10-1-1996

Non-Rayleigh signal statistics caused by relative motion during measurements

A. R. Jameson
RJH Scientific, Inc.

Alexander Kostinski
Michigan Technological University

Follow this and additional works at: https://digitalcommons.mtu.edu/physics-fp

Part of the Physics Commons

Recommended Citation

Follow this and additional works at: https://digitalcommons.mtu.edu/physics-fp
Part of the Physics Commons
Non-Rayleigh Signal Statistics Caused by Relative Motion during Measurements

A. R. JAMESON
RIJ Scientific, Inc., Alexandria, Virginia

A. B. KOSTINSKI
Michigan Technological University, Houghton, Michigan
(Manuscript received 5 July 1995, in final form 20 February 1996)

ABSTRACT

In order to reduce fluctuations, remote sensing devices such as radars and radiometers typically sample many times before forming an estimate. When mean values are stationary during this sampling period, the fluctuations in the amplitudes and intensities obey the same probability density functions (pdf) as those for each sample contributing to the estimate. However, it is shown in this work that when mean values change from sample to sample (i.e., pulse to pulse for most radars), the pdf’s of the amplitudes and intensities differ from those corresponding to the samples. Such changes can be inherent to the scatterers as, for example, the scatter of microwaves from an ocean surface, or they can be induced by factors such as antenna motion across gradients.

With respect to meteorological radars, it is routinely argued that the central limit theorem leads inexorably to zero-mean Gaussian distributions of the two components of the electric field phasor backscattered from precipitation because of the large number of independent scatterers in the sampling volume. Consequently, the net amplitudes and intensities obey Rayleigh and exponential probability density distributions, respectively. While apparently true for each pulse (sample) even when the reflectivity across the beam is not uniform, the authors show that, in general, the underlying statistics of the amplitudes and intensities are no longer Rayleigh nor exponential. This occurs because the number of scatterers and intensities change from sample to sample as, for example, when a radar beam moves while the mean intensity is changing. Consequently, non-Rayleigh statistics and deviations from Gaussian distributions are probably much more common than previously appreciated.

A statistical model is developed and confirmed from detailed Monte Carlo drop simulations of a radar sampling as the beam moves through a cloud. Theory and these model simulations show that the resultant pdf’s of the amplitude and intensity are mixtures of the pdf’s from each sample contributing to the estimate. This mixture of pdf’s also produces increased variance. Because of the general nature of these findings, it is likely that the effects of sampling through changing conditions (namely, biases and increased variances) probably also apply to many other types of remote sensing instruments including those using square-law detectors.

1. Introduction

The rapidly fluctuating intensities of radar echoes is one of the most noticeable characteristics of distributed meteorological targets such as precipitation. These fluctuations are the expression of constructive and destructive interference of waves scattered by each drop as they reshuffle because of different relative motions. While the basic physical understanding of coherent scatter has been known since the work of Rayleigh (1877) on the theory of sound, the application of this understanding to meteorological radar signal statistics gained general acceptance through the work of Marshall and Hitchens (1953).

Because the behavior of the amplitude of waves at a single frequency can be described in terms of a magnitude and a phase, it is often represented on a plane as a vector (referred to as a phasor) that moves in time. Associated with each position of the phasor there is also a projection of the X and Y components (also called the I and Q quadrature components in the literature) that fluctuate as the phasor moves. It is also well known that when there is a sufficient number of independent approximately equivalent scatterers and after adequate sampling, the central limit theorem leads to a zero-mean, equal variance, and uncorrelated Gaussian distributions of the X and Y components. As Marshall and Hitches (1953) then show in the meteorological context, the Rayleigh probability distribution for the amplitude A and the corresponding resulting probability distribution of the intensity I (\(I = X^2 + Y^2\)) follow and are given by:

\[
P(A) dA = \frac{2A}{(a^2)} \exp\left(-\frac{A^2}{(a^2)}\right) dA
\]

\[
P(I) dI = \frac{1}{(\langle I \rangle)} \exp\left(-\frac{I}{(\langle I \rangle)}\right) dI, \quad (1)
\]

Corresponding author address: Dr. Arthur R. Jameson, RIJ Scientific, Inc., 5904 Richmond Highway, Suite 401, Alexandria, VA 22303.
where the angle brackets denote the average. These concepts and distributions are the foundation of much of the treatment of meteorological radar signal statistics. Specifically, (1) describes the statistical basis for the so-called square-law radar power receiver that produces an output proportional to $I$. It is also used to derive expressions applicable to other power receivers such as the so-called linear receiver and logarithmic receiver that produce output signals proportional to the average $I^{1/2}$ and $\log(I)$, respectively.

Yet Rogers (1971) hypothesized that at times the distribution of intensity may deviate from exponential when a radar beam encompasses a region in which the radar reflectivity factor $Z$ varies. In his formulation he envisioned a sampling volume containing an assortment of different reflectivity factors occurring at a frequency governed by the amount of the cloud filled by any particular value of the reflectivity factor. He then hypothesized that the resulting density distribution of the amplitude measured for the entire sampling volume would be non-Rayleigh because the probability of any particular value would be conditioned on its frequency of occurrence in the cloud. At no point, however, was there a discussion of how this situation leads to a *microphysical* violation of the conditions satisfying the central limit theorem (CLT), a necessity if the amplitude distribution is to be non-Rayleigh (at least when the phase is uniformly distributed).

As we demonstrate here, this previous work is inconsistent with the actual process of the radar measurement. In this paper we present a physical model and mathematical formalism that avoids a fundamental weakness in prior work, namely a prediction of non-Rayleigh statistics for a sampling volume containing variable reflectivity factors even when the beam is stationary.

We claim that a *stationary* radar beam does not "see" all the different reflectivity values in a conditioned manner but only the one mean value corresponding to the entire sampling volume. The reason is that a beam containing widely different reflectivity factors simultaneously does not violate any of the conditions necessary for the CLT. In particular, the sum of independent random variables (in this case the components of the complex scattered amplitudes corresponding to each drop) tends to have a Gaussian density function if their number is large, they are weakly correlated, and none of the variables is dominant. These conditions are still satisfied even when inhomogeneities in the reflectivity are present. Specifically, a radar resolution volume typically contains a vast number (billions) of independent scatterers, no one drop dominates the signal, and the drops scatter independently, so that the amplitudes of the drops remain uncorrelated. Consequently, the conditions of the CLT are satisfied, and the net amplitude must remain Rayleigh distributed. This conclusion is confirmed independently using a detailed microphysical Monte Carlo simulation of radar measurements in clouds.

In addition, while appearing similar to prior formulations, our expressions are essentially different and converge to the proper value for stationary resolution volumes. In previous formulations it is the spatial distribution of the reflectivity factor, independent of the sampling strategy, that conditions the final probability distribution, whereas in our analysis it is the probability distribution of the *measured mean* values [from sample (pulse) to sample] that matters. These values and their distribution depend not only on the variance of the actual reflectivity factor but also on the beam (size and motion) as well.

And yet if prior formulations are incomplete as we claim, why then are there a number of observations (e.g., Schaffner et al. 1980, 1983; more recently by Scarchilli et al. 1986) that indeed show differences between square law and linear receivers in regions of strongly changing radar reflectivity factor? The challenge, then, is to understand how this comes about. In contrast to a priori mathematical conjectures, however, we begin in this paper by first considering the *microphysical* origin of the radar measurements.

In particular, it appears that zero-mean Gaussian distributions of the quadrature components and Rayleigh distributions of the amplitude are no longer valid once the beam moves during sampling because the number (and sizes) of scatterers may change substantially from one pulse to the next. Consequently, non-Rayleigh statistics are probably much more common than previously appreciated. We will show that whenever the radar beam is moving while the sample to sample *true* mean intensity is changing, the statistics of the amplitudes are most generally non-Rayleigh. Moreover, if the changes are occurring rapidly, there will be a significant increase in the variance and inaccuracies of the estimates.

Here it is important to review a few points. Deviations from the Rayleigh statistics have been observed in the propagation of laser light in the atmosphere (see Jakeman and Tough 1988) and in radar signals such as those scattered off the ocean surface (Goldstein 1951; Trunk 1972; Ward 1981). These appear to be explained in a direct, physical manner in terms of a rapidly varying number of specular reflections that lead to deviations from zero-mean Gaussian components of the phasor. Based upon a random walk having increments of varying length, Jakeman and Pusey (1976), in fact, derive a non-Rayleigh distribution (the $K$ distribution) that appears consistent with the ocean observations and approaches the Rayleigh distributions when the number of specular reflecting elements becomes infinite. Since in clouds radar volumes often include billions of drops, the results for the ocean would seem to be irrelevant for the atmosphere where the CLT should reign supreme. However, since precipitation is rarely uniform, it is reasonable to consider the possi-
bility that these nonuniformities might lead to changes in the number of scatterers during sampling and, hence, violate the conditions required to satisfy the CLT, generally violated when one deals with random sums, that is, sums of random variables with a random number of terms (e.g., Feller 1966, 156–157). Moreover, it is at least conceivable and worth consideration that such unbalanced changes in the number of drops during a sampling period may introduce deviations from Rayleigh statistics.

The objective of this paper, then, is to explore the potential for deviations of intensity measurements from those derived using the Rayleigh distribution of amplitudes, to better understand the physical origins of any deviations, to quantify the magnitude of any deviations, and to consider possible implications to the measurements of intensities and other parameters in precipitation by radars and other remote sensing instruments. This is done following the actual order of our research beginning with a sequence of numerical experiments to help us identify what is important (section 2). While these experiments provide us with the understanding necessary to formulate a theoretical framework, the resulting formal mathematical structure itself is independent of the any specific numerical simulations. This framework is discussed in section 3.

2. Numerical experiments

a. The approach

While differences between two distributions of intensities may sometimes be striking, often the dissimilarities are more subtle so that visual inspection will not always be a sufficient discriminator. It is, therefore, useful to follow the approach in Jakeman and Pusey (1978) and Jakeman and Tough (1988) of using normalized moments of intensity \( I \) given by \( \langle I^n \rangle / \langle I \rangle^n \) for \( n = 2, 3, 4 \cdots \), where angle brackets denote an average. Using the moment generating function, it is then easy to show that for the Rayleigh distribution

\[
\frac{\langle I^n \rangle}{\langle I \rangle^n} = n!,
\]

so that each normalized moment corresponds to a single number.

In order to gain insight into how non-Rayleigh statistics may come about, we use detailed microphysical Monte Carlo simulations of radar measurements in rain. Because such calculations are computer intensive, the approach is simplified somewhat but is more than sufficient to capture the essential elements of radar measurements. While the meteorological details of the various simulations are not essential to the conclusions of this study, an attempt is made to make them at least somewhat realistic albeit still simplified.

In the most general case we envision a 1-km-wide radar beam moving across a hypothetical storm and sampling a total volume of \( 10^4 \) m³. During the interval between each radar pulse, the beam moves some amount (initially 10 m in this study). Consequently, we divide the “storm” into 10-m increments so that for any radar sample (pulse) there are 100 of these subelements in the beam. For each 10-m subelement, a reflectivity factor \( Z \) is assigned to a 100-m³ volume (in order to get a large number of drops) containing an exponential size distribution of raindrops given by

\[
N(D)dD = N_0 \exp(-\Lambda D)dD,
\]

where \( N(D)dD \) is the concentration of drops of size \( D \) to \( D + dD \), \( N_0 \) is a constant, and \( \Lambda \) is the slope of the distribution. In this work these latter two parameters are characterized by the Sekhon and Srivastava (1971) relations for thunderstorms. In order to reduce the computation time to reasonable levels, however, the entire distribution of drops is not used. Rather, we use a representative drop size \( D_z \) corresponding to that diameter that contributes the most to the reflectivity factor. Since \( Z(D) \propto D^6 \), it is straightforward to show that for exponential drop size distributions extending from 0 to \( \infty \), \( D_z = 60/\Lambda \), where \( D_z \) is in millimeters and \( \Lambda \) is in inverse centimeters. Furthermore, the number \( N_z \) of drops of size \( D_z \) is then given by \( N_z = Z/D_z^2 \). Profiles of the reflectivity factor used in this study include uniform \( Z \) and increasing (decreasing) linear profiles of the logarithm of \( Z \) with slopes of 5, 10, 20, and 40 dBZ km⁻¹.

For each 10-m movement of the beam between pulses, the subelement farthest from the direction of motion is eliminated, while a new subelement closest to the direction of motion is added. All other 98 subelements are retained until the next pulse. This continues until the radar has “scanned” the entire storm.

After accumulating a certain number of these “hits” or samples (10 in these first simulations), the radar computes various averages of the intensity. [For the 1-km beamwidth used in these initial calculations, the rate of sampling represents a 1% change across the width of the beam per pulse (10 m) and corresponds quite closely to typical sampling rates of most meteorological radars that vary between 0.5% and 2% of the beamwidth per pulse.]

For the simulation, it is assumed that we are using frequency agility (such as that typically used in spaceborne and other airborne radars where rapid sampling is necessary, for example) or that the interpulse period is sufficiently long so that the signal from each successive pulse is independent. For each subvolume corresponding to each pulse, we then use a random number generator with a repeat length exceeding \( 2 \times 10^{18} \) to assign each drop (actually several drops at a time using binomial deviates) a phase position and then compute the net complex amplitude after weighting the contribution of each drop by \( D^2 \) (since the amplitude is proportional to \( D^2 \) and \( D^3 \) is different for each subelement). We then move the “beam” for the next pulse.
and repeat the process of assigning new phases and summing to derive the net complex amplitude over the entire beam volume of $10^4 \text{ m}^3$ for that pulse. For each pulse the appropriate input is added to the square-law, logarithmic, and linear "receivers." This is repeated until we finally average a specific number of samples (i.e., number of pulses) to produce an estimate (the output) from the three different receivers. In addition, corresponding to each radar averaging cycle, we calculate the true $Z$ for later comparisons to the receiver outputs, and we calculate the normalized second through fifth moments of the intensity for characterization of the intensity probability distribution. Finally, this entire process is repeated from 20 to 2000 times to derive measures of quantities such as mean bias and to study any convergence toward a Rayleigh distribution.

b. First simulations—Results for a stationary sampling volume

1) **Uniform Reflectivity Factor**

In order to verify the validity of the randomization procedure, the radar sampling volume is held stationary and is filled with a uniform reflectivity factor. In Fig. 1 the average intensity $\langle I \rangle$ as well as the second normalized moment converge close to the values appropriate for a Rayleigh distribution of amplitude in less than 30 independent samples, while the third moment is close by 50 samples. Naturally it takes more samples (200 samples or more) to converge to the fourth and fifth moments, but convergence is clearly evident. This example is the classic scenario leading to Rayleigh statistics.

2) **Linear Profiles of the Logarithm of Z**

If uniform Z leads to a Rayleigh solution for a stationary sampling volume, the next obvious test is to reconsider one of the Rogers’s (1971) calculations using a linear profile of the logarithm of $Z$ lying within a stationary sampling volume. According to Rogers’s macroscopic formulation, we should find a non-Rayleigh distribution of intensities. However, results from our microscopic simulation of a stationary beam encompassing a 40 dBZ km$^{-1}$ gradient in reflectivity factor across the entire radar beam show that the statistics of the resulting distribution is, in fact, Rayleigh at least up to the fifth normalized moments (Fig. 2). It appears that Rogers’s lack of a physical model to describe the radar measurements introduces a violation of the central limit theorem when, in fact, there is none. But then, what causes the apparently observed deviations from Rayleigh distributions in other scenarios? The prime objective of the remainder of this section is to address this question.

3) **Uniform but Fluctuating Number of Scatterers**

Next, suppose the number of drops all of the same size is allowed to fluctuate on a pulse to pulse basis an average $\langle \Delta N \rangle/\langle N \rangle$ where $\langle N \rangle$ is the average number of drops in a stationary sampling volume. An increase in $\langle \Delta N \rangle$ is, therefore, equivalent to an increase in the variance of $N$ and, hence, in the variance of the intensity. While perhaps of uncertain meteorological relevance for a stationary sampling volume, significant changes in $\langle N \rangle$ can be viewed as a simplification of

---

**Fig. 1.** Plot showing the convergence of the mean intensity and the normalized moments of the intensity toward their values corresponding to a Rayleigh distribution of amplitudes as the number of independent samples increases.

**Fig. 2.** Results of a simulation of radar observations in a sampling volume encompassing a linear profile of the logarithm of the radar reflectivity factor. While this replicates a scenario described by Rogers (1971), in contrast to those findings, these results correspond to a Rayleigh distribution.
what happens when a radar beam moves into or out of a cloud, for example. Before performing the latter full-fledged simulation, however, we first wish to explore just what magnitude of fluctuations are required to produce obvious deviations from the Rayleigh distribution. Therefore, while \( \langle N \rangle \) is held constant, we want to see what happens when \( \Delta N \) is allowed to increase as a percentage of \( \langle N \rangle \). This is illustrated in Fig. 3 where the normalized moments relative to those corresponding to a Rayleigh amplitude distribution are plotted using 5000 Monte Carlo simulations and a mean number of 1000 drops.

It appears that while the moments are indeed Rayleigh when \( \Delta N / \langle N \rangle \leq 15\%\), significant departures first appear once the average number fluctuation exceeds 15\% of the average number of scatterers. While it is unlikely that the number of drops will change that much during a dwell time in a stationary sampling volume, such changes should be fairly common when the radar sampling volume moves into different regions of a cloud. It is, therefore, probably no accident that Scarcelli et al. (1986), for example, found that the effects of changing reflectivity factor seemed to be enhanced once they moved the radar antenna. We, therefore, must investigate why antenna motion appears to be so important by considering a moving sampling volume.

c. Second simulation—A moving sampling volume

1) Step function from clear air into a region of uniform reflectivity factor

In this simulation imagine a beam 1 km wide leaving clear air and entering a cloud of uniform reflectivity factor. As the beam sweeps over this step function between no-cloud and cloud, the number of scatterers is increasing with each pulse as more of the cloud in included in the sampling volume. Furthermore suppose that the radar transmits a pulse corresponding to a sample of the cloud every 10 m and that the radar averages over ten such pulses to produce each estimate of intensity every 100 m (10 m per pulse time 10 pulses). In Fig. 4 the normalized moments from this simulation are plotted. Clearly, even when entering a uniform field of reflectivity from clear air, the distribution of intensities deviates considerably from that expected for the Rayleigh distribution as the number of drops increases from pulse to pulse.

2) The effect of gradients in \( Z \)

In this set of simulations the radar "scans" into and out of regions of gradients in the radar reflectivity factor. Performing the same calculations as in the previous section Fig. 5 shows that once again the statistics are non-Rayleigh whether \( Z \) is changing by 5, 10, 20 or 40 dBZ km\(^{-1}\). Moreover, the similarity between Figs. 4 and 5 suggest a common, but non-Rayleigh shape of the amplitude distribution corresponding to the radar estimate. In short it appears that zero-mean Gaussian distributed components and Rayleigh distributed amplitudes and associated intensities are perfectly valid until the radar antenna moves.\(^1\) Then these distributions will deviate from those expected using classical theory. In the next section we investigate the physical origin behind this effect.

3. The probability mixture model

Since the simulations involving a stationary sampling volume all agree with Rayleigh statistics even if the volume contains variable \( Z \), we have concluded so far that the motion of the radar is important for generating deviations from Rayleigh statistics. How does this happen?

Here we want to divorce our conceptual formulation of the pdf's from the process of actually estimating quantities, something we will then consider briefly later. Nevertheless, it is important to recall that radars combine a number of samples. Because the radar is essentially stationary at the instant of each pulse, the CLT implies that the amplitude corresponding to each pulse (sample) is drawn from a Rayleigh distribution

---

\(^1\) Alternatively, the same effect, in principle, could be produced if the precipitation moved through the beam, but normally that would require unrealistically rapid motion of those scatterers located several kilometers from the radar. For example, at 60-km range, a shift corresponding to 1° rotation of the antenna in 0.128 s, fairly typical for many radars, would require an advection velocity of the precipitation of about 490 km h\(^{-1}\) to produce the same effect for a stationary antenna.
Fig. 4. A scatterplot of the normalized moments of intensity as a function of the normalized second moment of the intensity as a 1-km-wide radar beam enters and exits a region of constant radar reflectivity factor. Even though the reflectivity is uniform, the statistics deviate significantly from the point values expected (double circles) if the amplitude distribution were Rayleigh.

having a mean consistent with the $Z$ (number and sizes of drops at the microscopic level) in that particular sampling volume; that is, for each pulse the amplitude is a random variable drawn from a Rayleigh distribution conditioned by the mean value. These samples are not identical, though, because the radar is moving and usually views different parts of the cloud from sample to sample. Thus, the mean value of each pdf (number and sizes of the drops at the microscopic level) changes from pulse to pulse. Consequently, the underlying distribution of the amplitudes must be non-Rayleigh, in general. The question is then "what is the underlying probability density function (stochastic model) of the amplitude (as well as intensities, quadrature components) in this nonstationary case?" While for some purposes one might be satisfied knowing only the mean and the variance, which can be calculated using a conditional probability approach, the entire pdf is much more satisfying since it contains all available statistical information about the random variable.

As the above discussion suggests, we can determine this pdf using a conditional probability approach. That is, if we let $X$ and $\mu_i$ be random variables, then for a mixture having $M$ components ($M$ resolution cells) the pdf is given (e.g., Feller 1966, specifically vol. 2, chapter 2, section 5; chapter 5, section 9; and p. 164) by

$$P(X)dX = \sum_{i=1}^{M} p_i P_i(X | \mu_i) dX,$$

where the $p_i$ and $\mu_i$ are random variables (representing, say, the measured intensities and mean intensities of the $i$th resolution cell) corresponding to the relative fraction that $P_i(X | \mu_i)$ contributes to the mixture, $\sum p_i = 1$, the $\mu_i$ are all different (mutually exclusive conditions), and the vertical bar denotes conditioning. Furthermore, for conditional random variables (e.g., Ochi 1990)

$$E_\mu[E(X | \mu)] = E(X)$$

and

$$\sigma_X^2 = E_\mu[\sigma^2(X | \mu)] + \sigma^2_\mu[E(X | \mu)].$$

where $E$ denotes the expected value while $E_\mu$ denotes the expected value with respect to random variable $\mu$ (i.e., the averaging takes place over all $\mu$). Hence, the mean of $X$ is simply the average of all mean $X$ corresponding to all the $\mu$. On the other hand, the variance of $X$ is the sum of two components, one being the mean variance of $X$ averaged over all the $\mu$ and the other being the variance of all the $\mu$.

If we let $X$ be the intensity measured by a radar and if $\mu$ is the mean intensity corresponding to a sample cell, then (5) says that the average intensity will be the average of all the $\mu$'s. In particular, if the beam is stationary, there is only one $\mu$ so that the average intensity is $\mu$, while the first term of the variance on the right-hand side of (5) is simply the variance of the exponential distribution corresponding to $\mu$ and the second-term vanishes. Thus, we end up with a mean and variance corresponding to the exponential distribution, in the case of intensity, or the Rayleigh distribution, in the case of amplitudes.

With regard to the entire pdf, if we let $X$ be the amplitude, then (4) becomes a mixture of $M$ Rayleigh distributions; that is,

$$P(A)dA = \sum_{i=1}^{M} p_i P_i(A | \mu_i) dA,$$

where

$$P_i(A | \mu_i) dA = \frac{2A}{\mu_i} \exp\left(-\frac{A^2}{\mu_i}\right) dA.$$

Here it is important to point out that $P(A)dA$ is the probability that an observed amplitude from one pulse lies between $A$ and $A + dA$. As we point out below, however, (6) does not represent the pdf of the estimated mean amplitude formed after summing $M$ pulses and dividing by $M$. 

\footnote{Nonstationarity implies the presence of both nonuniformities and significant relative motions.}
This formulation for the distribution of the amplitudes can be confirmed using our detailed microphysical Monte Carlo simulations. This is illustrated in Fig. 6 where results from this probability mixture are compared with those using the Monte Carlo drop simulation model of the radar entering a region containing a 30 dBZ km$^{-1}$ gradient in the reflectivity factor. In this figure the mixture pdf was determined by computing the Rayleigh distribution using the true mean value corresponding to each component.

Consequently, in Fig. 6a when the antenna moves very slowly through the reflectivity gradient (so slowly that there are 10 estimates over the distance of the beamwidth), the amplitude distribution remains nearly Rayleigh because $\mu$ changes very slowly from sample to sample. [This probably explains, for example, why measurements by Lhermitte and Kessler (1966) collected while slowly moving the antenna agreed with the expectations of the Rayleigh distribution.] The Rayleigh mixture model and drop simulation both capture this effect of slow antenna motion as Fig. 6a illustrates. However, when the radar moves more quickly so as to yield one estimate per beamwidth, as is typically done in practice, Fig. 6b shows that the amplitude distribution deviates significantly from that expected for a Rayleigh distribution associated with the same mean inten-
sity. Moreover, the agreement between the drop simulation and the conceptual mixture model is excellent so that we have very good reasons to believe that (6) is the correct mathematical representation of the distribution of sampled amplitudes.

Furthermore, using (5) or computing directly from (6), it follows that

$$
\langle A \rangle = \left( \frac{\pi}{4} \right)^{1/2} \sum_{i=1}^{M} p_i \mu_i^{1/2},
$$

(7)

while

$$
\sigma^2(A) = \sum_{i=1}^{M} p_i \mu_i - \frac{\pi}{4} \left( \sum_{i=1}^{M} p_i \mu_i^{1/2} \right)^2.
$$

(8)

Similarly, with respect to the intensity it follows from (6) that

$$
P(I) \, dI = \sum_{i=1}^{M} \frac{p_i}{\mu_i} \exp \left( -\frac{I}{\mu_i} \right) \, dI.
$$

(9)

Consequently,

$$
\langle I \rangle = \sum_{i=1}^{M} p_i \mu_i,
$$

(10)

while

$$
\sigma^2(I) = 2 \sum_{i=1}^{M} p_i \mu_i^2 - \left( \sum_{i=1}^{M} p_i \mu_i \right)^2.
$$

(11)

Again, when the antenna is stationary, all $\mu_i = \mu$ so that the mean and variances go to their respective values for Rayleigh and exponential distributions.

One consequence of these relations is that the variances of both the amplitudes and the intensities are increased beyond what they would be for Rayleigh and exponential distributions having equivalent mean values. This is seen by rewriting (8) and (11) so that

$$
\sigma^2(A) = \sigma^2_{\text{Ray}}(A) + \frac{\pi}{4} \sigma_{\mu}^2
$$

and

$$
\sigma^2(I) = \sigma^2_{\text{exp}}(I) + \sigma_{\mu}^2,
$$

where

$$
\sigma^2_{\text{Ray}}(A) = \left( 1 - \frac{\pi}{4} \right) \sum_{i=1}^{M} p_i \mu_i
$$

and

$$
\sigma^2_{\text{exp}}(I) = \sum_{i=1}^{M} p_i \mu_i^2,
$$

Hence, the variances of $A$ and $I$ may be interpreted as having a "statistical" component [first terms on the right-hand sides of (12) and (13)] and another "structural" component [second terms on the right-hand sides of (12) and (13)].
natural' component [second terms on the right-hand sides of (12) and (13)] arising from the variability of the reflectivity factor, \( \mu \). This division follows from (5). Thus, if it were possible to measure these pdf's, it would be possible to extract information about the subbeam scale structure. In particular, it turns out that the second terms on the right-hand side of (13) are only accessible by measuring the entire pdf's.

In other words, now that we have completed a formulation of the pdf's, it is reasonable to ask how we would go about measuring them. In (8) and (11), it is important to remember that \( M \) refers to the number of components in the mixture. In order to measure these pdf's, however, one would have to make \( N \), say, rapid scans over these \( M \) subvolumes. When \( N \) is large, the resulting histogram will approach the mixture pdf by the law of large numbers. In fact, that is precisely how the Monte Carlo simulation curves in Fig. 6 were generated.

While it may often be unlikely that such measurements would be feasible using conventional radars, it should certainly be possible if one were to use a radar such as the 'red ball' (Krehbiel and Brook 1979) that rotates at rates of several hundred rpm. Moreover, using such observations, it should be possible to reconstruct the subbeam scale structure from \( N \) sequences (time series) of \( M \) evenly spaced samples while the beam moves uniformly and the radar receives signal from the \( M \) different subvolumes. While of potential experimental interest, such measurements are not readily available so that one must instead resort to more conventional approaches such as looking at the arithmetic mean and variance of a sequence of evenly spaced samples, for example. This can be investigated by reinterpreting (4). In this case we now let the \( p_i \)'s be constants equal to \( 1/M \), and (4) is now a probability mixture but simply the arithmetic mean of a sum of conditional random variables,\(^3\) that is,

\[
P(I|\mu) = \frac{1}{M} \sum_{i=1}^{M} p_i (I|\mu_i),
\]

where \( I \) is the intensity and \( \mu_i \) is the mean value corresponding to the \( i \)th sample (subvolume). One can then define

\[
\langle I \rangle = \frac{1}{M} \sum_{i=1}^{M} P_i (I|\mu_i).
\]

Furthermore, if the \( P_i \)'s are statistically independent, then by the definition of independence, the variance of \( \langle I \rangle \) is simply the sum of the individual variances (Papoulis 1991) so that the variance will be dominated by the largest \( \mu \). It is also straightforward to derive various expressions for biases corresponding to different detectors.

\(^3\) The fact that Eq. (15) is a special case of our approach was clarified during discussions regarding possible bias that one of us (ABK) had with Dr. R. Meneghini (PIERS Conference, July 1995). We also thank Dr. T. Iguchi for a private communication with one of us (AJ) concerning the same point.
However, these variances depend upon several factors. For example, any action that causes \( \langle I \rangle \) to increase on a pulse (sample) to pulse (sample) basis will lead to increased variances (and increased uncertainty of the estimate). Among them are those that produce an increase in the separation distance between successive samples (pulses) in a region of gradient in the reflectivity factor. This is obvious since a bigger change in distance produces a bigger change in \( \langle I \rangle \) from pulse to pulse. For example, this can come about by reducing the pulse repetition frequency (PRF) while maintaining the same antenna scan rate, increasing the scan rate while keeping the PRF constant, or using a larger beamwidth for a constant scan rate and PRF. On the other hand, if all those factors are kept constant, any increase in the variances must arise from an increase in the variance of the intrinsic reflectivity factor as illustrated in Fig. 7 for different gradients of the linear profiles of the logarithm of \( Z \).

In this example the amplitudes and intensities are measured by a radar that has a 0.5-km beamwidth and generates one estimate using 51 pulses (samples) per beamwidth. The average \( \langle I \rangle \) is held constant in these calculations so that the variances remain fixed for Rayleigh and exponential distributions having means corresponding to \( \langle I \rangle \). Figure 7 illustrates that the variances of the mean intensities and mean amplitudes increase with increasing gradients (variance) in \( Z \), reaching values two to three times larger than would be expected for Rayleigh and exponential distributions. This is important because convergence to the proper mean will be slower; that is, it will require more samples to reach a given level of certainty, or, conversely, for a fixed number of pulses, the uncertainty in the estimate increases as the variability of \( Z \) increases.

Moreover, as a consequence of these relations biases can appear. This is illustrated in Fig. 8 for the previous simulations. Here the bias is defined as the logarithmic difference between the expected true mean intensity and the mean values actually measured by the various detectors. Obviously, the square-law detector remains unbiased, while linear and logarithmic detectors do show some bias because of the increased variance of the measured intensities.

However, this bias is not necessarily the actual bias associated with a particular measurement. That is, the measured mean \( \langle I \rangle \) itself is a random variable having a pdf with a certain dispersion (see Fig. 9a). Since in almost all cases an instrument scans a region only once \( (N = 1) \), the measurement of that mean power corresponds only to one realization from this pdf. Consequently, since the bias is defined to be the ratio of this random variable (mean intensity) to a constant (true mean intensity), the bias itself also becomes a random variable. Hence, the sample bias associated with any particular sample of the mean intensity is also only one realization from a distribution of such biases (Fig. 9b). While the sample bias will converge to an average sample bias (such as those illustrated in Fig. 8) after a sufficient number of realizations \( (N \gg 1) \), in general, the sample bias will deviate from the sample mean bias because \( N = 1 \).

There are several noteworthy points. First, with one realization \( (N = 1) \) it is possible even for a square-law detector to produce a biased sample estimate even though the mean sample bias is 0 dB. Second, because the standard deviation of all three distributions is about 0.8 dB, it will not be unusual for some observations (one in five) to have biases of at least \( \pm 1 \) dB in this example. (Actually, because of the processing methods used by most radars, there will usually be considerably fewer independent samples than in this example so that at times the variance of the bias is likely to be considerably larger than Fig. 9b suggests.) Third, even though these distributions appear nearly symmetric, they are not. For the square-law detector in this example, the bias is less than unity 62% of the time but greater than unity only 38% of the time. Moreover, there is certainly no guarantee that positive and negative biases will “average out” since any particular scene is observed only once \( (N = 1) \). That is, it is unlikely that any particular setting will be ever be reproduced exactly somewhere else, so that each new scene is different and, consequently, so are the pdf’s of the biases. This may preclude a definitive prediction of the amount of net residual bias left after combining an entire set of \( N = 1 \) samples into a “global” average. Finally, Fig. 9 suggests that the detection of such non-Rayleigh statistics using different receivers (e.g., Scarchilli et al. 1986) is better enhanced by using a combination of square law and logarithmic receivers rather than using a combination of square-law and linear receivers. This is not only because of the substantial difference in the mean sample biases (the usual argument), but also because that combination minimizes the overlap between the two pdf’s.

Moreover, during observations it would be useful to detect the existence of non-Rayleigh statistics. One method, as just discussed, is to use the difference between two receivers for such tests. However, this requires two methods of detection, one of them being

---

4 The variances of the intrinsic reflectivity factor are those associated with gradients of linear profiles of the logarithm of the radar reflectivity factor ranging from 5 to 40 dBZ km\(^{-1}\). It is important to note, however, that the sampling order of \( \mu \) is not important to the statistics so that these results apply to any reflectivity structure producing the indicated variance and correct mean value.

5 Normally one might call this an “error,” but we use “bias” here to give a more pertinent connotation to residual uncertainty than that implied by “error,” which usually suggests something that can be “averaged” to insignificance with a large number of samples, a condition not satisfied here.
square law processing. Alternatively, it is possible to detect deviations from Rayleigh and exponential distributions using only a square-law detector by considering the statistic \( \langle I^2 \rangle / \langle I \rangle^2 \) as discussed in earlier sections. It follows from the definition of the variance that

\[
\frac{\langle I^2 \rangle}{\langle I \rangle^2} = \frac{\sigma_I^2}{\langle I \rangle^2} + 1.
\]

When the amplitude distribution is Rayleigh, then \( \langle I^2 \rangle / \langle I \rangle^2 = 2 \) because the variance equals \( \langle I \rangle^2 \). Thus, deviations from 2 indicate deviations from a Rayleigh distribution of the amplitudes.

4. Summary and discussion

In this work we have investigated the physics leading to the violation of conditions required for application of the central limit theorem (CLT). In particular we found that when the number of scatterers changes sufficiently during sampling, it is possible to create non-Rayleigh and nonexponential distributions of the amplitude and intensity, respectively. A common reason for such changes is the movement of the beam during sampling across a region containing inhomogeneities in the reflectivity factor. It is shown that in such cases the distribution of the amplitudes and intensities is described by probability mixtures of the pdf’s associated with each of the components.

For some purposes such a detailed understanding and description would seem superfluous. For example, one may only to be concerned with the measurement of some mean value of, say, the intensity. In this case the full pdf is not required since the statistics of the sampling of the mean itself is properly viewed in terms of a linear sum of random variables drawn from different distributions having different mean values as discussed in section 3.

However, the broader perspective in this paper provides a new way of thinking about the measurements. In the past it was always presumed that the \( N \) samples contributing to an estimate all came from the one Rayleigh distribution under the implicit assumption that the mean values and pdf’s remained unchanged during sampling. We now think that this is not always a good assumption. Rogers (1971) was the first to point out that inhomogeneities in the reflectivity factor might produce deviations from the Rayleigh distribution of amplitudes. But while this previous study correctly identified the importance of spatial structure in the reflectivity factor to radar measurement statistics, it did not provide a correct understanding of how structure can affect radar estimation through beam motion. In this work, however, we begin at the microphysical level of drops and show that while the signals are entirely described by Gaussian and Rayleigh statistics when the radar remains stationary regardless of variations of the reflectivity factor within the sampling volume (in accordance with the CLT), deviations from these Gaussian and Rayleigh statistics do arise when the reflectivity factor changes during sampling. Moreover, the pdf’s of the amplitudes and intensities are described in terms of probability mixtures of the component pdf’s corresponding to each resolution volume.
In other words, there is now a new paradigm of sampling, namely that estimates are formed by sampling a sequence of random variables from different pdf's so that the pdf's of the amplitudes and intensities are no longer Rayleigh nor exponential, respectively, where conditions are changing sufficiently rapidly. For example, this new understanding suggests methods for getting at sub-beam-scale structure, something that was completely ignored under the old paradigm. By collecting a time series of rapid scan measurements across a region, for example, the resulting data can be processed in different ways to extract estimates of the variance of the parameters and, perhaps, even more detailed descriptions of the structure itself.

Moreover this new view shows how sampling as conditions change can increase the variance of the estimates and can introduce biases in all detectors. Why? Because we now understand that data contributing to the measurements collected in regions of variable parameters are usually sampled only once so that we are not really measuring an average mean value nor mean sample bias. Rather we are only looking at single realizations from the pdf's of the means and the sample biases. Consequently, for example, when forming an estimate based upon just one scan (N = 1) across a region of variability, as is usually the case, there may be a sample bias even for square-law detectors that are ordinarily considered untainted by bias. What we now know is that this is only true for square-law detectors in a "mean" sense (N ≥ 1); that, in fact, this condition is essentially never satisfied in practice because regions are rarely sampled over and over in rapid succession. Moreover, it is no longer clear how all these different biases add up when attempting to compute a global average from a set of several such "one-pass" measurements. It may well be that there is a much larger residual uncertainty (bias) than expected surrounding global mean values specifically chosen to minimize the importance of "fluctuations."

In addition, even if there were no large mean sample biases, changes in the reflectivity structure may still significantly increase variances and, hence, slow convergence to the proper mean. For example, one of the objectives of the Tropical Rainfall Measuring Mission (TRMM) is to collect rainfall measurements in the Tropics using an array of different spaceborne instruments including radiometers and a precipitation radar. One of the purposes of the precipitation radar is to collect measurements in warm-rain, intense showers. While nearly invisible to the TRMM radiometers, these showers produce much of the rainfall in parts of the tropics. Unfortunately, such showers are associated with strong gradients in Z. Since the TRMM radar will scan laterally to the direction of orbit motion at a rate sufficient to yield an estimate every beamwidth, the results of this study would seem to suggest that the signal statistics are often likely to be non-Gaussian and non-Rayleigh and that significant inaccuracies might be possible. Moreover, the sample and hold estimation technique (or mixture averaging) may cause a substantial, artificial broadening of the precipitation areas as well at times (see the appendix).

Fortunately, this turns out not to be the case for the TRMM radar. The reason is that the radar is steered electronically so that it uses what may be called "block" averaging (appendix) in which the beam is held stationary while each estimate is made. Block averaging also reduces beam filtering effects of a moving sample volume (appendix). On the other hand, the TRMM radiometers are not electronically steered and, therefore, may be subject to the non-Rayleigh effects described in this work.

While schemes such as block averaging may become more common with advanced electronically steered beams, mixture averaging is likely to remain a problem for most radars and many other remote sensing instruments as well. It becomes increasingly important, then, to evaluate under what conditions and when such deviations from Rayleigh statistics become significant and what adjustments to algorithms will be required for optimal estimation. It is clear from this study, however, that with the exception of the most boring circumstances the mean reflectivity of radars is almost always changing during sampling so that it is likely that non-Rayleigh statistics is far more prevalent than previously appreciated. This is not necessarily a "curse," however, because while such inhomogeneities may increase statistical uncertainty, they may, at the same time, also provide useful information about sub-beam-scale structure with proper sampling and processing.

More generally, for any remote sensing instrument that relies on averaging several samples to form an estimate, it is possible that the statistics of the estimate may deviate from those normally assumed for each sample when the measurements are collected in changing conditions. Consequently, these findings apply not only to radars but to other remote sensors such as radiometers and scatterometers as well.

Acknowledgments. Support for Jameson for this research was provided by the National Aeronautics and Space Administration (NASA) under Grant NAS5-32780 and the Jet Propulsion Laboratory (JPL-958437). Support for Kostinski came from the National Science Foundation under Grant ATM-9116075 and NASA under Grant NAS5-32780.

6 While irrelevant to TRMM, an alternative is to use a wide band of frequencies so that an estimate can be made with one pulse as the antenna moves. This, of course, has already been done for the red ball radar (Krehbiel and Brook 1979) but is of limited value since simultaneous Doppler and polarization measurements are nearly impossible.
APPENDIX

Comparison of Block and Mixture Averaging

The effect of finite beamwidth as a smoothing filter is well known. In the case of a uniform illumination function, beam smoothing is often envisioned as a running (uniform) average of the reflectivity factor, for example, over the dimension of the beam. Such a model is appropriate if a radar were capable of producing an estimate after each pulse [using a wide band of frequencies (see discussion at the end of the body of the paper)] or a series of pulses (block averaging) in which the antenna is held stationary (electronically or otherwise). However, such a model is fundamentally incorrect for a moving antenna and leads to an underestimation of the smoothing actually performed by the beam because of the "sample and hold" technique (mixture averaging) of estimation as illustrated in Fig. A1. (This is also what produces the deviations from Rayleigh statistics discussed in the body of this paper.)

At the top of the figure the rows of boxes or bins represent the distance the radar sampling volume moves per pulse. Each bin has an associated mean reflectivity factor. The left side of the figure represents the case if a radar were capable of block averaging, for example, with an "x" indicating when that box has been sampled. The weighting function for each estimate in the case of block averaging is uniform over the reflectivity factors in the sample volume [i.e., $P(x) = 1/B$, where $x$ is the distance moved during sampling over one beamwidth $B$]. In contrast on the right side of Fig. A1, we see that each reflectivity is weighted differently in the estimate because of the motion of the antenna. In the case when one estimate is produced after the sampling volume has moved one beamwidth (as is typical), the weighting function is now triangular and extends over a distance of twice the beamwidth.

From (10) in the text, it is clear that the estimate of the mean by averaging samples is equivalent to applying the weighting function $P(x)$ to the reflectivity factor in each bin. (That is, the estimate of the mean reflectivity factor is the same whether you average the sample means or average over each bin with the weighting function.) It is then simple to develop analytic expressions for the case of these two types of averaging applied to a linear profile of the logarithm profile of the reflectivity factor (constant gradient in $Z$). This is illustrated in Fig. A2 for "normal" mixture averaging (solid line) and for block averaging (dotted line) in the case when the gradient $G$ times the beamwidth equals 30 dBZ. As for any scientific instrument, we consider error simply to be the difference between the measured and actual reflectivity factor at a location.

It appears that block averaging is superior to mixture averaging. To illustrate, suppose $B = 1$ km. Then in
Fig. A2, $G = 30 \text{ dBZ km}^{-1}$ and the distortion of the echo and the errors (which can be quite significant) extend over a distance of 2 km. On the other hand, if $B = 4 \text{ km}$, then $G$ is only $7.5 \text{ dBZ km}^{-1}$ (a very reasonable value). Yet we still have the same echo distortion and the same magnitude of errors except that they now extend over a distance of about 8 km. Moreover, at the edge of the “storm,” the “measured” average reflectivity factor can exceed the “true” $Z$ (i.e., there is an error) by 10–15 dB in this case.

These figures, however, only apply to square-law detection and a uniform illumination function. If a logarithmic detector is used, then the errors in regions of lower reflectivity may be reduced, but then there would be increased error where $Z$ is greatest. Using a more realistic Gaussian illumination function may also tend to reduce the magnitude of the errors somewhat, but then the beam extends over an even wider domain and contains sidelobes so that it is not clear that there will be a significant (if any) reduction in the errors. Regardless, block averaging appears superior to mixture averaging.

REFERENCES


