\hat{H}_3$ is expected to provide much better (lower bias and variance) estimates of long-term exposure concentration than either $\hat{H}_1$ and $\hat{H}_2$.

In practice, wasteload allocation seeks to estimate the allowable TMDL L, based on a fixed concentration C_0, and a design flow H. The allowable TMDL is estimated from

$$\hat{L} = C_0 \cdot \hat{H} \tag{21}$$

The load L is also a random variable with properties similar to H because C_0 is a constant. Therefore, upward bias associated with the estimator $\hat{H}$ in (21) will lead to upward bias in estimated allowable pollutant loads. Therefore, existing EPA (1991) guidelines lead to upward bias in estimated allowable pollutant loads.

Use of the estimator $\hat{H}_3$ in (20) will lead to a nearly unbiased estimate of the long-term exposure concentration; however, it will not lead to an unbiased estimate of the wasteload L when used in (21). If the interest is in an unbiased estimate of the wasteload allocation L in (21) for a fixed exposure concentration C_0, then a more sensible design streamflow would be the arithmetic mean μ because the expected load is given by $E[L] = C_0E[X] = C_0\mu$.

APPLICATION TO ESTIMATION OF EFFECTIVE GEOPHYSICAL PROPERTIES

Previous studies (Gomez-Hernandez and Gorelick 1989) have shown that the harmonic mean provides a lower limit to the effective hydraulic conductivity of an aquifer system. Here, an effective value is defined as the equivalent value of the parameter that produces an output variable that replicates the mean behavior of the system over a range of variability. For example, effective hydraulic conductivity could be defined as that value of hydraulic conductivity that reproduces the mean specific discharge or the average ground-water outflow to a river channel. Gomez-Hernandez and Gorelick (1989) argue that effective hydraulic conductivity near a well lies between the geometric and harmonic means. In geophysical reservoir engineering studies, the harmonic mean has been widely used for a long time. For example, Muskat (1937, pp. 402–404) suggests the use of $\hat{H}_1$ as an appropriate estimate of the effective permeability for a layered medium with cross-layer flow. Jensen (1998) summarizes numerous studies that evaluate different p-norms [(6)] in various physical situations. A review of the literature reveals that it is unclear which value of p in (6) is best suited for the estimation of the effective geophysical properties of a heterogeneous system. Nevertheless, it is clear

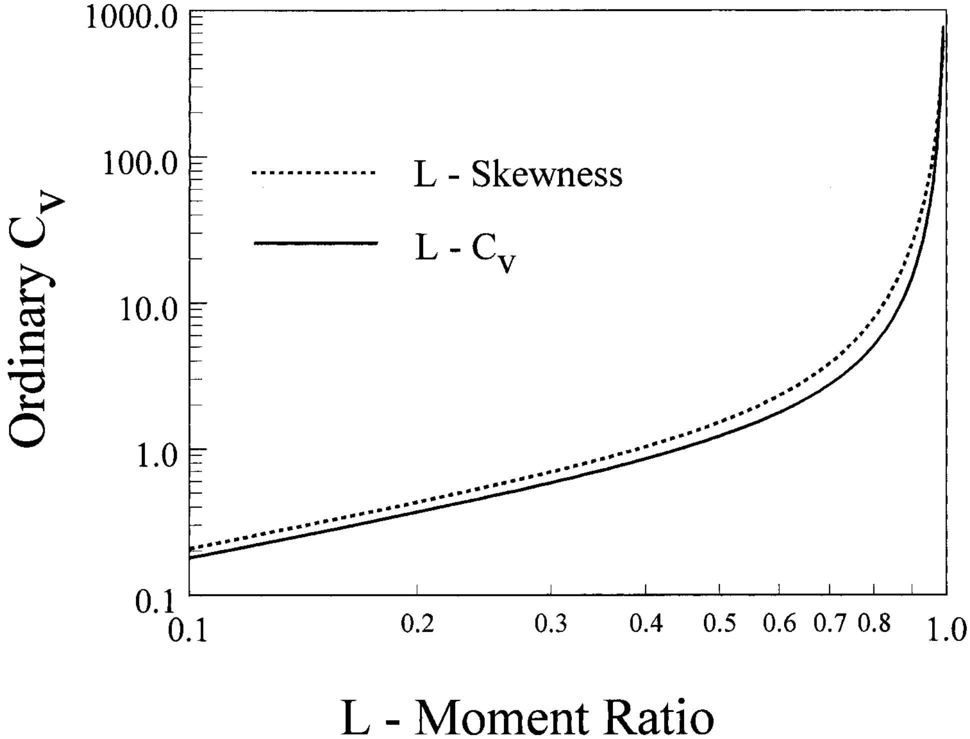

FIG. 8. Relationship between Ordinary C_v and Two L-Moment Ratios for LN2 Distribution

from this study, that an MLE of the p-norm in (6) provides an attractive alternative to the moment estimators currently recommended.

Ababou and Wood (1990) and Jensen (1998) report that the generalized mean or p-norm [defined in (6)] of an LN2 random variable can be written as

$$\mu_p = \exp(p\mu_y + p^2\sigma_y^2/2) \tag{22}$$

Jensen (1998) recommended the p-norm estimator

$$\hat{\mu}_p = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^p\right)^{1/p} \tag{23}$$

which reduces to $\hat{H}_1$ when $p = -1$ and to the arithmetic mean when $p = 1$. A preferred estimator of the p-norm for LN2 observations is the MLE given by

$$\hat{\mu}_p^* = \exp(p\hat{\mu}_y + p^2\hat{\sigma}_y^2/2) \tag{24}$$

where $\hat{\mu}_y$ and $\hat{\sigma}_y$ are given in (15b) and (15c), respectively. One expects the estimator $\hat{\mu}_p^*$ to have roughly the same advantages over $\hat{\mu}_p$ as the estimator $\hat{H}_3$ had over $\hat{H}_1$. Note that Crow and Shimizu (1988, Section 2.8) derive the sampling properties of various alternative estimators of the statistic $\theta = \exp(a\mu_y + b\sigma_y^2)$. Note that $\theta = \mu_p$ for $a = p$ and $b = p^2/2$, so that their results are useful for evaluation of the performance of alternative estimators of the p-norm of an LN2 variate.

Numerous investigators have evaluated the goodness-of-fit of the LN2 distribution to samples of hydraulic conductivity. Law (1944) first introduced the LN2 distribution for describing the distribution of permeability. Bensen (1993) reviews numerous studies that recommend the LN2 distribution for modeling the distribution of permeability and conductivity observations for naturally deposited soils and aquifers and for compacted soils and soil liners. Bensen (1993) uses the probability plot correlation coefficient hypothesis test and L-moment diagrams to compare the goodness-of-fit of several probability distributions to the conductivity of compacted soils. Bensen (1993) found that the LN3 and GEV distributions are often preferred over the LN2 distribution for characterizing the probability distribution of the hydraulic conductivity of compacted soils.

CONCLUSIONS

The harmonic mean is a natural measure of central tendency of random variables in engineering applications relating to: (1) estimation of the design streamflow for wasteload allocations (Rossman 1990); (2) estimation of the effective or aggregate properties of geophysical media (Gomez-Hernandez and Gorelick 1989); and (3) the design of declining rate filter beds (Saatci 1989). Unlike other measures of central tendency, such as the mean and median, no literature exists describing the statistical properties of alternative estimators of the harmonic mean. Previous investigations have employed a standard moment estimator [see $\hat{H}_1$ in (12)] that has been shown by Jensen et al. (1997) and Jensen (1998) to exhibit significant bias, particularly for large values of C_v. To avoid bias, Jensen et al. (1997) and Jensen (1998) introduce an approximately unbiased estimator of H, where the unbiasing factor depends on a sample estimate of C_v. Unfortunately, their estimator will not really be unbiased in practice, because Jensen et al. (1997) and Jensen (1998) ignore the downward bias associated with sample estimates of C_v (Vogel and Fennessey 1993) in their derivations. Furthermore, their unbiasing factors introduce additional variance (due to the need to estimate C_v), making their resulting estimator unstable, particularly for large C_v and small n.

This study introduces an MLE for the harmonic mean [(15b)] and the p-norm [(23)], which are nearly unbiased estimators and are generally more efficient (lower RMSE) than the alternative estimators introduced previously, for lognormal observations. We found that the MLEs can lead to improvements in our ability to estimate harmonic mean streamflows and p-norms of geophysical properties.

ACKNOWLEDGMENTS

The writers are grateful to Ian Wilson for computing the L-moments of the mean daily flow observations, and to Tim Cohn, Jerry L. Jensen, and two anonymous reviewers whose comments led to important improvements in this paper.

APPENDIX. REFERENCES

Ababou, R., and Wood, E. F. (1990). ''Comment on 'Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge,' by J. J. Gomez-Hernandez and S. M. Gorelick.'' *Water Resour. Res.*, 26(8), 1843–1846.

Aitchison, J., and Brown, J. A. C. (1969). *The lognormal distribution with special reference to its uses in economics.* Cambridge University Press, New York.

Benjamin, J. R., and Cornell, C. A. (1970). *Probability, statistics, and decision for civil engineers.* McGraw-Hill, New York.

Benson, C. H. (1993). ''Probability distributions of hydraulic conductivity of compacted soil liners.'' *J. Geotech. Engrg.*, ASCE, 119(3), 471–486.

Burk, F. (1985). ''By all means.'' *The Am. Math. Monthly*, 92(1), 50.

Crow, E. L., and Shimizu, K. (1988). *Lognormal distributions: Theory and applications.* Marcel Dekker, New York.

EPA. (1991). ''Technical support document for water quality-based toxics control.'' *EPA/505/2-90-001*, Office of Water.

Finney, D. J. (1941). ''On the distribution of a variate whose logarithm is normally distributed.'' *Supplement to the J. Royal Statistical Soc. (Industrial and Agric. Res. Sec.)*, VII(2), (Issued with Part IV, Vol. 104), 155–161.

Gomez-Hernandez, J. J., and Gorelick, S. M. (1989). ''Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge.'' *Water Resour. Res.*, 25(3), 405–419.

Hosking, J. R. M. (1990). ''L-moments: Analysis and estimation of distributions using linear combinations of order statistics.'' *J. Royal Statistical Soc., Ser. B.*, 52(1), 105–124.

Jensen, J. L. (1998). ''Some statistical properties of power averages for lognormal samples.'' *Water Resour. Res.*, 34(9), 2415–2418.

Jensen, J. L., Thomas, S. D., and Corbett, P. W. M. (1997). ''On the bias and sampling variation of the harmonic average.'' *Mathematical Geology*, 29(2), 267–276.

Kendall, S. M., Stuart, A., and Ord, J. K. (1977). *The advanced theory of statistics*, 4th Ed., Vol. 1, Macmillan, New York.

Landwehr, J. M. (1978). ''Some properties of the geometric mean and its use in water quality standards.'' *Water Resour. Res.*, 14(3), 467–473.

Law, J. (1944). ''Statistical approach to the interstitial heterogeneity of sand reservoirs.'' *Trans. AIME*, 155, 202.

Martin, G. R., and Ruhl, K. J. (1993). ''Regionalization of harmonic-mean streamflows in Kentucky.'' *Water Resour. Investigations Rep. 92-4173*, U.S. Geological Survey, Louisville, Ky.

Muskat, M. (1937). *The flow of homogeneous fluids through porous media.* McGraw-Hill, New York.

Rossman, L. A. (1990). ''Design stream flows based on harmonic means.'' *J. Hydr. Engrg.*, ASCE, 116(7), 946–950.

Saatci, A. M. (1989). ''Harmonic mean conductivity in declining rate filters.'' *J. Envir. Engrg.*, ASCE, 115(2), 462–466.

Slack, J. R., Lumb, A. M., and Landwehr, J. M. (1993). ''Hydro-climatic data network (HCDN).'' *Water Resour. Investigations Rep. 93-4076*, U.S. Geological Survey, Washington, D.C. ⟨http://wwwrvares.er.usgs.gov/hcdn_cdrom/1st_page.html⟩.

Stedinger, J. R. (1980). ''Fitting lognormal distributions to hydrologic data.'' *Water Resour. Res.*, 16(3), 481–490.

Vali, M. A. (1942). ''On the sampling distribution of harmonic means.'' *Bull. Calcutta Math. Soc.*, 34, 87–91.

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

Vogel, R. M., and Fennessey, N. M. (1994). ''Flow-duration curves. I: New interpretation and confidence intervals.'' *J. Water Resour. Plng. and Mgmt.*, ASCE, 120(4), 485–504.