A Study of Sampling-Variability Effects in Raindrop Size Observations

PAUL L. SMITH AND ZHONG LIU*

Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, South Dakota

JURG JOSS

Swiss Meteorological Institute, Locarno-Monti, Switzerland

(Manuscript received 21 October 1991, in final form 4 January 1993)

ABSTRACT

Because of the randomness associated with sampling from a population of raindrops, variations in the data reflect some undetermined mixture of sampling variability and inhomogeneity in the precipitation. Better understanding of the effects of sampling variability can aid in interpreting drop size observations. This study begins with a Monte Carlo simulation of the sampling process and then evaluates the resulting estimates of the characteristics of the underlying drop population. The characteristics considered include the liquid water concentration and the reflectivity factor; the maximum particle size in each sample is also determined. The results show that skewness in the sampling distributions when the samples are small (which is the usual case in practice) produces a propensity to underestimate all of the characteristic quantities. In particular, the distribution of the sample maximum drop sizes suggests that it may be futile to try to infer an upper truncation point for the size distribution on the basis of the maximum observed particle size.

Resulting paired values, for example, of Z and W for repeated sampling, were plotted on the usual type of log–log scatterplots. This yielded quite plausible-looking Z–R and Z–W relationships even though the parent drop population (and, hence, the actual values of the quantities) was unchanging; the "relationships" arose entirely from the sampling variability. Moreover, if the sample size is small, the sample points are shown to be necessarily displaced from the point corresponding to the actual population values of the variables. Consequently, any assessment of the "accuracy" of a Z–R relationship based on drop size data should include some consideration of the numbers of drops involved in the samples making up the scatterplot.

1. Introduction

The measurement of hydrometeor characteristics such as number concentrations, sizes, and size distributions is important to the investigation of precipitation physics. Determination of bulk quantities such as liquid water concentration (LWC) or radar reflectivity factor, and of the relationships between pairs of such variables, is often a part of such studies. Because of the randomness associated with sampling from a population of raindrops, such observations typically involve some undetermined mixture of sampling variability and inhomogeneity in the precipitation.

Establishing a clear distinction between sampling variability in the observations and inhomogeneity in the precipitation is very difficult in any practical situation, partly because it is difficult to determine what constitutes homogeneous precipitation. Suppose one postulates the existence of a region of the atmosphere containing precipitation homogeneous in the sense that the underlying raindrop size distribution is fixed, while the sizes and locations of individual drops within the region vary at random. One can envision taking an increasingly large sample of the drops from within this region, and the result ideally would asymptotically approach the underlying drop size distribution. Then differences among any smaller samples taken from different subregions manifest the effects of sampling variability.

Now suppose the same sampling process could be repeated for an adjacent region, also homogeneous in the above sense but possibly involving a different underlying drop size distribution. If the same asymptotic distribution (within the limits determined by the sample size and drop statistics) were found, then the precipitation would be considered homogeneous over the combined region; if not, it would be considered inhomogeneous. In either case, individual smaller samples from subregions of the second region would differ among themselves, and the range of variation would

* Current affiliation: Byrd Polar Research Center, Ohio State University, Columbus, Ohio.

Corresponding author address: Paul L. Smith, Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, 501 E. St. Joseph Street, Rapid City, SD 57701-3995.

overlap to a greater or lesser degree the range of variation of small samples from the first region.

In a practical situation, one may have two observed samples (which may or may not be contiguous in space and time) and wish to know whether they represent subregions from a region of homogeneous precipitation (in which case, any difference is due entirely to sampling variability) or two different regions of inhomogeneous precipitation (in which case, the difference represents a mixture of sampling variability and the inhomogeneity). No general answer to such a question is available, but it seems clear that added information about the effects of sampling variability in a situation with assumed homogeneous precipitation would aid in interpreting differences among observed raindrop samples. Thus, better knowledge of the uncertainties associated with sampling from a fixed population of raindrops would be useful.

Estimation of these uncertainties requires understanding of the statistics of the particle-sampling process. Several studies of the statistics of hydrometeor sampling have been reported (e.g., Cornford 1967; Joss and Waldvogel 1969; Gertzman and Atlas 1977). They generally start with an assumed (usually exponential) size distribution for the population of particles and use Poisson statistics to determine the sampling variations for such quantities as the LWC. The results are typically expressed in the form of sampling standard deviations.

This new study approaches the question from a different point of view by first performing a Monte Carlo simulation of the sampling process (on a particle-by-particle basis) and then evaluating the resulting estimates of the population characteristics. The characteristics considered include the LWC, R^* (defined as the fourth moment of the drop size distribution, a proxy quantity for the rainfall rate as introduced in Joss and Gori 1978), the reflectivity factor Z, and the mass-weighted mean particle diameter; the maximum particle size in each sample is also determined. The study examines the sampling distributions for these quantities, and their behavior as the sample size varies. The results complement the earlier ones by showing the effect of skewness in the sampling distributions when the samples are small (which is frequently the case in practice).

This approach also permits investigation of the effects of sampling variability on the inferred relationships between pairs of these variables—something not treated in most of the earlier studies. Mueller and Sims (1966) did note the inherent correlation between values of R and Z determined from measured drop size distributions. Richards and Crozier (1983) presented data indicating the presence of such a correlation between R and Z values determined from drop size observations, which was not reflected with concurrent Z values determined with a radar system. Chandrasekar and Bringi (1987) considered this correlation explicitly, using an assumed gamma drop size distribution function. However, their analysis determined the correlation for linear-scale plots of Z versus R, and not for the more familiar log–log plots.

2. Simulation procedure

We begin with an assumed exponential size distribution for the population of drops and use Poisson statistics to determine the number of drops in any given sample volume. The former assumption follows Marshall and Palmer (1948), while the latter follows Sasyo (1965). Lovejoy and Schertzer (1990) have recently suggested fractal behavior in connection with the latter issue; this might mean that a distribution broader than the Poisson should be used for simulating the sample numbers. However, with increasing sample size, the observed drop size distributions tend to become more nearly exponential (Joss and Gori 1978). We do not expect the substantive findings of this paper to change significantly with other, broader-type distributions, so no other type has been investigated.

a. Normalization procedure and equations

The Marshall–Palmer exponential size distribution function $n(D) = n_0 e^{-\Lambda D}$ may be written as

$$n(D) = \left(\frac{4N_t}{D_m}\right) e^{-4D/D_m}$$

with

$$D_m = \frac{4}{\Lambda}; \quad N_t = \frac{n_0}{\Lambda} = \frac{n_0 D_m}{4}. \tag{1}$$

Here D_m is the mass-weighted mean particle diameter and N_t is the total drop concentration.

The population value Q_p for any bulk quantity Q expressible as a function of D can be derived from (1):

$$Q_p = \int_0^\infty Q(D)n(D)dD = \int_0^\infty Q(D)\left(\frac{4N_t}{D_m}\right) e^{-4D/D_m} dD. \tag{2}$$

The quantities of interest here are expressible as power-law functions of D; that is, $Q(D) = cD^n$ (e.g., $n = 3$ for liquid water concentration, $n = 6$ for radar reflectivity factor). If the population is sampled by observing the particles in some sampling volume V_s, the sample value of the quantity is

$$Q_s = \frac{\Sigma Q(D)}{V_s}. \tag{3}$$

The summation extends over all the particles in the sample, and the sample volume is here assumed to be independent of D. The ratio of sample to population value is

$$\frac{Q_s}{Q_p} = \left[\frac{\Sigma Q(D)/V_s}{\int_0^\infty Q(D)(4N_t/D_m)e^{-4D/D_m}dD}\right] = \left[\frac{\Sigma Q(D)}{4\int_0^\infty Q(D)(N_tV_s)e^{-4D/D_m}d(D/D_m)}\right]. \quad (4)$$

The quantity N_tV_s is the expected total number of particles $\langle C_t\rangle$ in the sampling volume V_s. Thus the simulation can be organized in terms of values of $\langle C_t\rangle$ and the normalized particle diameter D/D_m.

b. Quantities calculated

1) CHARACTERISTIC SIZES

The simplest quantity to determine in a sample is the maximum drop-size $D_{\max}$, expressed here in normalized form as $D_{\max}/D_m$. For the mass-weighted mean diameter, one notes that the population value is defined as

$$D_m = \frac{\int_0^\infty n(D)D^4dD}{\int_0^\infty n(D)D^3dD} = \frac{4}{\Lambda} \quad (5)$$

in agreement with (1). The sample estimate is

$$(D_m)_s = \frac{\Sigma D^4/V_s}{\Sigma D^3/V_s} = \frac{\Sigma D^4}{\Sigma D^3}. \quad (6)$$

Thus, the normalized estimate is

$$\frac{(D_m)_s}{D_m} = \frac{\Lambda\Sigma D^4}{4\Sigma D^3} = \frac{\Sigma(4D/D_m)^4}{4\Sigma(4D/D_m)^3}. \quad (7)$$

Introducing the normalized size variate $x = 4D/D_m$, (7) may be written as

$$\frac{(D_m)_s}{D_m} = \frac{\Sigma x^4}{4\Sigma x^3}. \quad (8)$$

2) BULK QUANTITIES (LWC, R^*, Z)

For the liquid-water concentration $Q(D)$ is $\pi\rho D^3/6$, ρ being the density of water (or of ice, if appropriate). The population value obtained from (2) is

$$W_p = \frac{\pi\rho N_tD_m^3}{64}. \quad (9)$$

From (3), the sample value is

$$W_s = \frac{\pi\rho\Sigma D^3}{6V_s}. \quad (10)$$

Thus the normalized value is

$$\frac{W_s}{W_p} = \frac{64\Sigma D^3}{6N_tV_sD_m^3} = \frac{\Sigma x^3}{6\langle C_t\rangle}. \quad (11)$$

A similar analysis leads to

$$R_p^* = \frac{\pi N_tD_m^4}{64}; \quad R_s^* = \left(\frac{\pi}{6V_s}\right)\Sigma D^4 \quad (12)$$

and

$$\frac{R_s^*}{R_p^*} = \frac{\Sigma x^4}{24\langle C_t\rangle} \quad (13)$$

for the "normalized rainfall rate," and

$$Z_p = \frac{45N_tD_m^6}{256}; \quad Z_s = \frac{1}{V_s}\Sigma D^6 \quad (14)$$

and

$$\frac{Z_s}{Z_p} = \frac{\Sigma x^6}{720\langle C_t\rangle} \quad (15)$$

for the reflectivity factor.

Note from (11), (13), and (15) that these results of the simulations will be expressible in terms of just the expected number of particles in the sample.

c. Two-stage Monte Carlo simulation method

A two-stage Monte Carlo method was used for the simulations. If the drops are Poisson-distributed in space, the drop count C_t in a sample volume V_s is also Poisson-distributed with expected value $\langle C_t\rangle$. A look-up table based on $\langle C_t\rangle$ and the associated distribution was used to determine the C_t value for each sample as the random number generator produced numbers in the first stage of the simulation.

To determine the specific sizes for the C_t drops in the sample, another look-up table expressed in terms of the normalized diameter variate x and based on the assumed Marshall–Palmer exponential size distribution function was used. This table was truncated at $x = 10$ ($D = 2.5D_m$) because particles larger than $2.5D_m$ are rarely observed in nature. [Such truncation reduces the population values slightly from those given in (5), (9), (12), and (14).] The particle sizes were determined in the second stage by drawing exactly C_t random numbers and referring to this table.

After these two stages, the (normalized) sample values for $D_{\max}$, $(D_m)_s$, LWC, R^*, and Z can be determined. This process was repeated over 1000 or more runs for each chosen value of $\langle C_t\rangle$. Runs have been made for $\langle C_t\rangle$ = 10, 20, 50, 100, 200, 500, and 800 drops. For a basis of comparison, we note that the literature contains examples of sample values based on single drops; typical 1-s samples from airborne Particle Measuring Systems, Inc. (PMS) optical array probes

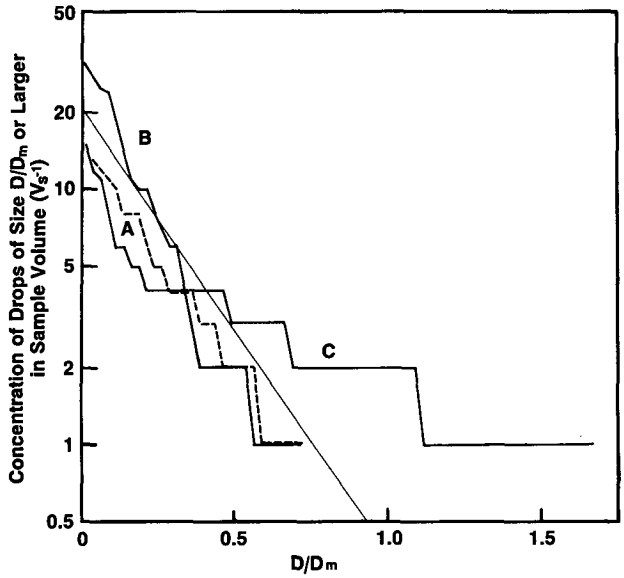

FIG. 1. Examples of "observed" drop size distributions from simulation for $\langle C_t \rangle = 20$. Thin line shows population distribution $\langle C_t \rangle \exp(-4D/D_m)$. A: $C_t = 13$ drops; B: $C_t = 31$; C: $C_t = 15$.

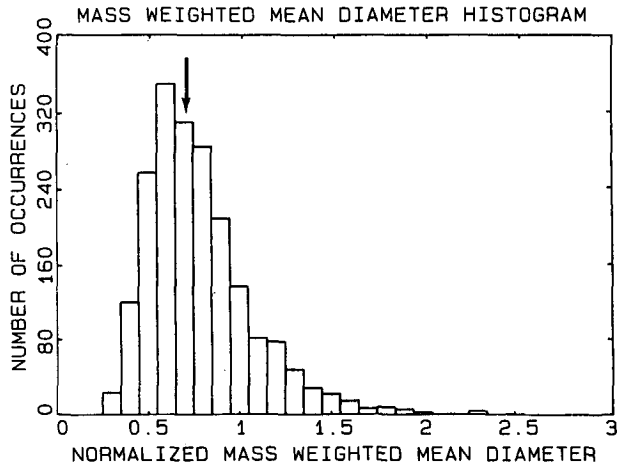

FIG. 2. Histogram of sample values of the (normalized) mass-weighted mean particle diameter, for sample volumes yielding an expected $\langle C_t \rangle = 20$ drops. The population value D_m falls at 1.0 on the abscissa; arrow indicates median sample value.

involve of the order of 100 drops; and 1-m^3 samples from intense subtropical rains involve many thousands of drops.

Figure 1 shows three examples of "observed" drop size distributions from the simulation for $\langle C_t \rangle = 20$. Some small samples (A in the figure) approximate the population distribution fairly well, although the absence of any large particles causes this one to underestimate the population LWC ($W_s/W_p = 0.40$). Other samples (B) also appear to be roughly exponential but do not conform well to either the population size distribution or the LWC; here $W_s/W_p = 0.39$, so the greater number of small drops in this sample contributes essentially nothing to the value of W_s. For samples of this size, the average number of drops larger than D_m is less than 0.4, so a sample that does contain one or two such drops (C in the figure) gives a greatly distorted view of both the size distribution and the LWC (here $W_s = 3.35W_p$).

3. Sampling distributions

Specific sampling distributions are shown for $\langle C_t \rangle = 20$, and the behavior of sample median values is then presented as a function of expected sample size.

a. Characteristic particle sizes

Figure 2 shows the sampling distribution of mass-weighted mean particle diameter $(D_m)_s$ for $\langle C_t \rangle = 20$. The distribution is markedly skewed toward low values; the median of $(D_m)_s$ is about two-thirds of the population value (1.0 on the normalized scale), and about 83% of the values fall below the population value. The most common observational result will therefore be an underestimate of the true population D_m. As the (expected) sample size increases, the median estimate $(D_m)_s$ also increases (Fig. 3), but it takes a sample volume yielding a $\langle C_t \rangle$ of more than 100 drops for the median estimate to reach 90% of the population value.

Figure 4 shows the frequency distribution of maximum observed drop size from each sample for $\langle C_t \rangle = 20$. Although $D_{\max} = 2.5D_m$ for the population was set in the simulation, the median in Fig. 4 is only $0.86D_m$. In other words, the median maximum drop size in the samples is only about one-third of the true

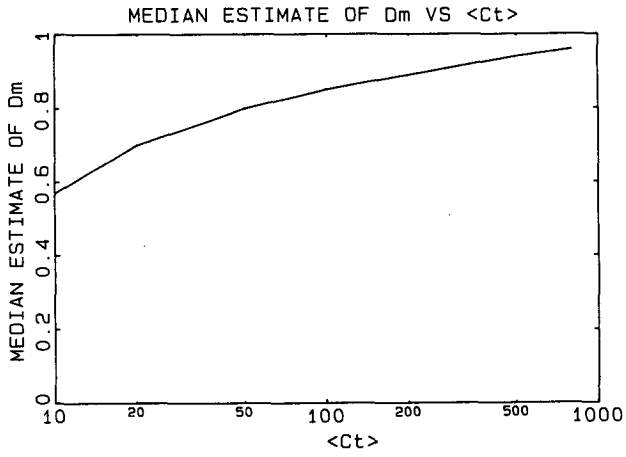

FIG. 3. Variation of median sample estimate of mass-weighted mean particle diameter with expected sample size $\langle C_t \rangle$.

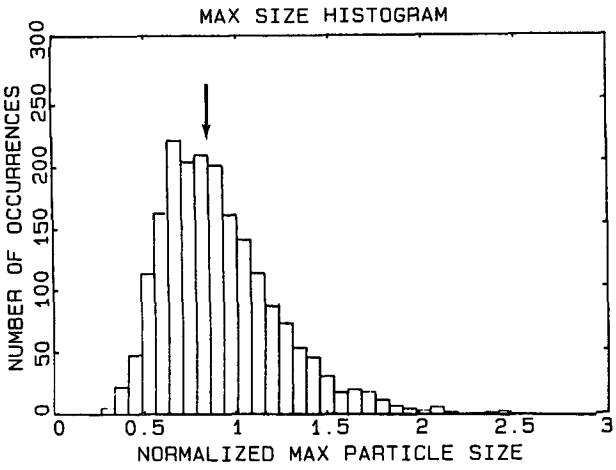

FIG. 4. Histogram of sample values of the (normalized) maximum particle diameter, for sample volumes yielding an expected $\langle C_t \rangle = 20$ drops. The population value falls at 2.5 on the abscissa; arrow indicates median sample value.

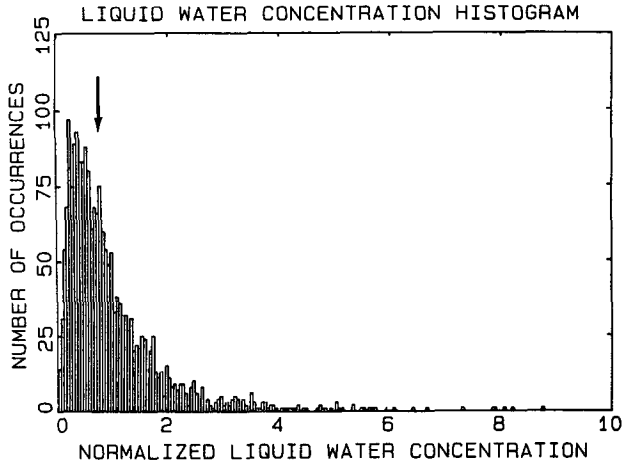

FIG. 6. Histogram of sample values of the (normalized) liquid water concentration, for sample volumes yielding an expected $\langle C_t \rangle = 20$ drops. The population value falls at 1.0 on the abscissa; arrow indicates median sample value.

value, and is actually smaller than the population mass-weighted mean diameter. Here, too, the usual observational result will be a drastic underestimate of the true maximum particle size. Even when $\langle C_t \rangle$ increases to 500 drops, the sample median $D_{\max}$ is still less than $1.7 D_m$ (Fig. 5). This suggests that it is futile to try to infer an upper truncation point for the size distribution on the basis of the maximum observed particle size in a small or even moderate-sized sample.

b. Liquid water concentration

Figure 6 displays the sampling distribution of liquid water concentration for $\langle C_t \rangle = 20$. The skewness in this distribution is even more marked. The median is only 0.78 times the population value and one observation exceeds ten times the population value. This indicates a general tendency of the samples to underestimate the LWC, with an occasional large overestimate occurring. This propensity to underestimate the population value shows up in all variables simulated, and is more pronounced for the higher-order moments of the samples. The median sample estimate only reaches 90% of the population value with (expected) samples larger than 100 drops (Fig. 7).

Direct experimental evidence of this behavior can be found in the output of airborne evaporator-type sensors for total water concentration, that have small entrance apertures. For example, Fig. 4 in Pinty and

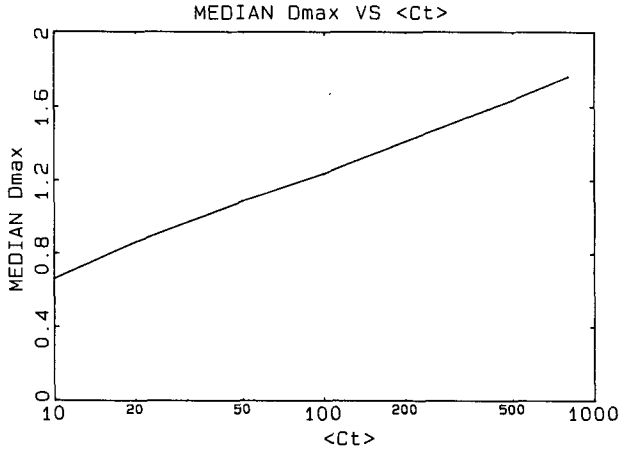

FIG. 5. Variation of median value of maximum particle diameter in samples with expected sample size $\langle C_t \rangle$.

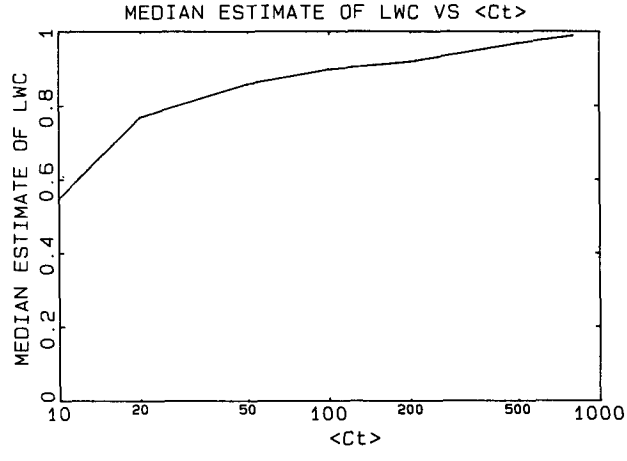

FIG. 7. Variation of median sample estimate of liquid water concentration with expected sample size $\langle C_t \rangle$.

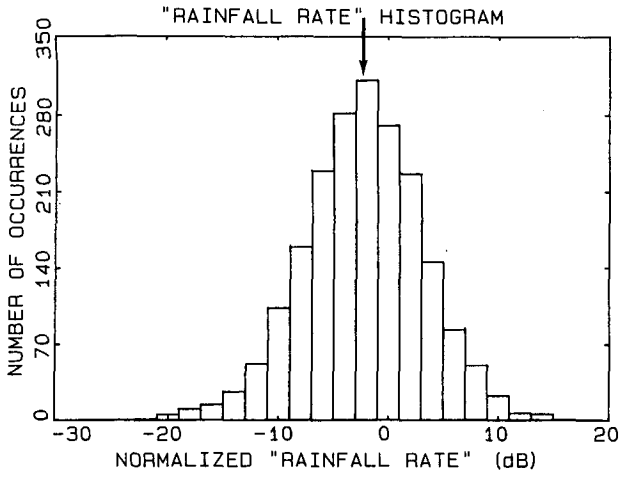

FIG. 8. Histogram of sample values of the "rainfall rate" R^*, for sample volumes yielding an expected $\langle C_t \rangle = 20$ drops. The population value R_p^* falls at 0 dB on the abscissa; arrow indicates median sample value.

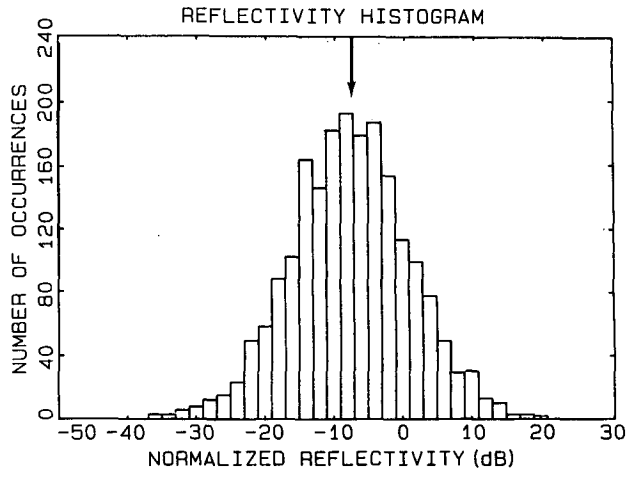

FIG. 10. Histogram of sample values of the (normalized) radar reflectivity factor, for sample volumes yielding an expected $\langle C_t \rangle = 20$ drops. The population value falls at 0 dB on the abscissa; arrow indicates median sample value.

Rosset (1983) displays evidence of individual large particles entering such an instrument (as suggested in Smith 1976). Averages over short time intervals (thus involving small sampling volumes) would have the kind of skewed distribution indicated by this simulation.

c. *Rainfall rate R**

We chose to use R^*, the fourth moment of the drop size distribution as introduced by Joss and Gori (1978), as opposed to R itself so that we could employ normalized variables; this greatly simplified the simulation procedure. The behavior of R values, and of the Z–R relationships discussed in section 4, should be bracketed by the results involving W and R^*. Figure 8 gives the sampling distribution for the values of R^*, for the case $\langle C_t \rangle = 20$; it uses a logarithmic scale for R^* because of the wide range of sample values involved. The distribution, like the others, is markedly skewed toward low values, but on this logarithmic scale it looks roughly normal. (In fact, a chi-square test showed no significant deviation from a lognormal distribution.) Here about 75% of the samples fall below the population value, which is 0 dB on the logarithmic scale. The median is 2.6 dB below the population value; in other words, half the samples underestimate R^* by a factor of 1.8 or more. The maximum exceeds the population value by a factor of 20 (13 dB), while the minimum is 0.005 (or −23 dB) times the population value.

Figure 9 displays the variation of the median sample estimate of R^* with (expected) sample size $\langle C_t \rangle$. It shows again that the tendency to underestimate the population value diminishes as the sample size increases. However, to obtain a median estimate that is within 1 dB of the population value of R^* would require a sample of more than 100 drops.

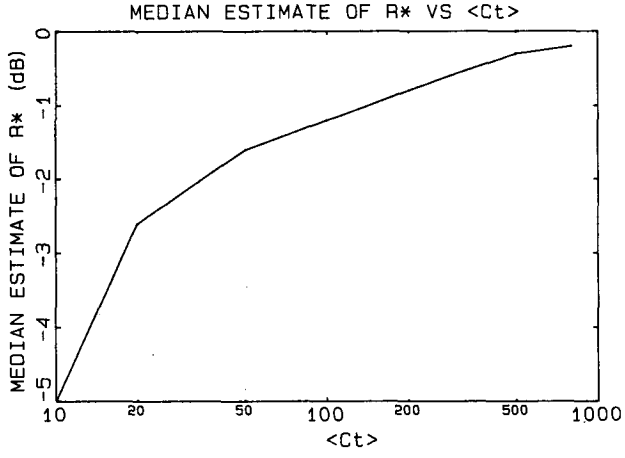

FIG. 9. Variation of median sample estimate of R^* with expected sample size $\langle C_t \rangle$.

d. *Radar reflectivity factor*

Figure 10 illustrates the radar reflectivity factor sampling distribution for $\langle C_t \rangle = 20$. This distribution also appears to be lognormal, and here, too, a chi-square test showed no significant deviation from a lognormal form.[1] Once again the usual observational re-

[1] A reviewer suggested that it might be possible to show analytically that these sampling distributions should be lognormal, but we have not succeeded in doing so.

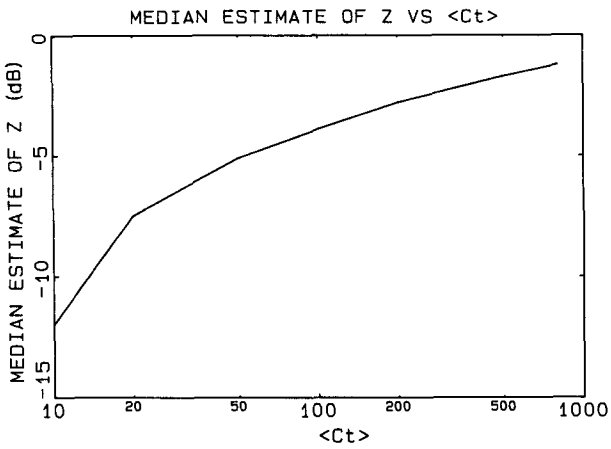

FIG. 11. Variation of median sample estimate of radar reflectivity factor with expected sample size $\langle C_t \rangle$.

sult is a drastic underestimate; almost 85% of the values fall below the population value. The sample median is 7 dB lower than the population value. It increases with $\langle C_t \rangle$ (Fig. 11), but it takes an expected sample of at least 100 drops to achieve a median within 3 dB of the population value.

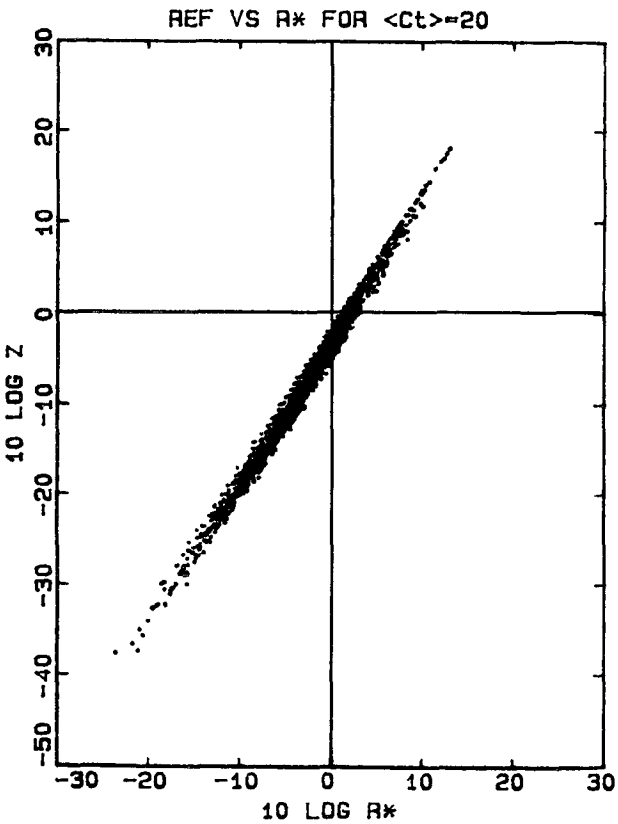

FIG. 12. Scatterplot of sample values of normalized radar reflectivity factor versus normalized rainfall rate, for $\langle C_t \rangle$ = 20; run size = 2000.

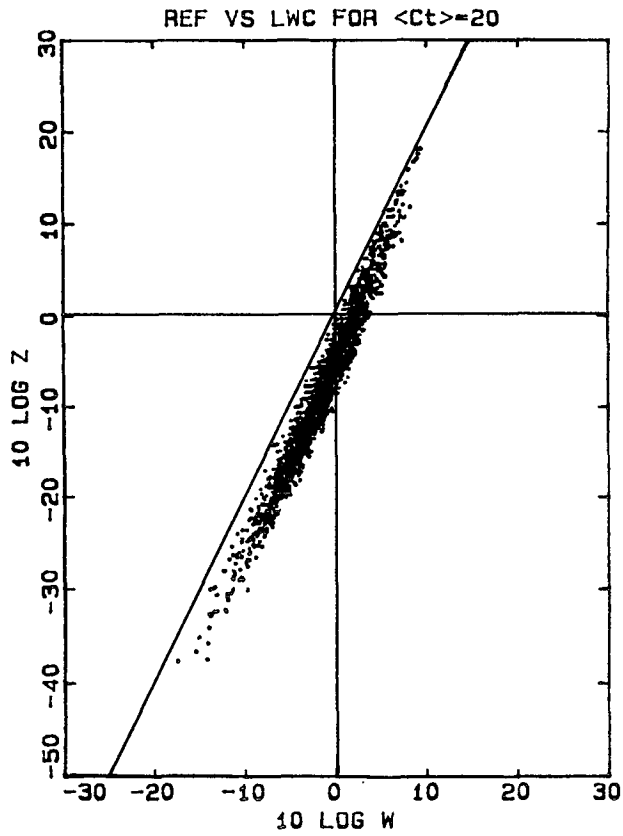

FIG. 13. Scatterplot of sample values of Z versus liquid water concentration, for $\langle C_t \rangle$ = 20; run size is 2000. Diagonal line represents upper bound established in section 5a.

4. Apparent Z–R^* and Z–W relationships

The most interesting results of these simulations concern the apparent Z–R^* and Z–W relationships that result from the sampling variability. The existence of these relationships becomes evident when the individual sample values are plotted on the usual type of log–log scatterplots. Figure 12 shows an example of a resulting "Z–R^* relationship," while Fig. 13 shows a "Z–W relationship." Note that in each case, the apparent relationship is an artifact of the correlation inherent in the computation of the different moments of the same raindrop samples. The parent population (and, hence, the actual value of R^*, W, or Z) does not change at all during the repetitive sampling process. Thus the "relationships" arise entirely from the sampling variability—a characteristic that was noted by Chandrasekar and Bringi (1987).

The scatterplots and corresponding regression relationships have characteristics reminiscent of those ob-

TABLE 1. Linear regression parameters for Z_s–W_s and Z_s–R_s^* log–log scatterplots; $\langle C_t \rangle = 20$; run size is 2000.

	Z_s–W_s	Z_s–R_s^*
Correlation coefficient	0.975	0.994
Standard error	0.190	0.094
Slope (exponent)	2.165	1.580
Intercept	−4.912	−3.305

tained from observations of natural rain (see Table 1). The sample reflectivity factors cover a range of about 60 dB, while the ranges of the other variables correspond to a factor of about 10^3. The regression slopes, which would correspond to the exponents in power-law Z–R^* or Z–W relationships, are 1.58 and 2.17, respectively. The correlations are higher and the standard errors of estimate lower than those commonly found in experimental studies of these relationships.

As the sample sizes increase, the sample ranges of Z and the other variables narrow. Also, the chance that a single drop will dominate a sample (thus leading to a trivial correlation among the quantities) decreases. Consequently, the correlations on the scatterplots decrease (Table 2). The regression slopes tend to increase in the process, and the values become implausible (or at least inconsistent with what is known about Z–R and Z–W relationships) for samples larger than about 100 drops.

It is recognized that no empirical Z–R relationship has been obtained by repeated sampling from a population of drops known to be fixed over the sampling interval. Thus the reported observations always contain some mixture of the sampling variability examined here and actual inhomogeneities in the population drop size distribution. The present results suggest, however, that the sampling-variability component of the reported relationships may be considerably greater than has usually been supposed.

5. Displacement of scatterplot from population values

In the simulation leading to Figs. 12 and 13, the (normalized) population value of each variable is 0 dB. However, as can be seen in the figures, the point (0, 0) falls outside the array of points obtained in the sampling simulation. In other words, no point corresponding to the population values Z_p, R_p^*, or Z_p, W_p will be obtained for any sample of this size, no matter how many times the sampling is repeated. This rather surprising suggestion can be confirmed by the following analysis.

a. Analysis

We work first with the liquid water concentration, for convenience. For any given sample value W_s, the greatest possible value of Z_s would occur if all of the liquid in the sampling volume were contained in a single drop (see the Appendix). If the given sample value W_s were due to a single drop, its diameter would be

$$D_s = \left(6\frac{V_s W_s}{\pi\rho}\right)^{1/3}. \tag{16}$$

The corresponding sample reflectivity factor Z_s would be

$$Z_s = \frac{36 V_s W_s^2}{(\pi\rho)^2}. \tag{17}$$

This defines a $\log Z_s$–$\log W_s$ relationship having a slope of 2.0. To find an intercept value, note that when W_s has the population value from (9), or 0 dB on Fig. 13, Z_s has the value

$$Z_s(W_s = W_p) = \frac{9\langle C_t \rangle N_t D_m^6}{1024}. \tag{18}$$

The ratio of this value to the population value from (14) would be

$$\frac{Z_s(W_s = W_p)}{Z_p} = \frac{\langle C_t \rangle}{20}. \tag{19}$$

This ratio will be less than unity, unless $\langle C_t \rangle \geq 20$. Thus we arrive at the interesting conclusion that the array of points on the Z–W scatterplot cannot possibly include the population value (W_p, Z_p; or 0, 0 on the log scales of Fig. 13) unless the sampling volume is large enough to give an expected sample size of at least 20 drops.

TABLE 2. Regression parameter variation with (expected) sample size.

$\langle C_t \rangle$	10	20	50	100	200	500	800
Z_s–W_s							
Correlation coefficient	0.988	0.979	0.952	0.924	0.886	0.858	0.822
Slope	2.056	2.176	2.384	2.515	2.694	2.980	3.096
Z_s–R_s^*							
Correlation coefficient	0.997	0.995	0.988	0.981	0.970	0.959	0.944
Slope	1.530	1.581	1.677	1.754	1.855	2.018	2.026

A corresponding generalized analysis for any Z–Q relationship (again with $Q = cD^n$) leads to the limiting single-drop results

$$Z_s = c^{-6/n} V_s^{(6-n)/n} Q_s^{6/n} \tag{20}$$

and

$$\frac{Z_s(Q_s = Q_p)}{Z_p} = \frac{(n!)^{6/n} \langle C_t \rangle^{(6-n)/n}}{6!}. \tag{21}$$

This last quantity must be less than unity unless

$$\langle C_t \rangle^{6-n} \geqslant \frac{(6!)^n}{(n!)^6}. \tag{22}$$

Consequently, for the case of R^* ($n = 4$), the slope of the limiting Z–R^* line is 1.5 and the array of points cannot include the point corresponding to the population value, unless $\langle C_t \rangle \geqslant 37.5$ drops.

The line plotted in Fig. 13 is that determined by the point indicated by (19) and the slope indicated by (17). This line clearly represents an upper bound on the points in the scatterplot, confirming that the simulation results are consistent with this finding.

b. *Implications*

What does this imply with respect to empirical determinations of Z–R or Z–W relationships? We customarily proceed by plotting experimental sample values of the two variables on a log–log scatterplot, sometimes (but not often enough) with error bars to represent the associated uncertainties. Taking into account the correlation inherent in the sample values, it would be more nearly correct to think in terms of error ellipses, which would lie more or less parallel to the overall regression line. One instinctively wants to center each ellipse on the point corresponding to the observed sample values of the variables. However, the foregoing analysis shows that for small (expected) sample sizes, the ellipse should actually be placed so that the observed point lies *outside* it. In other words, the expected (two-dimensional) distribution of the point corresponding to the population values does not include the sample point itself.

The array of points in Figs. 12 and 13 is displaced from the point (0, 0) corresponding to the population values of the simulated variables for the reasons discussed in section 5a and the Appendix. We would therefore expect the error ellipse to be displaced in the opposite sense from the observed point on the experimental scatterplot. This displacement would decrease as the sample sizes (number of drops) increase, and so would tend to be smaller as values of the variables (W, R, and Z) increase. Thus the greatest displacements would be expected to occur toward the lower end of the scatterplots.

The net effect of ignoring this displacement when analyzing raindrop observations should be to shift the resulting experimental regression line toward the vertical, leading to a higher exponent in the corresponding power-law relationship. This tendency should diminish as the sampling volume increases, suggesting that the exponent might be observed to decrease as the sampling volume increases. We have reviewed examples of raindrop disdrometer data (all with sample volumes of at least 0.5 m^3) for evidence of this tendency, and find only slight indication of it. However, these examples generally do not involve sample sizes as small as those indicated to be of concern here (no data for rainfall rates lower than about 0.1 mm h^{-1} were included), and inhomogeneities in the underlying drop population of the natural rain would tend to mask this effect.

6. Discussion

The results of this simulation of the raindrop sampling process show the strong skewness in the sampling distributions for the maximum and mass-weighted mean particle sizes as well as for the liquid water concentration, rainfall rate, and radar reflectivity factor. This skewness produces a tendency to underestimate these quantities, the tendency being more pronounced for higher-order moments of the size distribution (e.g., the reflectivity factor). It decreases as the (expected) sample size increases.

For most of the quantities, the skewness becomes small enough that the median sample values approach reasonably close to the population values for sample volumes yielding (expected) sample sizes greater than 100 or so drops. This would occur, for instance, with a mean drop concentration of 500 m^{-3} when the sampling volume exceeds 0.2 m^3. However, the sampling standard deviations in the estimates of the population quantities [as given, say, by Gertzman and Atlas (1977)] may still be so large that the uncertainties are unacceptable for many purposes. In the notation used here, the fractional standard deviation (FSD, or coefficient of variation) for sample values of a quantity Q from this simulation, as given by Gertzman and Atlas, becomes

$$\frac{\sigma_Q}{Q} = \frac{1}{\langle C_t \rangle^{1/2}} \frac{[\Gamma(2n + 1, 10)]^{1/2}}{\Gamma(n + 1, 10)}, \tag{23}$$

where $\Gamma(a, b)$ represents the incomplete gamma function. For each quantity of interest, Table 3 shows the

TABLE 3. Sampling fractional standard deviations (FSD) for bulk quantities.

Quantity	n	FSD ($\times 1/\langle C_t \rangle^{1/2}$)	$\langle C_t \rangle$ for FSD = 1	FSD for $\langle C_t \rangle$ = 100
W	3	3.93	15.4	0.39
R^*	4	5.75	33	0.57
Z	6	26.3	692	2.63

numerical value for the ratio of the gamma functions in (23), the value of $\langle C_t \rangle$ for which the FSD is unity, and the FSD for $\langle C_t \rangle = 100$ drops. The calculated sampling FSDs are significantly reduced by the truncation of the drop size distribution used in the simulation at $D/D_m = 2.5$, even though only about one drop in 20 000 would be larger than this in a full exponential distribution.

For the maximum particle size, even an expected sample of 500 drops yields a median estimate that is still far below the maximum size in the population being sampled. This suggests that determination of maximum sizes, or truncation values for the particle size distributions, may not be possible with practical sampling volumes.

Some instruments do not respond well (or even at all) to small raindrops, say, ones less than 0.3–0.5 mm in diameter. Examples include the Joss–Waldvogel raindrop disdrometer and the Illinois State Water Survey (ISWS) raindrop camera. A part of the exponential drop size distribution important in terms of the numbers of drops involved is therefore not included in observations made with such instruments. A first approximation to the effects of this omission on the results presented here can be obtained by noting that 37% of the drop concentration N_t involves drops larger than $0.25D_m$; and 14%, drops larger than $0.5D_m$. Multiplying the $\langle C_t \rangle$ scales in the various figures by the corresponding factor (0.37 or 0.14, respectively, in these examples) will give approximate results applicable for such cases.

We have also examined the correlations inherent among values of R^*, W, and Z determined from observation of a fixed raindrop size distribution by repetitive sampling. The results show that these correlations are higher than might be intuitively expected, which can make it difficult to separate sampling variations from inhomogeneities in the precipitation in the determination of empirical Z–R (or Z–W) relationships. They also show that the sampling variations can introduce a kind of bias in the scatterplots, which is greater for smaller sample sizes. Thus any assessment of the "accuracy" of a Z–R relationship based on drop size data should include some consideration of the numbers of drops involved in the samples making up the scatterplot.

It is difficult to assess the practical importance of these results. Pronounced effects would occur mainly with very small sample sizes and therefore would be manifest primarily in observations involving low liquid water concentrations or rainfall rates. In a "Marshall–Palmer" rainfall with a rate of 0.1 mm h^{-1}, the mass-weighted mean drop size D_m would be 0.6 mm and the total drop concentration N_t would be 1200 m^{-3}. A 1-m^3 sampling volume would therefore involve $\langle C_t \rangle = 1200$, which is already beyond the range of values considered in these simulations. It is not difficult to restrict attention to samples of sufficient size to render the effects discussed here negligible for most purposes. However, this can be done more easily with ground-based measurements than with airborne instruments operating in highly variable environments with generally quite restricted sampling apertures.

On the other hand, raindrop size data come from many sources and involve a wide range of both drop populations and sampling volumes. The following things suggest that these sampling-variability effects may be of concern in at least some instances:

1) Typical 1-s samples from airborne PMS probes involve around 50–100 raindrops.

2) One threshold used to accept samples obtained with the ISWS raindrop camera was eight drops in a 1 min (or 1 m^3) sample (Jones 1992).

3) Some Z–R scatterplots have included points for estimated rainfall rates well below 0.01 mm h^{-1}, and even some points based on a single raindrop in a sample [an example appears in Smith et al. (1974)].

Thus the sampling variability probably plays some role in the great variations among published Z–R relationships.

These results therefore shed new perspective on the reasons for some of the variability appearing in the wide range of Z–R relationships reported in the literature. The best way to deal with this basic problem is to make sure that only samples of substantial size are used in establishing the Z–R relationships. In general, this will require that the fractional standard deviations of the quantities of interest, as given by Gertzman and Atlas (1977), be of the same order as (or smaller than) the natural variations to be investigated.

Acknowledgments. This research was supported by the U.S. National Science Foundation under Grant ATM-8720252.

APPENDIX

Upper Limit on Z_s

To demonstrate that for any given sample value of the liquid water concentration (LWC), the greatest possible value of Z_s would occur if all of the liquid were contained in a single drop, consider any two drops in the sample. Let D be the diameter of the larger one and $D(1 - \alpha)$ that of the smaller. The total volume of the two drops is $(\pi D^3/6)[1 + (1 - \alpha)^3]$ and they contribute $D^6[1 + (1 - \alpha)^6]$ to Z_s. Now let the two be combined into a single drop, whose diameter would be $D[1 + (1 - \alpha)^3]^{1/3}$. Its contribution to the reflectivity factor would be

$$D^6[1 + (1 - \alpha)^3]^2 = D^6[1 + 2(1 - \alpha)^3 + (1 - \alpha)^6], \quad \text{(A1)}$$

which clearly exceeds that of the original pair of drops.

Repeating this "coalescence" process over the whole

drop sample, we arrive at the conclusion that for a given sample LWC, the greatest possible value of Z_s would occur if it were due to a single drop in the sampling volume.

REFERENCES

Chandrasekar, V., and V. N. Bringi, 1987: Simulation of radar reflectivity and surface measurements of rainfall. *J. Atmos. Oceanic Technol.*, **4,** 464–478.

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.

Jones, D. M. A., 1992: Raindrop spectra at the ground. *J. Appl. Meteor.*, **31,** 1219–1225.

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

——, and A. Waldvogel, 1969: Raindrop size distribution and sampling size errors. *J. Atmos. Sci.*, **26,** 566–569.

Lovejoy, S., and D. Schertzer, 1990: Multifractals, universality classes, and satellite and radar measurements of cloud and rain fields. *J. Geophys. Res.*, **95,** 2021–2034.

Marshall, J. S., and W. McK. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5,** 165–166.

Mueller, E. A., and A. L. Sims, 1966: The influence of sampling volume on raindrop size spectra. *Proc. 12th Conf. Radar Meteor.*, Norman, Amer. Meteor. Soc., 135–141.

Pinty, J. P., and R. Rosset, 1983: L'appareil de Ruskin pour la mesure de la teneur en eau totale nuageuse: Avantages et limitations. *J. Rech. Atmos.*, **17,** 67–76.

Richards, W. G., and C. L. Crozier, 1983: Precipitation measurement with a C-band weather radar in southern Ontario. *Atmos.–Ocean,* **21,** 125–137.

Sasyo, Y., 1965: On the probabilistic analysis of precipitation particles. *Proc. Int. Conf. Cloud Physics,* Sapporo, Japan, 254–259.

Smith, P. L., Jr., 1976: Comments on "The measurement of water content by an evaporator." *J. Appl. Meteor.*, **15,** 189–191.

——, K. R. Hardy, and K. M. Glover, 1974: Applications of radar to meteorological operations and research. *Proc. IEEE,* **62,** 724–745.