P1A.7 THE BIAS IN MOMENT ESTIMATORS FOR PARAMETERS OF DROP-SIZE DISTRIBUTION FUNCTIONS

Paul L. Smith* and Donna V. Kliche
IAS, South Dakota School of Mines and Technology, Rapid City, SD

1. INTRODUCTION

Investigators frequently acquire observations of raindrop sizes and seek to describe the drop-size distributions (DSDs) by analytical expressions, the exponential or gamma function being most common. Moment methods to estimate parameters for the DSD functions have become more or less traditional, even though Haddad *et al.* (1996, 1997) have pointed out that the methods are biased. The intuitive appeal of such methods is almost irresistible, and the associated mathematical manipulations lend a convincing aura. However, the methods are indeed biased – in the statistical sense that the expected values of the "fitted" parameters differ from the parameters of the underlying raindrop populations – and so lead to erroneous inferences about the characteristics of the DSDs being sampled.

The bias in the moment methods can be demonstrated by testing their ability to recover parameters of known DSDs from which samples are taken. This must be done by computer simulation, as the DSDs in nature are inherently unknown. The simulations herein use a Monte Carlo simulation procedure similar to that described in Smith *et al.* (1993); the Appendix below summarizes the procedure.

2. CHARACTERISTICS OF SAMPLING DISTRIBUTIONS

Sampling from long-tailed DSDs of the exponential or gamma types exhibits certain general features. Sample values of the DSD moments M_i (i = 3 gives LWC; i = 6 gives Z) are unbiased: the expected, or mean, sample value of M_i corresponds to that of the drop population being sampled. However, the sampling distributions are skewed, as shown for exponential DSDs in Smith *et al.* (1993). The skewness is greater for higher-order moments and for smaller sample sizes. Sampling from gamma distributions with positive shape parameter (μ) displays the same general features, but the skewness is reduced. Figure 1 shows an example of the skewness in the Z sampling distribution for a gamma distribution with μ = 2.

Sampling the small drops can be a major instrumental problem for exponential DSDs, but is of less concern for gamma DSDs. However, adequately sampling the relatively rare large drops remains a concern. Regardless of the (population) value of μ, fewer than one drop in 100 in a DSD is larger than D = 1.5 D_m, where D_m is the mass-weighted mean diameter, and fewer than one in 1500 is larger than D = 2 D_m. However, for μ = 2 the drops larger than 1.5 D_m contribute more than ten percent of the LWC and almost half the reflectivity factor. Consequently, the relatively large but relatively rare drops tend to be important in determining the moments of physical significance.

Corresponding author address: P.L. Smith, 501 East Saint Joseph, Rapid City, SD 57701-3995;
e-mail: paul.smith@ias.sdsmt.edu

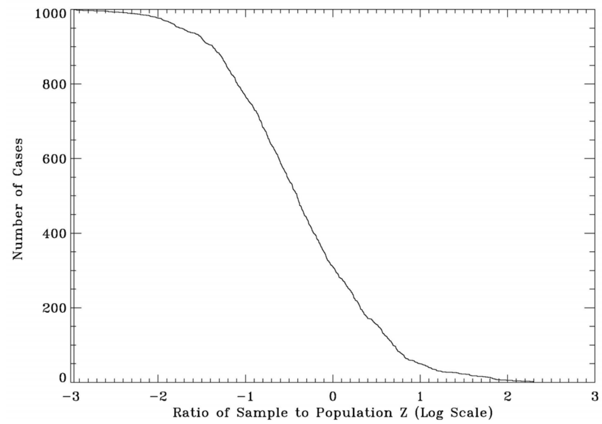

Fig. 1: Example of sampling distribution for moment M_6 (Z), showing cumulative number of cases with sample Z larger than value indicated on abscissa. Population DSD: gamma, μ = 2. Mean sample size: 50 drops. Median value –4.1 dB; (logarithmic) mean value –3.84 dB.

Figure 5 of Smith *et al.* (1993) showed that the maximum drop size in an exponential DSD is rarely approached in samples of even hundreds of drops. Figure 2 here shows a histogram of the largest drop x_{max} (= D_{max}/D_m) in samples averaging 50 drops from a gamma DSD with μ = 2. There is clearly no basis for assuming truncation of the underlying DSD at the maximum observed drop diameter, with samples of such sizes.

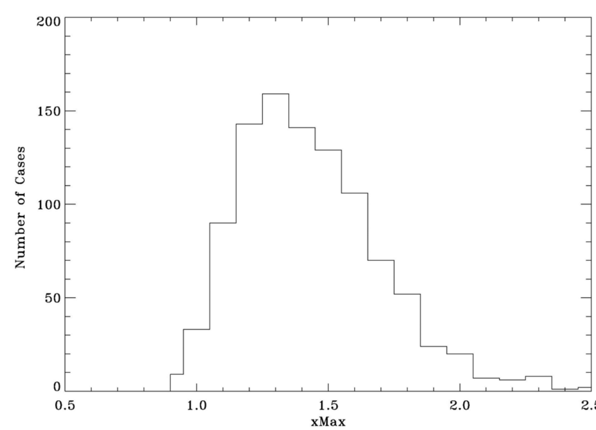

Fig. 2: Example of sampling distribution for maximum drop size $x_{max} = D_{max}/D_m$. Population DSD: gamma, μ = 2. Mean sample size: 50 drops. Median value 1.34; mean value 1.380.

3. MOMENT ESTIMATORS

The use of moment methods to estimate parameters for DSD functions evidently began with Waldvogel's (1974) paper on the "N_0 jump" of DSDs. He used moments M_3 and M_6 (i.e. LWC and Z) to determine pairs of parameters for exponential functions that purportedly represented the observed DSDs. This was surely one of the greatest mistakes he ever made, for distributions "fitted" in this way do not look at all like the samples upon which they are based. Figure 3 shows an example from raindrop camera data illustrating the large discrepancies involved.

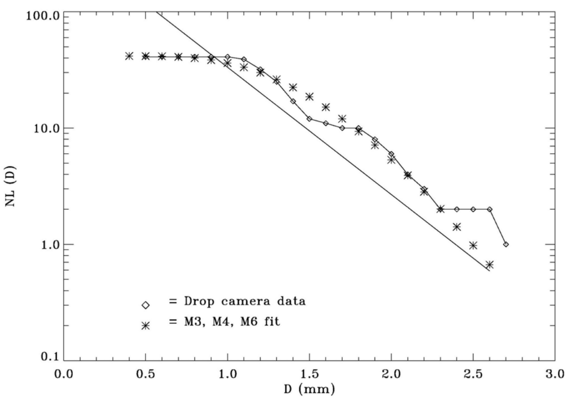

Fig. 3: Example of drop-size data from Illinois State Water Survey raindrop camera (Majuro, Marshall I., 0930 16 May 1959). Ordinate shows "inverse cumulative" number concentration $N_L(D)$ of drops of diameter D or larger. Unlabelled straight line shows "Waldvogel fit" based on M_3 and M_6. Asterisks indicate "moment-method fit" based on M_3, M_4, and M_6.

More recently, estimators involving three moments have been used in attempts to fit gamma distributions to DSD observations. Examples include Ulbrich (1983); Kozu and Nakamura (1991); Smith (1993); Tokay and Short (1996); and Ulbrich and Atlas (1998). These procedures can produce closer fits to the observations, at least for part of the size spectrum (Fig. 3). However, the apparent closeness of fit is misleading, because the fitted distributions typically do not resemble the original drop populations from which the samples are taken.

4. THE BIAS IN MOMENT ESTIMATORS

As noted, Haddad *et al.* (1996, 1997) point out that moment estimators are biased, and worthy statisticians are appalled at the idea of using this approach. That estimates of DSD parameters obtained in this way are biased was actually indicated in Smith *et al.* (1993). Figure 4, reproduced from that paper, compares sample estimates of D_m to the value of D_m for the exponentially-distributed population from which the (simulated) samples were drawn. In this example, more than 80% of the values are underestimates, and the mean (the expected value) is about 78% of the population value. In terms of the more familiar exponential slope parameter Λ (= $4/D_m$), this means that Λ is generally overestimated and the "fitted" DSDs will contain too many small drops and too few large ones.

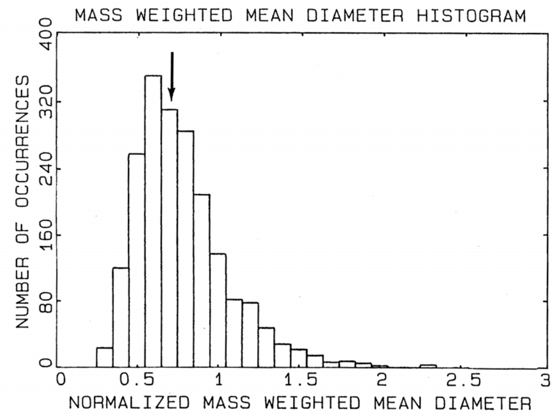

Fig. 4: Example of sampling distribution for mass-weighted mean diameter D_m. Population DSD: exponential. Mean sample size: 20 drops. (From Smith et al., 1993).

The bias in the moment estimators for gamma distributions is also substantial. Figure 5 shows the distribution of estimated values of D_m for samples averaging 20 drops (i.e. N_T = 20) from a gamma distribution with μ = 2. Here just 74% of the values are underestimates, and the expected value is 90.5% of the population value. The bias decreases, along with the skewness of the sampling distributions, as the population shape parameter μ increases, and also decreases with increasing sample size.

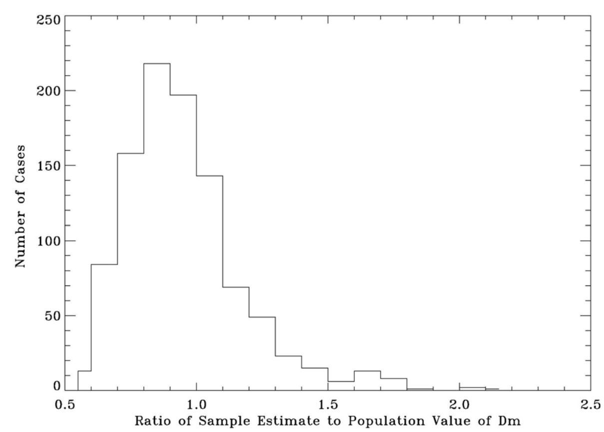

Fig. 5: Example of sampling distribution for "moment-method fit" of D_m. Population DSD: gamma, μ = 2. Mean sample size: 20 drops. Median value 0.87; mean value 0.905.

Of greater concern is the bias in the estimates of the gamma shape parameter μ. Figure 6 shows a distribution of values of μ estimated using the moments M_3, M_4, and M_6, as employed by Ulbrich (1983), Kozu and Nakamura (1991), and Tokay and Short (1996). In this simulation the population PDF had μ = 2 and nearly all of the sample values of μ are overestimates – most drastically so. As the bias in the moment estimators tends to overestimate μ and underestimate D_m, the tendency will be to overestimate the more customary "slope" (scale) parameter λ. The hybrid approach used by Testud *et al.* (2001) does not employ a moment-based calculation to determine μ, but the estimators for their other gamma parameters (D_m and N_o*) are biased.

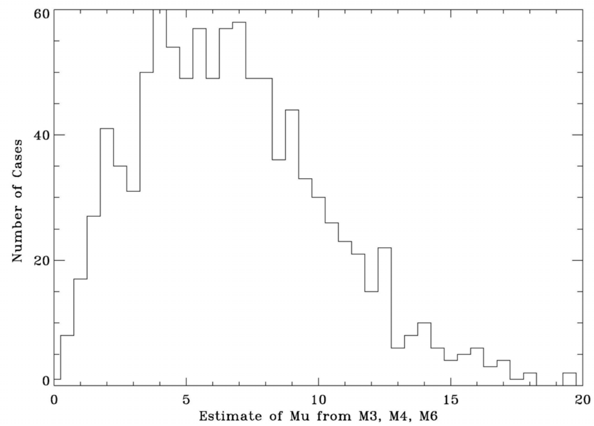

Fig. 6: Example of sampling distribution for gamma shape parameter μ, based on M_3, M_4, and M_6. Population DSD: gamma, μ = 2. Mean sample size: 50 drops. Median value 6.6; mean value 7.04.

This bias can be quite misleading. One can sample from what is actually an exponential DSD; use moment methods in an attempt to "fit" parameters of a gamma distribution to the observations; find (biased) high values of μ; and conclude, quite erroneously, that the population DSD was gamma after all. Figure 7 shows an example where a sample of 50 drops was drawn from an exponential DSD. While the gamma M_3, M_4, M_6 fit matches part of the sample distribution reasonably well, it in no way corresponds to the population from which the sample was drawn. The "fitted" value of μ is 7.66;

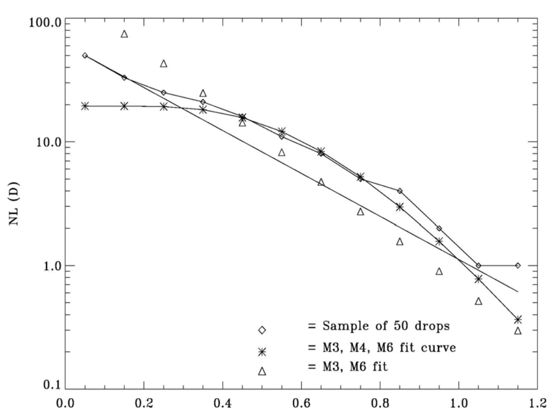

Fig. 7: A sample of 50 drops drawn from an exponential population (population DSD indicated by unlabelled straight line). Gamma curve with asterisks indicates "moment-method fit" based on M_3, M_4, and M_6. Triangles indicate "Waldvogel fit" based on M_3 and M_6. Ordinate as in Fig. 3.

the expected value of μ for samples of this size, drawn from an exponential DSD (μ = 0), is about μ = 7.7. Thus, sampling the DSD with samples of this size will not reveal that the underlying DSD is exponential.

As the sample size increases, the bias diminishes, so that with very large samples the moment methods may give approximations of the population parameters that become sufficiently accurate for practical purposes. Further simulations are exploring the sample sizes required for this to occur, but thus far it appears that the samples required are in the hundreds, if not thousands, of drops. It is also worth noting that Joss and Gori (1978) found that observed DSDs approach the exponential form as the sample size is increased by averaging more DSD data.

5. RELATED FINDINGS

Ulbrich (1983) and others have noted a correlation between the gamma distribution parameters n_1 and λ in data from DSD observations. In the present notation,

$$n_1 = N_T\,(\mu+4)^{\mu+1} / (\mu!\,D_m^{\mu+1})$$

$$\lambda = (\mu + 4)/D_m$$

The sampling variation of D_m is relatively small (Fig. 5) compared to that of μ (Fig. 6), so "fitted" values of λ will depend mostly on μ. Figure 8 illustrates a strong correlation between n_1 and μ arising from sampling variations alone – i.e. there is no physical significance to this correlation.

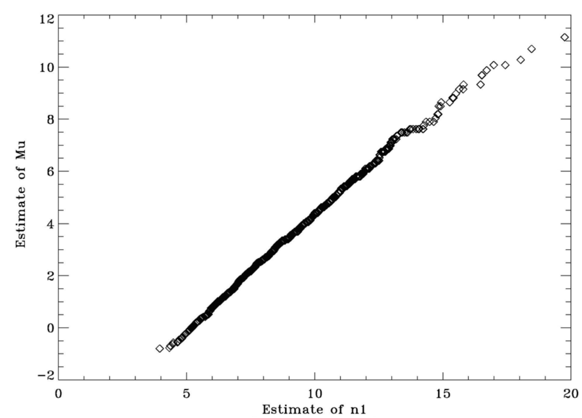

Fig. 8: Illustration of correlation (resulting strictly from sampling variation) between gamma parameters n_1 and μ from "moment-method fit" based on M_3, M_4, and M_6. Population DSD: gamma, μ = 2. Mean sample size: 50 drops.

Figure 9 shows another strong relationship found in the simulations, between the moment estimate of μ and the largest drop x_{max} in the sample. Larger values of x_{max} lead to smaller estimates of μ, which are generally closer to the population value. This relationship appears to be insensitive to the sample size, and if it turns out to be equally insensitive to the population value of μ (or if values of μ in nature vary little) it would be subject to experimental verification. Larger samples (i.e. higher N_T) generally lead to larger values of x_{max}, so the value of increased sample size in reducing the bias in estimates of μ is evident. Drop samples sufficiently large to describe the large-drop end of the size spectrum adequately are therefore essential if moment methods are to yield reasonable estimates of the DSD parameters.

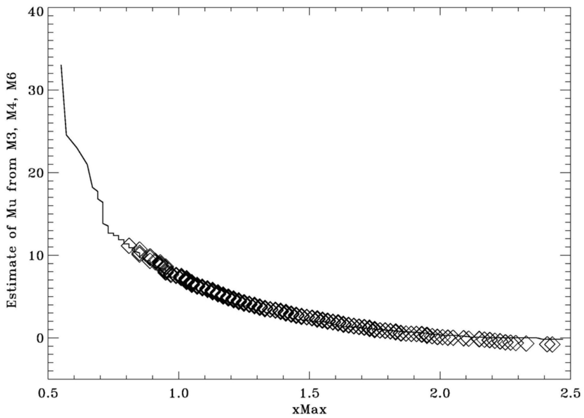

Fig. 9: Illustration of correlation between "moment-method fit" of gamma shape parameter μ and maximum drop size in the sample. Population DSD: gamma, μ = 2. Mean sample sizes: 20 drops (solid line), 50 drops (diamonds).

6. CONCLUSIONS

Moment estimators for parameters of DSD functions are inherently biased. They tend to give erroneous values of the DSD parameters unless the drop samples are much larger than those commonly available. In particular, estimates of the gamma shape parameter μ tend to be far larger than the shape parameter of the underlying DSD from which the samples are taken. The bias is greatest for small sample sizes, and also greater when higher-order moments of the observed DSDs are used in the fitting process (not demonstrated here).

Moment methods might provide estimates of DSD parameters of sufficient accuracy if very large samples (hundreds, perhaps thousands) of drops were available. Failing that, some alternative approach to fitting the observed DSDs must be used. Haddad *et al.* (1996, 1997) suggested a maximum likelihood approach, which may be satisfactory if the sampling errors are not too great.

Acknowledgment: This material is based upon work supported by the National Science Foundation under Grant No. ATM-9907812.

REFERENCES

Haddad, Z.S., S.L. Durden and E. Im, 1996: Parameterizing the raindrop size distribution. *J. Appl. Meteor.,* **35**, 3-13.

Haddad, Z.S., D.A. Short, S.L. Durden, E. Im, S. Hensley, M.B. Grable and R.A. Black, 1997: A new parameterization of the rain drop size distribution. *IEEE Trans. Geosci. Remote Sensing,* **35**, 532-539.

Joss, J., and E.G. Gori, 1978: Shapes of raindrop size distributions. *J. Appl. Meteor.,* **17**, 1054-1061.

Kozu, T. and K. Nakamura, 1991: Rainfall parameter estimation from dual-radar measurements combining reflectivity profile and path-integrated attenuation. *J. Atmos. Oceanic Tech.,* **8**, 259-270.

Smith, J.A., 1993: Marked point process model of raindrop-size distributions. *J. Appl. Meteor.,* **32**, 284-296.

Smith, P. L., Z. Liu and J. Joss, 1993: A study of sampling-variability effects in raindrop-size observations. *J. Appl. Meteor.*, **32**, 1259-1269.

Testud, J., S. Oury, R.A. Black, P. Amayenc and X. Dou, 2001: The concept of "normalized" distribution to describe raindrop spectra: A tool for cloud physics and remote sensing. *J. Appl. Meteor.,* **40,** 1118-1140.

Tokay, A. and D.A. Short, 1996: Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds. *J. Appl. Meteor.,* **35**, 355-371.

Ulbrich, C.W., 1983: Natural variation in the analytical form of the raindrop size distribution. *J. Climate Appl. Meteor.,* **22**, 1764-1775.

Ulbrich, C.W. and D. Atlas, 1998: Rainfall microphysics and radar properties: Analysis methods for drop size spectra. *J. Appl. Meteorol.,* **37**, 912-923.

Waldvogel, A., 1974. The N_0 jump of raindrop spectra. *J. Atmos. Sci.,* **31**, 1067-1078.

Appendix: Summary of Simulation Procedure for Raindrop Sampling

First write the drop-size distribution n(D) as the product of the (average) total drop concentration N_T (m^{-3}) and the probability density function (PDF) of drop diameter D:

$$n(D) = N_T \text{ (PDF)}$$

According to Mielke (1976), the gamma PDF is

$$\text{PDF} = (\lambda^{\mu+1} / \mu!)\, D^{\mu} \exp(-\lambda D)$$

In terms of the mass-weighted mean diameter D_m, this becomes

$$\text{PDF} = \frac{(\mu+4)^{\mu+1}}{\mu!} \frac{D^{\mu}}{D_m^{\mu+1}} e^{-(\mu+4)D/D_m}$$

Define x = D/D_m; then

$$n(x) = N_T \frac{(\mu+4)^{\mu+1}}{\mu!} x^{\mu} e^{-(\mu+4)x}$$

To simulate sampling from a DSD with a specified value of μ (μ = 0 gives an exponential DSD), specify the value of N_T and assume a 1 m^3 sampling volume. (This is equivalent to a combination of any arbitrary N_T and a sample volume that will lead to the same average number of drops in a sample.) Use a Poisson random number generator with mean value N_T to generate values of the drop count C in individual samples, followed by a gamma random number generator for each sample to generate C values of the normalized drop diameter x from the gamma PDF. The latter is readily represented (for integer values of μ) in a look-up table.
