Sampling Issues in Estimating Radar Variables from Disdrometer Data

PAUL L. SMITH

South Dakota School of Mines and Technology, Rapid City, South Dakota

(Manuscript received 8 February 2016, in final form 23 August 2016)

ABSTRACT

Simulation of sampling from gamma-distributed raindrop populations demonstrates that significant biases and substantial errors can occur in estimates of polarimetric radar variables based on samples of raindrop populations obtained with disdrometers. Biases and RMS errors of 0.5 dB or more in estimates of differential reflectivity Z_{dr} can occur with samples of even a few hundred drops; significant biases and errors also occur in estimates of reflectivity Z_H or specific differential phase K_{dp}. The results indicate that very large samples would be required to obtain adequate representation of the population characteristics for many radar applications. They also suggest that greater attention is needed to the sample sizes in the disdrometer data used in developing polarimetric rainfall-rate estimators or hydrometeor classification algorithms.

1. Introduction

Sampling issues are ubiquitous in attempts to estimate radar variables, such as reflectivity, differential reflectivity, or specific differential phase, from raindrop samples. Examples of previous studies of these issues include Cornford (1967), Joss and Waldvogel (1969), Gertzman and Atlas (1977), Smith et al. (1993), Smith and Kliche (2005), Kliche (2007), Cao and Zhang (2009), and Smith et al. (2009). A key issue is to determine the biases and uncertainties in estimates of the polarimetric radar variables from raindrop disdrometer data. Estimating the variables is a statistical problem of estimating the characteristics of a population from a sample or samples drawn from that population. In the radar context, the population of interest is the array of raindrops present in the volume of the atmosphere that contributes to the radar echo observed at some given instant. The true characteristics of that population are never known, so the biases and uncertainties cannot be determined from actual disdrometer data. Instead, one must resort to sampling simulations in which a population is postulated and samples are drawn therefrom according to some plausible sampling procedure.

The purpose of the present work is to use such a simulation procedure to assess the biases and errors involved in estimating values of the reflectivity Z_e, differential reflectivity Z_{dr}, or specific differential phase K_{dp} from disdrometer data. A statistical estimate of population characteristics is known to improve as the size of the sample increases, so an important focus will be on sample sizes. The ideal outcome would be guidelines that indicate such things as how large a sample is needed to estimate, say, Z_{dr} with bias and uncertainty no greater than some desired values; or how to assess the uncertainty in such estimates for a given sample. Unfortunately, the range of possible drop populations is too wide to allow any simple guidelines; however, the present results provide useful indications of the magnitude of the sampling problems. They further suggest that samples from available disdrometers often fall short of what might be desired.

2. Simulation of the raindrop sampling process

The straightforward way to investigate these sampling issues is to postulate a raindrop population having a drop size distribution (DSD) function with known properties; simulate the drawing of repeated random samples from that distribution; calculate the sample estimates of the variables of interest; and then compare those estimates with the known population values by examining the appropriate sampling statistics. The

Corresponding author address: Paul L. Smith, South Dakota School of Mines and Technology, 501 E. St. Joseph St., Rapid City, SD 57701.
E-mail: paul.smith@sdsmt.edu

DOI: 10.1175/JTECH-D-16-0040.1

familiar gamma DSD model provides a convenient basis for this work:

$$n(D_{eq}) = N_T \frac{(\mu+4)^{\mu+1}}{\mu!} \frac{D_{eq}^{\mu}}{D_m^{\mu+1}} \exp[-(\mu+4)D_{eq}/D_m]. \quad (1)$$

The parameters in (1) have a clear physical significance: D_{eq} is the diameter of a sphere with the same volume as a raindrop, N_T represents the total drop number concentration, μ is the distribution shape parameter, and D_m denotes the mass-weighted mean drop diameter. The simulation procedure is similar to that followed in Kliche (2007) and Smith et al. (2009). For a given case, we specify values for the parameters μ and D_m; because one main interest is in the effects of sample size, we organize the simulations in terms of sample sizes rather than N_T. For differential reflectivity this is easy enough, because the value of Z_{dr} is independent of the drop number concentration. For reflectivity, using a logarithmic (dBZ) scale allows those results to be organized in terms of sample sizes. For differential phase, using percentage values for the errors in K_{dp} accomplishes the same objective.

We assume the drops to be randomly (i.e., Poisson) distributed in space. The validity of this assumption has been questioned (e.g., Kostinski and Jameson 1997), and here we note only that deviations from the Poisson distribution would exacerbate the sampling problems. Thus, the biases and errors determined from these simulations represent lower bounds on what might be encountered in practice. We then draw 100 random volume samples[1] from this DSD for each specified sample size. For each sample, we calculate values of Z_{dr} and K_{dp}, along with those of Z_H (the reflectivity factor for horizontal polarization), using parameters for water drops at 2.8 GHz and 10°C. We then examine the sampling statistics [mean, median, standard deviation, and root-mean-square (RMS) error values] to evaluate the biases and errors in the estimates of those variables.

To establish a reference baseline, we first use the parameter values $\mu = 2$ and $D_m = 2$ mm and initial sample sizes of 500 drops. Section 3 presents the simulation results for the sampling biases and errors in estimates of Z_{dr}, first for this baseline case and then for a range of sample sizes, values of μ, and values of D_m. Section 4 provides similar results for the estimates of Z_H and K_{dp}; section 5 discusses some relationships that emerged from examination of the simulation results. Section 6 highlights the sample size problem, and section 7 concludes with a discussion of the implications of these results for efforts to estimate these polarimetric radar variables from disdrometer data.

[1] The appendix discusses how to convert between surface and volume samples.

TABLE 1. Baseline population variables: $\mu = 2$, $D_m = 2$ mm ($\lambda = 3$ mm^{-1}), $N_T = 800$ m^{-3}.

Base variables	
Z_H (dBZ)	44.04
Z_{dr} (dB)	1.585
K_{dp} (deg km^{-1})	0.42
Weighted mean diameters (mm)	
K_{dp} weighted	2.59
Z weighted	3.00
Z_H weighted	3.07
Z_{dr} weighted	4.43

3. Estimates of Z_{dr}

Table 1 summarizes the characteristics of the baseline drop population; for purposes of illustration, the table uses a drop number concentration of 800 m^{-3}, though this value has no direct bearing on the simulations. The Z_{dr} -weighted mean diameter was calculated as follows:

$$D_{Z_{dr}} = \frac{\sum_{D_{eq}} D_{eq} \Delta Z_H(D_{eq})}{\sum_{D_{eq}} \Delta Z_V(D_{eq})}, \quad (2)$$

where the ΔZs are the contributions of all the drops of size D_{eq} to the respective reflectivity values. The comparison of the weighted mean diameters suggests that sampling problems will increase as we go from K_{dp} to Z_H to Z_{dr}; that is borne out in the results presented below.

a. *Biases and errors in sample estimates of Z_{dr} for the baseline population*

Figure 1 (open circles) shows the errors in the sample estimates for the baseline case $\mu = 2$, $D_m = 2$ mm, and sample sizes of 500 drops. Table 2 (middle column) shows the corresponding sampling statistics; the mean value (bias) is −0.15 dB and the RMS error is 0.46 dB. The median error is −0.25 dB and 74% of the sample values are underestimates. These facts illustrate the skewness that typifies disdrometer sample estimates of radar variables with the usual heavy-tailed DSDs (e.g., Smith et al. 1993; Uijlenhoet et al. 2006; Smith et al. 2009). The basic reason for this is not difficult to understand; most of the contribution to Z_{dr} comes from large

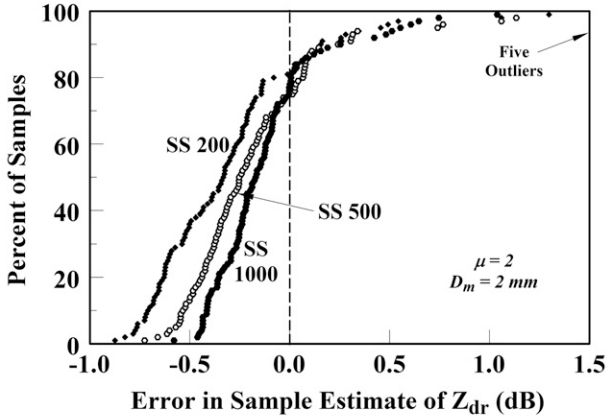

FIG. 1. Cumulative distributions of errors in sample estimates of Z_{dr} for different sample sizes. Samples from baseline gamma DSD with parameters $\mu = 2$, $D_m = 2$ mm.

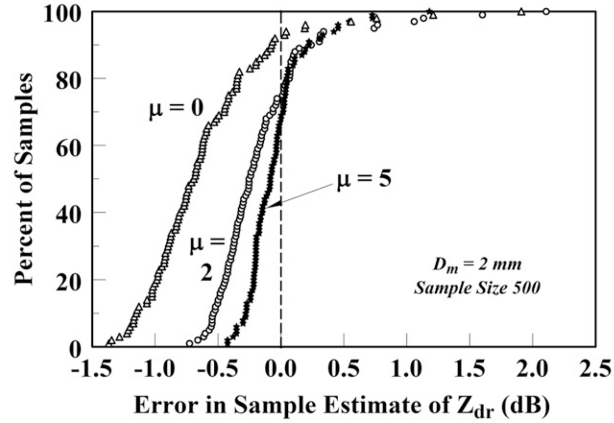

FIG. 2. Cumulative distributions of errors in sample estimates of Z_{dr} for different values of μ. Samples from gamma DSD with $D_m = 2$ mm; sample sizes 500 drops.

drops that are relatively rare. Such drops are few, or even absent, in most of the samples.

The bias and the RMS error raise questions about the basis for the oft-cited need for 0.1-dB uncertainty in radar measurements of Z_{dr} (e.g., Ryzhkov et al. 2005). Only 19 of the 100 sample estimates of Z_{dr} are within 0.1 dB of the population value, and there is a 19% chance of an error greater than 0.5 dB. Only seven of the samples gave estimates of Z_{dr} within 0.1 dB and also of Z_H within 1 dB.

To examine the effect of sample sizes on the estimates, we next drew 100 similar random samples of sizes 200 drops and then 1000 drops. Samples of such sizes are common in surface disdrometer data collected at 1-min intervals. The expected decrease in the errors with increasing sample size is evident (Table 2), but even for 1000 drops the bias is still more than 0.1 dB and the RMS error is about 1/3 dB. The skewness is little diminished, with about three-quarters of the samples still yielding underestimates.

b. *Biases and errors in sample estimates of Z_{dr} for different population parameters*

We then repeated the simulations for sample sizes of 500 drops with ranges of values of μ and then D_m to explore the behavior of the sampling errors with narrower or wider DSDs. Figure 2 and Table 3 summarize the results for varying μ. The bias and errors diminish as μ increases and the DSD narrows; for values of μ of perhaps 5 or greater, the bias (and even the median error) becomes essentially negligible, while the RMS error also decreases—though remaining well in excess of 0.1 dB with this sample size. The skewness of the sampling distribution also becomes less pronounced. In the limit of a monodisperse distribution ($\mu \to \infty$), it would require only a single drop to estimate Z_{dr}.

Figure 3 and Table 4 summarize the results for varying D_m. The bias increases slowly and the RMS error more markedly as D_m increases and the DSD widens. The skewness of the sampling distribution varies little with the value of D_m, and the median error increases only slightly as D_m increases.

4. Biases and errors in estimates of other radar variables

a. *Baseline population—Effects of sample size*

Figure 4 and Table 5 summarize the results for the errors in the sample estimates of Z_H values, for samples of different sizes drawn from the baseline population. These error values also exhibit the typical skewness

TABLE 2. Effect of variations in sample size on biases and errors in sample estimates of Z_{dr}; $\mu = 2$, $D_m = 2$ mm.

Sample size	200	500	1000
Bias (dB)	−0.25	−0.15	−0.11
RMS error (dB)	0.51	0.46	0.34
Median error (dB)	−0.33	−0.25	−0.17
Underestimates (%)	81	74	77

TABLE 3. Effect of variations in μ on biases and errors in sample estimates of Z_{dr}; $D_m = 2$ mm, sample size 500 drops.

μ	0	2	5
Bias (dB)	−0.61	−0.15	−0.04
RMS error (dB)	0.80	0.46	0.26
Median error (dB)	−0.69	−0.25	−0.08
Underestimates (%)	92	74	67

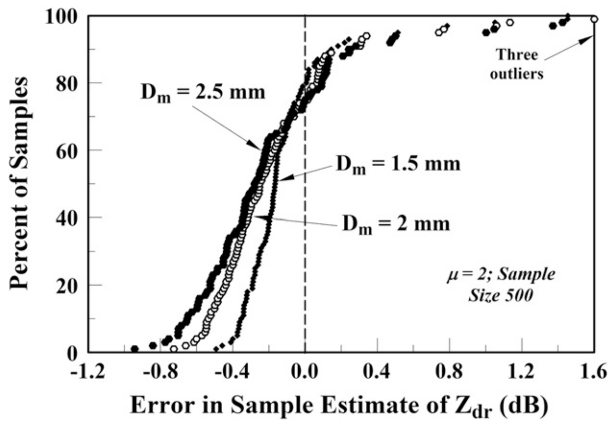

FIG. 3. Cumulative distributions of errors in sample estimates of Z_{dr} for different values of D_m. Samples from gamma DSD with $\mu = 2$; sample sizes 500 drops.

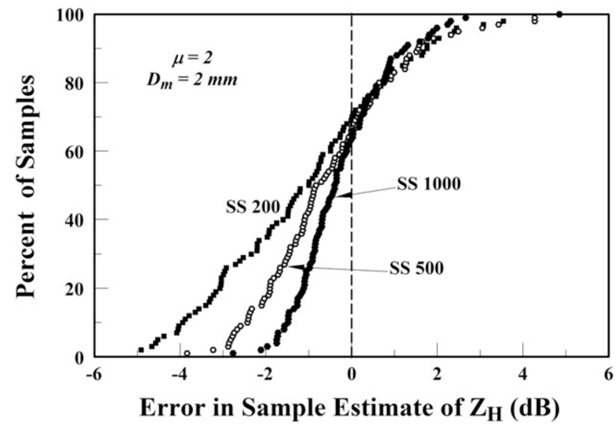

FIG. 4. Cumulative distributions of errors in sample estimates of Z_H for different sample sizes. Samples from baseline gamma DSD with parameters $\mu = 2$, $D_m = 2$ mm.

discussed in Smith et al. (1993), Uijlenhoet et al. (2006), and Smith et al. (2009). The errors again diminish with increasing sample size, but the RMS error exceeds 1 dB even with a sample size of 1000 drops. The median errors are in good agreement with the results of the simulation reported in Smith et al. (2009, their Fig. 1). With a sample size of 500 drops, there is a bias in the Z_H estimates of −0.46 dB; the RMS error is 1.8 dB. Only 41% of the estimates are within 1 dB of the population value, and there is a 6% chance of an error greater than 3 dB.

Figure 5 and Table 5 summarize the corresponding results for the errors in the sample estimates of K_{dp} values. These error values are also skewed, but to a lesser extent, and diminish with increasing sample size. The apparent bias is slight, less than 5% with a sample size of 200 drops and only 1% with a sample of 1000 drops; being small compared to the RMS error, the bias here might be considered negligible. The RMS error is more significant, at 15% even with a sample size of 1000 drops. With a sample size of 500 drops, the RMS error is 21%; only 32% of the estimates are within 10% of the population value, and there is a 23% chance of an error greater than 25%.

b. Effects of varying population shape parameter

As the gamma shape parameter μ increases, the drop size distribution narrows and the relative importance of the large drops decreases. Consequently, the errors in sample estimates of some variables strongly impacted by the large drops also decrease. Figure 6 and Table 6 illustrate this behavior for the estimates of Z_H with $D_m = 2$ mm and a sample size of 500 drops. With the widest distribution considered ($\mu = 0$), there is a bias of more than −2 dB and the RMS error exceeds 3 dB. With the narrowest distribution ($\mu = 5$), most of the estimates (69%) are within 1 dB of the population value; the bias is less than 0.25 dB and the RMS error only slightly exceeds 1 dB. Samples from populations with still larger values of μ would give even smaller errors.

The situation with respect to the effect of μ on sample estimates of K_{dp} is similar (Fig. 7; Table 6). With the widest distribution, there is a bias of −10% and the RMS error is 45%. For values of $\mu = 2$ and larger, the bias has essentially disappeared. With $\mu = 5$ the RMS error is down to 12%, and 58% of the estimates are within 10% of the population value.

c. Effects of varying mass-weighted mean diameter

The effect of variation in D_m on the drop size distribution is the reverse of that for μ, in that as D_m increases

TABLE 4. Effect of variations in D_m on biases and errors in sample estimates of Z_{dr}; $\mu = 2$, sample size 500 drops.

D_m (mm)	1.5	2.0	2.5
Bias (dB)	−0.09	−0.15	−0.16
RMS error (dB)	0.34	0.46	0.56
Median error (dB)	−0.16	−0.25	−0.27
Underestimates (%)	79	74	74

TABLE 5. Effect of variations in sample size on biases and errors in sample estimates of Z_H and K_{dp}; $\mu = 2$, $D_m = 2$ mm.

Sample size		200	500	1000
Bias	Z_H (dB)	−0.93	−0.46	−0.21
	K_{dp} (%)	−4.5	−2	1
RMS error	Z_H (dB)	2.78	1.81	1.20
	K_{dp} (%)	34	21	15
Median error	Z_H (dB)	−1.00	−0.77	−0.38
	K_{dp} (%)	−10	−5	−2
Underestimates (%)	Z_H	69	66	63
	K_{dp}	63	60	57

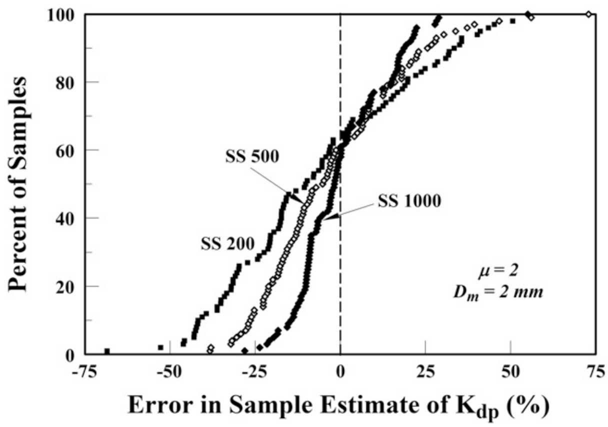

FIG. 5. Cumulative distributions of errors in sample estimates of K_{dp} for different sample sizes. Samples from baseline gamma DSD with parameters $\mu = 2$, $D_m = 2$ mm.

the DSD widens and the numbers of large drops increase. Consequently, one might anticipate the errors in sample estimates of variables strongly impacted by the large drops, such as reflectivity, to increase as well. However, Smith et al. (2009) showed that (1) can be reformulated in terms of the dimensionless ratio D_{eq}/D_m. Then if the reflectivity values are calculated using $(D_{eq})^6$, the sampling statistics are actually independent of D_m. The contributions of large drops to Z_H differ somewhat from $(D_{eq})^6$, but as Fig. 8 shows, the variation of the error distribution with D_m is almost indiscernible. The bias (around −0.4 dB; Table 7), the RMS error (around −1.8 dB), and the median error vary little as D_m increases from 1.5 to 2.5 mm.

The behavior with respect to the effect of D_m on sample estimates of K_{dp} is not greatly different (Table 7). With $\mu = 2$ and samples of 500 drops, the bias is small and varies little as D_m increases from 1.5 to 2.5 mm; the RMS error (around 20%) decreases slightly, as does the median error.

TABLE 6. Effect of variations in μ on biases and errors in sample estimates of Z_H and K_{dp}; $D_m = 2$ mm, sample size 500 drops.

μ		0	2	5
Bias	Z_H (dB)	−2.06	−0.46	−0.22
	K_{dp} (%)	−10	−2	−2
RMS error	Z_H (dB)	3.65	1.81	1.06
	K_{dp} (%)	45	21	12
Median error	Z_H (dB)	−2.12	−0.77	−0.18
	K_{dp} (%)	−17	−5	−3
Underestimates (%)	Z_H	75	66	59
	K_{dp}	71	60	54

5. Other findings of interest

Exploratory analysis of the various simulation results has elucidated some other sampling issues that arise in estimating radar variables from disdrometer data. These include the significance of the largest drop in a sample (D_{max}) and the correlations among the errors in the estimates of different variables.

a. The significance of D_{max}

The estimates of the radar variables and, consequently, the errors in those estimates, are strongly correlated with the size of the largest drop in a sample (e.g., Fig. 9). In this example compositing the simulation results for the baseline population with $\mu = 2$, $D_m = 2$ mm, and three different sample sizes, the nonlinear correlation coefficient between the error in the estimate of Z_{dr} and the sample D_{max} is 0.94. The coefficients for the individual sample sizes are slightly higher (0.96–0.98), and there are small differences in the parameters of the fitted

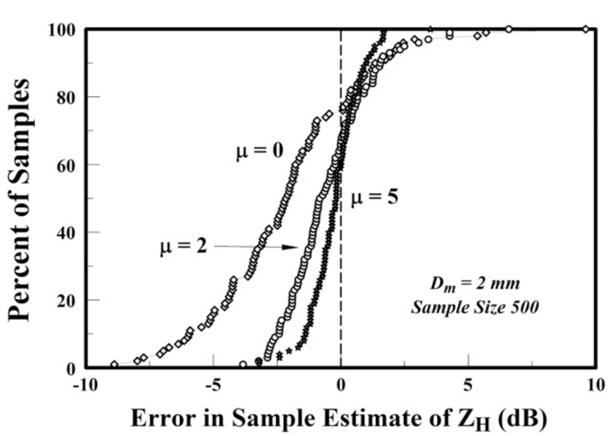

FIG. 6. Cumulative distributions of errors in sample estimates of Z_H for different values of μ. Samples from gamma DSD with $D_m = 2$ mm; sample sizes 500 drops.

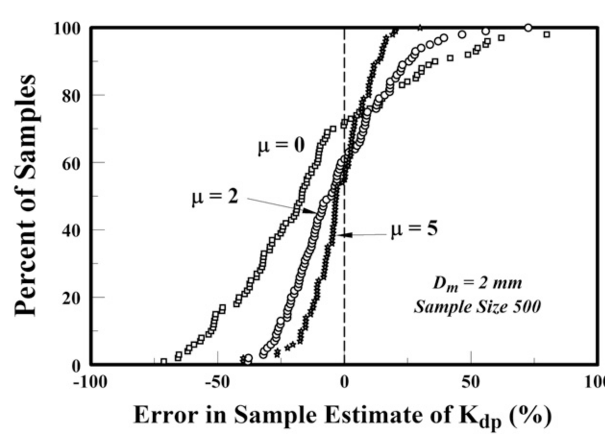

FIG. 7. Cumulative distributions of errors in sample estimates of K_{dp} for different values of μ. Samples from gamma DSD with $D_m = 2$ mm; sample sizes 500 drops.

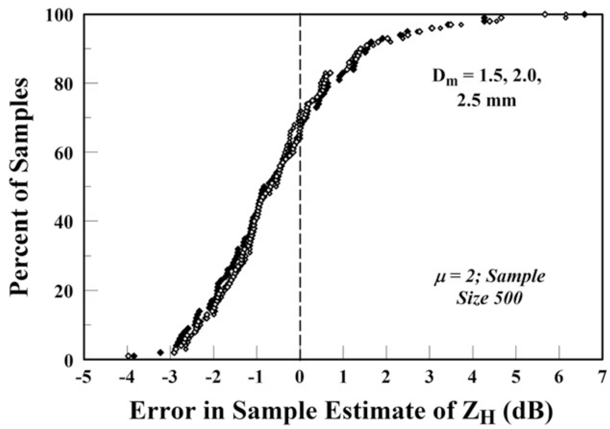

FIG. 8. Cumulative distributions of errors in sample estimates of Z_H for different values of D_m. Samples from gamma DSD with $\mu = 2$; sample sizes 500 drops.

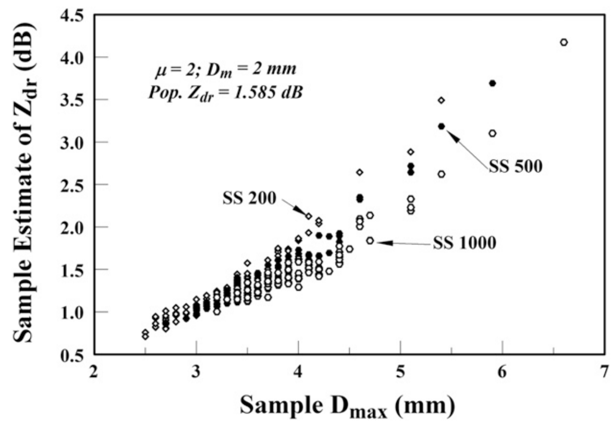

FIG. 9. Composite plot of sample estimates of Z_{dr} vs sample D_{max} for three different sample sizes. Samples from baseline gamma DSD with population parameters $\mu = 2$, $D_m = 2$ mm.

relationships. Interestingly, Carey and Petersen (2015) found similar importance of D_{max} in determining a relationship between Z_{dr} and D_0, the median volume diameter.

The drop size resolution used in these simulations is 0.1 mm, and this strong correlation suggests that there would be value in determining the sizes of the large drops to similar resolution. Unfortunately, some commonly used disdrometers categorize the large drops in size bins of widths as great as 1 mm (e.g., Yuter et al. 2006).

There are similar, though weaker, nonlinear correlations between the errors in the estimates of Z_H and K_{dp} and the sample D_{max} values. The correlation coefficient for the baseline population $\mu = 2$, $D_m = 2$ mm, and the same three different sample sizes is 0.87 for the Z_H estimates and 0.73 for the K_{dp} estimates.

Returning to the Z_{dr} estimates, with the population $D_m = 2$ mm and a sample size of 500 drops, there is similar correlation between those estimates and errors and the sample D_{max} values for different values of μ (Fig. 10). However, the correlation becomes somewhat weaker as μ increases and the DSD narrows. Interestingly, most of the differences in the errors for different values of μ are in samples with values of D_{max} less than about 4 mm, that is, less than twice the population D_m value. To examine the situation for different values of D_m with $\mu = 2$ and a sample size of 500 drops, we have to normalize the D_{max} scale by plotting D_{max}/D_m. Having done that, there is some difference in the scatter of points for different values of D_m; for the composite plot (Fig. 11) the nonlinear correlation coefficient is 0.95. For the individual values of D_m, the nonlinear correlations are actually higher (0.98–0.99). The plot appears to suggest that samples with $D_{max}/D_m \approx 2$ would give relatively small errors in the estimates of Z_{dr}; however, that ratio varies for other values of μ.

Finally, looking at the effects of varying μ on the relationship between the errors in the estimates of Z_H and the sample D_{max} we find for the case $D_m = 2$ mm and a sample size of 500 drops a nonlinear correlation coefficient

TABLE 7. Effect of variations in D_m on biases and errors in sample estimates of Z_H; $\mu = 2$, size 500 drops.

D_m (mm)		1.5	2.0	2.5
Bias	Z_H (dB)	−0.39	−0.46	−0.42
	K_{dp} (%)	−1.5	−2	−2
RMS error	Z_H (dB)	1.79	1.81	1.73
	K_{dp} (%)	24	21	18
Median error	Z_H (dB)	−0.54	−0.77	−0.66
	K_{dp} (%)	−6	−5	−4
Underestimates (%)	Z_H	71	66	66
	K_{dp}	62	60	60

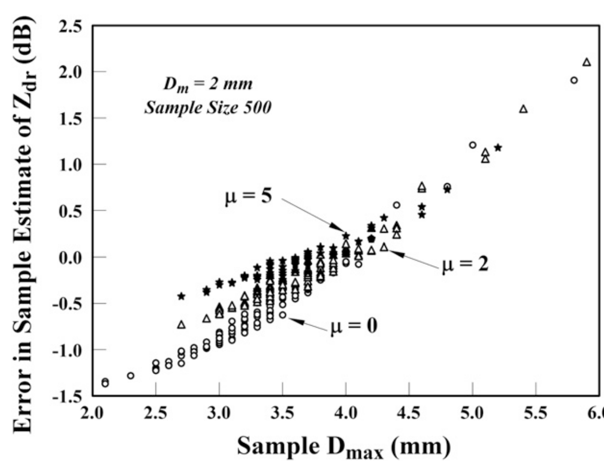

FIG. 10. Composite plot of errors in sample estimates of Z_{dr} vs sample D_{max} for three different values of μ. Samples from gamma DSD with population $D_m = 2$ mm; sample sizes 500 drops.

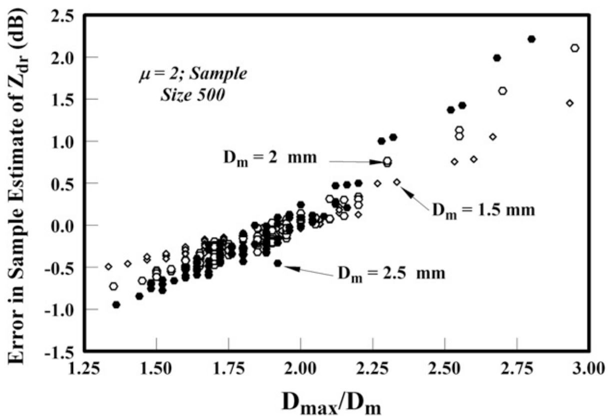

FIG. 11. Composite plot of errors in sample estimates of Z_{dr} vs ratio of sample D_{max} to D_m for three different values of D_m. Samples from gamma DSD with population $\mu = 2$; sample sizes 500 drops.

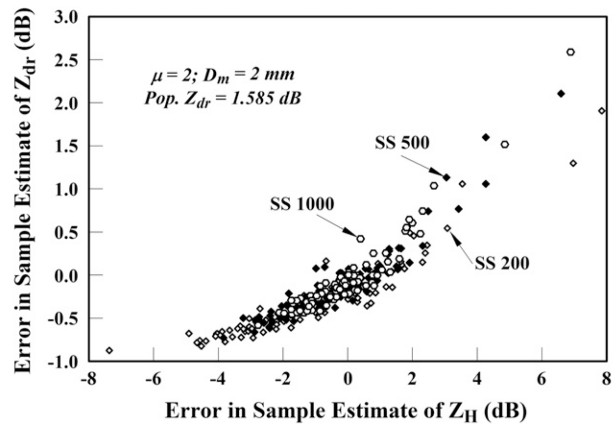

FIG. 12. Composite plot of errors in sample estimates of Z_{dr} vs errors in sample estimates of Z_H for three different sample sizes. Samples from baseline gamma DSD with population parameters $\mu = 2$, $D_m = 2$ mm.

of 0.91. For the case $\mu = 2$ and a sample size of 500 drops, the same correlation applies between the errors in the Z_H estimates and the D_{max}/D_m ratio. With respect to errors in the estimates of K_{dp}, the corresponding coefficients are 0.79 and 0.78, respectively.

b. Correlations among the errors in the estimates

Analysis of the results of the sampling simulations highlights the fact that the errors in the estimates of any pair of the three radar variables (Z_H, Z_{dr}, or K_{dp}) are correlated. This is not surprising, because as illustrated in the preceding section the estimates of all these variables, and therefore the errors involved, depend strongly on the largest drop(s) in the samples. Figure 12 illustrates this correlation between the errors in the estimates of Z_{dr} and Z_H for the baseline case $\mu = 2$ and $D_m = 2$ mm; the nonlinear correlation coefficient is 0.94. The corresponding error correlation coefficients are 0.97 for K_{dp} and Z_H, and 0.81 for Z_{dr} and K_{dp}.

Turning to the behavior of the error correlations when μ varies ($D_m = 2$ mm and sample sizes of 500 drops), we find the composite error correlation coefficients to be essentially identical to those listed in the preceding paragraph, for the three μ values considered. One difference noted is that the correlation coefficient for the errors in the Z_{dr} and K_{dp} estimates decreases from 0.90 when $\mu = 0$ to 0.84 when $\mu = 2$ and 0.73 when $\mu = 5$.

When D_m varies from 1.5 to 2.5 mm ($\mu = 2$ and sample sizes of 500 drops), the error scatterplots are quite similar. The error correlation statistics for the Z_{dr} –Z_H and K_{dp}–Z_H cases are virtually identical to those noted in the preceding paragraph for variations of μ. For the Z_{dr} –K_{dp} case, the error correlation coefficients fall within the same range but there is less variation with D_m.

The importance of these error correlations lies in their impact upon efforts to determine relationships among the radar variables and others from disdrometer data. With error correlations as strong as those seen here, the regression analyses used to identify such relationships should be using regression procedures that take account of the error covariances.

6. The sample size challenge

All three radar variables (Z_H, Z_{dr}, and K_{dp}) depend to a significant degree on the large drops in the population, and as shown in section 5a the estimates of those quantities likewise depend upon the large drops in the samples. Figure 13 shows the cumulative fraction (linear values) of the population variables Z_H and K_{dp} (no similar cumulative linear value exists for Z_{dr}) versus drop size for the baseline population DSD with $\mu = 2$ and $D_m = 2$ mm (the curves represent integrals over the analytical DSD function and in effect assume an infinite sample size). A similar curve for the rainfall rate is included for reference.[2] The curves indicate that well over half the values of these radar variables are contributed by drops larger than D_m (e.g., 85% for Z_H). However, those large drops are relatively rare; only about 6% of the drops in a population with $\mu = 2$ are larger than D_m and even larger drops are the primary contributors to the radar variables. The dashed vertical lines in the figure indicate the drop sizes for which only one drop in

[2] That the K_{dp} curve lies closer to the R curve indicates why K_{dp} measurements are more useful than reflectivity measurements in estimating rainfall rates.

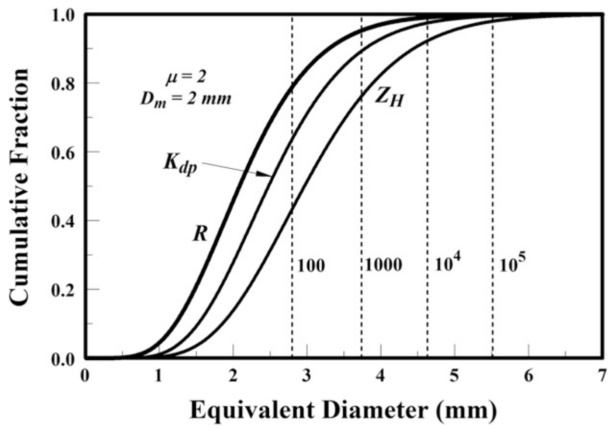

FIG. 13. Graph showing the cumulative fraction of rainfall rate R, K_{dp}, and Z_{H} for the baseline gamma DSD with parameters $\mu = 2$, $D_m = 2$ mm. Vertical dashed lines indicate drop sizes for which only one drop in 100, 1000, etc. is present in the population.

100, 1000, . . . in the population is larger, and provide a further indication that a sample of a few hundred drops is far too small to provide adequate estimates of the radar variables considered here.

Another illustration reinforces the importance of large sample sizes in estimating the radar variables. With the baseline $\mu = 2$ and $D_m = 2$ mm, the differential reflectivity-weighted mean drop diameter is 4.43 mm (Table 1); only 0.017% of the drops in the population (about 1 drop in 5763) are larger. The probability of having even a single larger drop in a sample of 200 is only 3.4%, and even that would not yield a good estimate of Z_{dr}.

7. Discussion and conclusions

The results of these raindrop sampling simulations illustrate the biases and errors that can occur in trying to estimate polarimetric radar variables from disdrometer data. The most serious difficulties arise with the estimates of differential reflectivity, for which biases and RMS errors greater than 0.5 dB were noted in some of the simulation results. Less important, but still significant, biases and errors appeared in some of the estimates of reflectivity factor and specific differential phase.

The greatest problems occur with wide drop size distributions (DSDs), including relatively large proportions of large drops, and these are the situations where the polarimetric variables are thought to be most useful. In addition, small sample sizes (e.g., a few hundred drops) lead to difficult problems with the estimates of the radar variables unless the DSDs are rather narrow (e.g., with a gamma distribution shape parameter greater than 5). The difficulties are greater for lighter rainfall rates, where accumulating a large sample is a challenge; increasing the sampling interval runs the risk of encountering inhomogeneities in the drop population. The simulation results provide mean biases and RMS errors, but they cannot provide specific values (such as "the RMS error in the estimate of Z_{dr} for this sample is 0.XY dB") for a given sample. Instead they give broader indications, such as "a sample of 200 drops with a fairly broad size distribution is likely to give an RMS error of 0.5 dB or more in the estimate of Z_{dr}" (cf. Tables 2–4).

Estimates of these radar variables from disdrometer data provide the basis for development of polarimetric rainfall-rate estimators and part of the basis for hydrometeor classification algorithms. Except for setting some minimum threshold (usually far less than 200 drops), the sample sizes and these sampling issues are seldom considered in the development of such algorithms [a notable exception appears in Zhang et al. (2003)]. The present results indicate that samples of a few hundred drops can often be essentially worthless for estimating the variables with the accuracy typically desired; for example, the NEXRAD specifications call for an accuracy of 1 dB in Z and 0.1 dB in Z_{dr}. Greater attention should be paid to the sample sizes and related biases and errors in the DSD-derived values used in developing such algorithms.

The simulations employed a gamma distribution model for the raindrop size distributions, with limited ranges of the population parameters. The results are therefore not universally applicable, but they do demonstrate that significant biases and substantial errors can occur in the estimates of the radar variables. These biases and errors can be mitigated only by using very large samples.

Acknowledgments. V. Chandrasekar provided guidance in the calculation of the radar variables for large drops. Support for publication of this work was provided through Navy Cooperative Agreement N00244-15-2-0009 with funding from the National Science Foundation.

APPENDIX

Relationship between Surface and Volume Samples

Use of volume samples in the simulations herein facilitates examining separately the effects of sample size and those of variations in the DSD parameters μ and D_m. Volume samples were collected from rain with the Illinois State Water Survey drop camera (Jones 1992), and airborne precipitation probes also collect volume samples.

However, much of the available DSD data comes from surface disdrometers and it is useful to relate the present results to such data. Johnson et al. (2014) outline a procedure for doing this.

As shown there, a volume DSD of form (1) would produce (using a very good approximation to empirical fall speed vs diameter data) a gamma surface DSD. The surface DSD would have the same general form:

$$n_s(D_{\mathrm{eq}}) = N_S \frac{(\mu_s + 4)^{\mu_s + 1}}{\mu_s!} \frac{D_{\mathrm{eq}}^{\mu_s}}{D_{ms}^{\mu_s + 1}} \exp[-(\mu_s + 4)D_{\mathrm{eq}}/D_{ms}].$$

Here the parameter N_s is the total number of drops per unit area per unit time. Johnson et al. (2014) show that the other parameters would be $\mu_s = \mu + 1$ and (after some algebra)

$$D_{ms} = \frac{(\mu + 5)D_m}{\mu + 4 + (0.193\,\mathrm{mm}^{-1})D_m}.$$

Thus, for example, the sampling statistics for volume samples from the baseline DSD with $\mu = 2$ and $D_m = 2$ mm would be applicable to surface samples from a surface population DSD with $\mu_s = 3$ and $D_{ms} = 2.19$ mm.

REFERENCES

Cao, Q., and G. Zhang, 2009: Errors in estimating raindrop size distribution parameters employing disdrometer and simulated raindrop spectra. *J. Appl. Meteor. Climatol.*, **48**, 406–425, doi:10.1175/2008JAMC2026.1.

Carey, L. D., and W. A. Petersen, 2015: Sensitivity of C-band polarimetric radar–based drop size estimates to maximum diameter. *J. Appl. Meteor. Climatol.*, **54**, 1352–1371, doi:10.1175/JAMC-D-14-0079.1.

Cornford, S. G., 1967: Sampling errors in measurements of raindrop and cloud droplet concentrations. *Meteor. Mag.*, **96**, 271–282.

Gertzman, H. S., and D. Atlas, 1977: Sampling errors in the measurement of rain and hail parameters. *J. Geophys. Res.*, **82**, 4955–4966, doi:10.1029/JC082i031p04955.

Johnson, R. W., D. V. Kliche, and P. L. Smith, 2014: Maximum likelihood estimation of gamma parameters for coarsely binned and truncated raindrop size data. *Quart. J. Roy. Meteor. Soc.*, **140**, 1245–1256, doi:10.1002/qj.2209.

Jones, D. M. A., 1992: Raindrop spectra at the ground. *J. Appl. Meteor.*, **31**, 1219–1225, doi:10.1175/1520-0450(1992)031<1219:RSATG>2.0.CO;2.

Joss, J., and A. Waldvogel, 1969: Raindrop size distribution and sampling size errors. *J. Atmos. Sci.*, **26**, 566–569, doi:10.1175/1520-0469(1969)026<0566:RSDASS>2.0.CO;2.

Kliche, D. V., 2007: Raindrop size distribution functions: An empirical approach. Ph.D. dissertation, South Dakota School of Mines and Technology, 211 pp.

Kostinski, A. B., and A. R. Jameson, 1997: Fluctuation properties of precipitation. Part I: On deviations of single-size drop counts from the Poisson distribution. *J. Atmos. Sci.*, **54**, 2174–2186, doi:10.1175/1520-0469(1997)054<2174:FPOPPI>2.0.CO;2.

Ryzhkov, A. V., S. E. Giangrande, V. M. Melnikov, and T. J. Schuur, 2005: Calibration issues of dual-polarization radar measurements. *J. Atmos. Oceanic Technol.*, **22**, 1138–1155, doi:10.1175/JTECH1772.1.

Smith, P. L., and D. V. Kliche, 2005: The bias in moment estimators for parameters of drop-size distribution functions: Sampling from exponential distributions. *J. Appl. Meteor.*, **44**, 1195–1205, doi:10.1175/JAM2258.1.

——, Z. Liu, and J. Joss, 1993: A study of sampling-variability effects in raindrop size observations. *J. Appl. Meteor.*, **32**, 1259–1269, doi:10.1175/1520-0450(1993)032<1259:ASOSVE>2.0.CO;2.

——, D. V. Kliche, and R. W. Johnson, 2009: The bias and error in moment estimators for parameters of drop size distribution functions: Sampling from gamma distributions. *J. Appl. Meteor. Climatol.*, **48**, 2118–2126, doi:10.1175/2009JAMC2114.1.

Uijlenhoet, R., J. M. Porra, D. Sempere Torres, and J.-D. Creutin, 2006: Analytical solutions to sampling effects in drop size distribution measurements during stationary rainfall. *J. Hydrol.*, **328**, 65–82, doi:10.1016/j.jhydrol.2005.11.043.

Yuter, S. E., D. E. Kingsmill, L. B. Nance, and M. Löffler-Mang, 2006: Observations of precipitation size and fall speed characteristics within coexisting rain and wet snow. *J. Appl. Meteor. Climatol.*, **45**, 1450–1464, doi:10.1175/JAM2406.1.

Zhang, G., J. Vivekanandan, E. A. Brandes, R. Meneghini, and T. Kozu, 2003: The shape–slope relation in observed gamma raindrop size distributions: Statistical error or useful information? *J. Atmos. Oceanic Technol.*, **20**, 1106–1119, doi:10.1175/1520-0426(2003)020<1106:TSRIOG>2.0.CO;2.