Spatial correlation of hydrometeor occurrence, reflectivity, and rain rate from CloudSat

Roger Marchand[1]

Received 3 August 2011; revised 20 January 2012; accepted 23 January 2012; published 17 March 2012.

[1] This paper examines the along-track vertical and horizontal structure of hydrometeor occurrence, reflectivity, and column rain rate derived from CloudSat. The analysis assumes hydrometeors statistics in a given region are horizontally invariant, with the probability of hydrometeor co-occurrence obtained simply by determining the relative frequency at which hydrometeors can be found at two points (which may be at different altitudes and offset by a horizontal distance, Δx). A correlation function is introduced (gamma correlation) that normalizes hydrometeor co-occurrence values to the range of 1 to −1, with a value of 0 meaning uncorrelated in the usual sense. This correlation function is a generalization of the alpha overlap parameter that has been used in recent studies to describe the overlap between cloud (or hydrometeor) layers. Examples of joint histograms of reflectivity at two points are also examined. The analysis shows that the traditional linear (or Pearson) correlation coefficient provides a useful one-to-one measure of the strength of the relationship between hydrometeor reflectivity at two points in the horizontal (that is, two points at the same altitude). While also potentially useful in the vertical direction, the relationship between reflectivity values at different altitudes is not as well described by the linear correlation coefficient. The decrease in correlation of hydrometeor occurrence and reflectivity with horizontal distance, as well as precipitation occurrence and column rain rate, can be reasonably well fit with a simple two-parameter exponential model. In this paper, the North Pacific and tropical western Pacific are examined in detail, as is the zonal dependence.

Citation: Marchand, R. (2012), Spatial correlation of hydrometeor occurrence, reflectivity, and rain rate from CloudSat, *J. Geophys. Res.*, *117*, D06202, doi:10.1029/2011JD016678.

1. Introduction

[2] Since their advent, scanning radar systems have been used to study the spatial structure of the atmosphere using statistical techniques. Much of this research has focused on the structure of the observed Doppler velocities, derived wind fields, or "clear-air" reflectivity obtained from UHF or VHF radar systems (that respond to variations in the index of refraction due to variability in temperature and moisture [e.g., *Pollard et al.*, 2000; *Lohou et al.*, 1998]). However, the spatial structure of hydrometeor (cloud and precipitation) reflectivity and Doppler velocities from centimeter and millimeter wavelength radar have also been investigated [e.g., *Smythe and Zrnic*, 1983; *Heymsfield*, 1976; *Lothon et al.*, 2005]. *Fliflet and Manheimer* [2006] used a high-power scanning millimeter-wavelength (94 GHz, W-band) radar to study the spatial correlation of reflectivity from cirrus clouds. They found the reflectivity autocorrelation function decreased with scale as $r^{2/3}$ (equivalent to a wave number spectrum of $k^{-5/3}$) for scale lengths ranging from 30 m to 10 km. Perhaps more prominently, spatial statistics of near-surface reflectivity or rainfall (derived from near-surface reflectivity) have long been a topic of great interest. This includes categorizing and characterizing precipitation structures of various sizes [e.g., *Austin and Houze*, 1972; *Rickenbach and Rutledge*, 1998], calculating reflectivity or rain rate correlation functions or power spectra [*Zawadzki*, 1973; *Crane*, 1990; *Lee et al.*, 2009; *Overeem et al.*, 2009], and examining scaling (multifractal) properties [*Gupta and Waymire*, 1993; *Lovejoy et al.*, 2008; *Mandapaka et al.*, 2009]. In addition to providing insights into atmospheric dynamics, understanding the correlation structure of rainfall has long been of interest to the hydrological community in attempting to estimate rain accumulation (and uncertainty in such) at the basin scale [e.g., *North and Nakamoto*, 1989] and in recent years a number of papers have focused on the correlation of rain rates on small spatial scales [e.g., *Gebremichael and Krajewski*, 2004; *Ciach and Krajewski*, 2006; *Moreau and Bouar*, 2009].

[3] Launched in late April 2006, the NASA CloudSat mission uses a near nadir-pointing millimeter-wavelength (94 GHz) radar to probe the vertical structure of clouds and precipitation from space [*Stephens et al.*, 2008]. A primary objective of the CloudSat mission is to provide global observations of cloud and precipitation structure that can be used in the analysis of climate models. Accordingly,

[1]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington, USA.

0148-0227/12/2011JD016678

CloudSat observations are being used to construct joint histograms of radar reflectivity with height and these data are being compared with model output using radar instrument simulators [*Bodas-Salcedo et al.*, 2008; *Marchand et al.*, 2009a]. A CloudSat radar simulator (whose purpose is to generate reflectivity-height histograms from climate model output) is included in the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) and will be used in the analysis of climate model output as part of the upcoming Intergovernmental Panel on Climate Change Fifth Assessment [*Bodas-Salcedo et al.*, 2011]. While the CloudSat radar does not scan, it does generate a curtain or two-dimensional cross section through the atmosphere as the satellite moves along its orbital trajectory and the CloudSat observations contain much more information on the structure of hydrometeor fields then is captured in the reflectivity-height histograms. In this paper we examine the along-track vertical and horizontal correlation of hydrometeor occurrence, reflectivity, and column rain rate derived from CloudSat observations using simple (two-point) spatial correlation statistics.

[4] The objective of this paper is to provide a first look at hydrometeor and rain spatial structure as revealed by CloudSat and undertake an initial assessment on the utility of correlation statistics to capture regional differences that can provide insights into the nature of cloud and precipitation structure and that can be used in quantitative model evaluations. The paper is organized as follows. In section 2 we describe the retrieval data sets and in section 3 we define the two point statistics used in this study. In section 4 we examine and contrast results for the North Pacific and the tropical western Pacific. This is followed in section 5 with an examination of zonally averaged results with conclusions and additional discussion in section 6.

2. Description of Data and Analysis

[5] The two data sets analyzed in this paper are the CloudSat geometric profile product (2B-GeoProf) and the column precipitation product (2C-Precip), described in detail by *Marchand et al.* [2008] and *Haynes et al.* [2009], respectively. The data analyzed are for CloudSat data release R04, the most current data available at the time of this study. Changes in the next release (R05) of 2B-GeoProf are expected to have little (if any) impact on the results reported here. Likewise the detection of precipitation reported in the 2C-Precip product is not likely to change significantly. However, the derived rain rates are in a preliminary state of development and large changes may well occur in the future. Therefore we caution that the rain rate related results are provisional.

[6] The 2B-GeoProf product contains the calibrated reflectivity and a hydrometeor detection mask that indicates which radar resolution volumes are likely to contain hydrometeors (at several confidence levels). In the CloudSat data processing, radar pulses are averaged about 0.16 s along the nadir track yielding a horizontal sampling distance of about 1.1 km. However, we note that the effective (6 dB) cross-track by along-track footprint of the radar at the surface is about 1.4 km $\times$ 1.7 km [*Tanelli et al.*, 2008], and so our estimate of horizontal correlation may be biased high at the smallest increment (Δx = 1.1 km). While the CloudSat measured return power (reflectivity) is sampled at a rate equivalent to about 240 m in range, the CloudSat radar transmits pulses with an approximate duration of 3.3 μs and so measures hydrometeor reflectivity with a vertical range resolution of approximately 480 m. That is, the reflectivity data are over sampled in the vertical. To avoid introducing vertical correlation in our analysis driven by the oversampling and because our ultimate objective is to compare these statistics with model output (which will likely have a vertical grid spacing larger than the CloudSat data), we reduce the vertical resolution of the CloudSat data to approximately 1 km (960 m) by averaging three consecutive CloudSat samples (range bins), in order to obtain the mean reflectivity for a span of 960 m, skip or ignore the following sample, and then average the next three samples to obtain the mean reflectivity for the next 960 m span. The averaging is accomplished by converting the reflectivity stored in dBZe to Ze, averaging Ze (with a value of 0 when no hydrometeor has been detected), and then converting back to dBZe. An example is given in Figure 1a, which shows a short segment of CloudSat data with full 240 m vertical sampling and below this (Figure 1b) the same segment reduced to 960 m vertical resolution. As one might expect, reducing the vertical resolution blurs or expands the apparent area filled by hydrometeors (because of partially filled volumes) and does have a small effect on the derived statistics. The statistics we examine here are (of course) an unavoidable function of the radar spatial resolution and sensitivity. In this analysis, we include all hydrometeor detections with a confidence level of 20 and apply a minimum reflectivity threshold of -27 dBZe, both of which ensure a low false detection rate. Details on the operational CloudSat hydrometeor detection algorithm, including false detection rates for various confidence levels is given by *Marchand et al.* [2008].

[7] This analysis is focused on the occurrence of hydrometeors and their reflectivity values rather than retrieved liquid or ice water content. This is because the observed radar reflectivity is a strong function of particle size and water content with contributions from both cloud and precipitating sized hydrometeors. On the basis of reflectivity observations alone, it is impossible to unambiguously identify where there are cloud-sized hydrometeors, precipitation-sized hydrometeors or both contributing to the reflectivity. Nor can one determine the water content without making some assumption about the particle size. In short, retrievals of water content based only on radar data are both highly uncertainty and are unlikely to capture the correct variability in water content since the retrievals have no way to separate what portion of the variability in the observed reflectivity is due to variations in water content rather than variations in particle sizes.

[8] Most of the time, the CloudSat radar pulse is able to penetrate clouds and precipitation and reflect off the surface. One can therefore estimate the path integrated attenuation (PIA) between the radar and the surface as the difference between the observed surface return and the expected (precipitation-free) value. In the R04 release of the CloudSat 2C-Precip product, results are limited to ocean areas only with the expected surface return estimated using the surface wind speed (obtained from AMSR-E; passive microwave emissions) and the sea surface temperature (obtained from ECMWF analysis). In future releases the product will be expanded to cover land surfaces. The CloudSat column rain

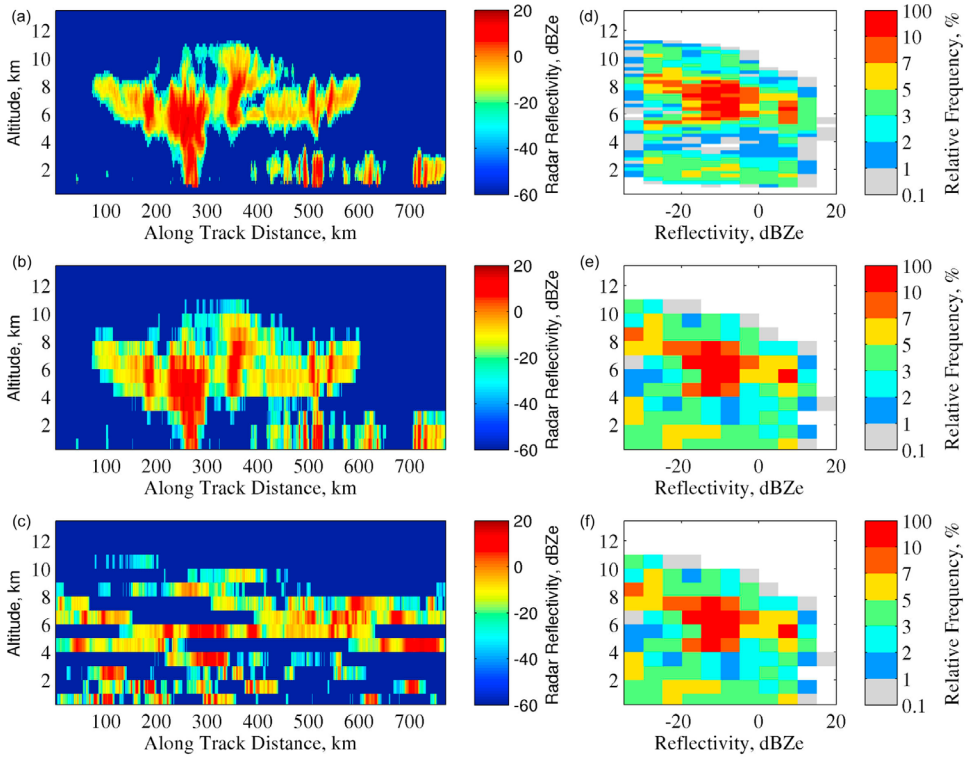

Figure 1. (a–c) Examples of CloudSat observed reflectivity and (d–f) associated reflectivity-height histograms: (top) full vertical sampling, (middle) 1 km vertical resolution, and (bottom) scrambled reflectivity field (see text). That scrambling the reflectivity field has no effect on the associated reflectivity-height histogram is demonstrated.

rate retrieval uses a combination of the observed reflectivity field and the PIA. The PIA is used to estimate the column rain rate while the observed reflectivity is used to ensure the presence of precipitation near the surface. Near surface reflectivity thresholds (after correction for rain attenuation) of −15, −7.5, and 0 dBZe are used to define possible, probable, and certain surface precipitation. While there is a strong relationship between PIA and the rain rate, there is a maximum value of PIA after which there is no surface return. For CloudSat the PIA limit is approximately 40 dB, which is equivalent to a max rain rate of 3.5 mm/h to 25 mm/h depending on details of vertical distribution. Above this threshold we are confident that there is precipitation in the column, and we assume at the surface. The algorithm also makes a variety of assumptions, including assuming the precipitation rate is constant with height. Precipitation is partitioned into snow, rain, cloud water and mixed-phase components (all of which contribute to PIA) in a complex way that depends on ECMWF analysis temperature profiles. A detailed description and sensitivity analysis are given by *Haynes et al.* [2009], with the studies by *Ellis et al.* [2009] and *Berg et al.* [2010] also providing insights into algorithm performance.

3. Two-Point Spatial Statistics

[9] Several recent papers have highlighted the value of using reflectivity-height histograms in climate model evaluation [*Bodas-Salcedo et al.*, 2008; *Marchand et al.*, 2009a]. While observed reflectivity-height histograms are certainly valuable in evaluating models, these histograms are single-point statistics and contain no information on the spatial correlation of the observations. That is, one can arbitrarily reorder the radar profiles in any CloudSat data segment (destroying all along-track or horizontal correlation) or shift the observed reflectivity (at a given height) an arbitrary along-track distance relative to the layer above or below (destroying vertical correlation) and obtain the same reflectivity-height joint histogram. The latter effect is demonstrated in Figures 1c and 1f, which show that in spite of scrambling the reflectivity field we obtain the same reflectivity-height histogram.

[10] In this paper we examine some simple two-point statistics, that is statistics that depend on the simultaneous co-occurrence (yes or no) of hydrometeors at two points in space or some value associated with the hydrometeors (in this case, radar reflectivity or rain rate) at two points. We assume that the statistics are horizontally invariant such that we can define the two-point probability of hydrometeor co-occurrence function, P_{12} ($h1$, $h2$, Δx), as the probability of a hydrometeor being detected at some point with altitude $h1$ and at some second point with altitude $h2$ and offset by Δx in the horizontal, without loss of generality. Likewise the two-point probability density function of radar reflectivity becomes a five-dimensional variable, R_{12} ($h1$, $h2$, Δx, $Z1$, $Z2$), which gives the probability that (when a hydrometeor is present at the two points specified by $h1$, $h2$ and Δx) the reflectivity at the first point will be $Z1$ and at the second point will be $Z2$. We show examples of the functions P_{12} and R_{12} in section 4.

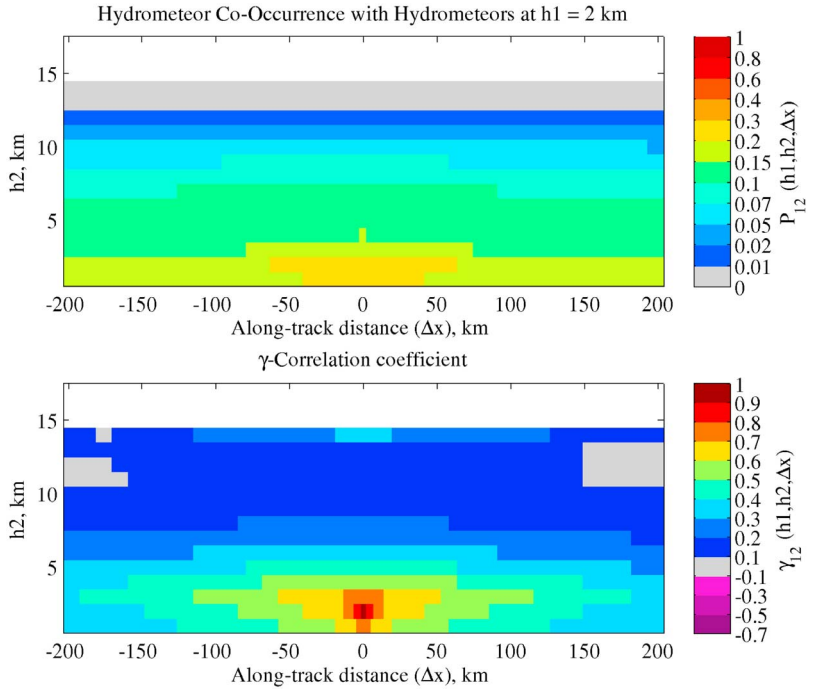

Figure 2. North Pacific. (top) Two-point probability of hydrometeor co-occurrence, P_{12} ($h1$, $h2$, Δx), where $h1$ = 2 km. (bottom) Gamma correlation coefficient.

[11] While we can estimate the function R_{12} from CloudSat observations, it is difficult to visualize or compare five-dimensional variables and so we will also use the traditional linear (or Pearson) correlation coefficient, ρ_{12}, defined as

$$\rho_{12}(h1,h2,\Delta x) \equiv \frac{\mathrm{Cov}(Z1,Z2)}{\sqrt{\mathrm{Var}(Z1)\mathrm{Var}(Z2)}} = \frac{\langle Z1Z2\rangle - \langle Z1\rangle\langle Z2\rangle}{\sqrt{(\langle Z1^2\rangle - \langle Z1\rangle^2)(\langle Z2^2\rangle - \langle Z2\rangle^2)}}, \quad (1)$$

where $Z1$ is the radar reflectivity at the first point, $Z2$ is the radar reflectivity at the second point, and the angle brackets $\langle\rangle$ indicate the mean value. We will see in section 4 that ρ_{12} provides a useful summary of the relationship between radar reflectivity (and rain rate) at two points in the horizontal and for short distances in the vertical.

[12] Also, while P_{12} describes the spatial structure of hydrometeor occurrence, in some respects it is more insightful to examine how much more or less likely one is to have a hydrometeor at both points than would be the case if the probability of co-occurrence at the two points were uncorrelated or independent of each other. To that end, we introduce the following correlation coefficient (which hereafter we will call gamma correlation):

$$\gamma_{12}(h1,h2,\Delta x) = \frac{P_{12}(h1,h2,\Delta x) - P(h1)P(h2)}{\min[P(h1),P(h2)] - P(h1)P(h2)} \quad (2a)$$

if $P_{12}(h1, h2, \Delta x) \geq P(h1)P(h2)$, and

$$\gamma_{12}(h1,h2,\Delta x) = \frac{P_{12}(h1,h2,\Delta x) - P(h1)P(h2)}{1 - \max[P(h1)+P(h2),1] + P(h1)P(h2)}, \quad \text{otherwise.} \quad (2b)$$

Here $P(h)$ is the (single point or background) probability of hydrometeor occurrence at altitude h (which is equivalent to P_{12} (h, h, Δx = 0)). We discuss the development of the gamma correlation coefficient in more detail in Appendix A and merely note here that gamma coefficient is similar to the linear coefficient (ρ) except the denominator has been changed to ensure gamma maintains a value between 1 and −1, where 1 indicates that P_{12} has the largest possible value that it could have and −1 indicates that P_{12} has the smallest possible value it could have. For simplicity, we use the terms "correlation" and "offset" in this paper when referring to the gamma function (equation (2)) and Δx, respectively, rather than the terms "autocorrelation" and "lag," which are sometimes used and are equivalent in the present case.

[13] All of the results shown in this paper are constructed using all CloudSat observations gathered during the 4 year period, December 2006 through November 2010. The CloudSat data gathered from the start of the mission (May 2006) through August 2006 is of lower quality because of problems related to radar beam pointing, which resulted in periods with enhanced ground clutter (from specular surface reflection) and second trip echoes. These problems where solved by early August 2006. Nonetheless we have opted not to include the period August 2006 through October 2006 in order to ensure that each season is equally weighted in the results presented here. The objective of this paper is to provide a first examination of hydrometeor and rain spatial structure; the analysis of seasonal and annual cycles is left to future research.

4. The North Pacific and Tropical Western Pacific

[14] In this section we contrast the hydrometeor spatial structure of the North Pacific (NP; 30°N to 60°N, 160°E to

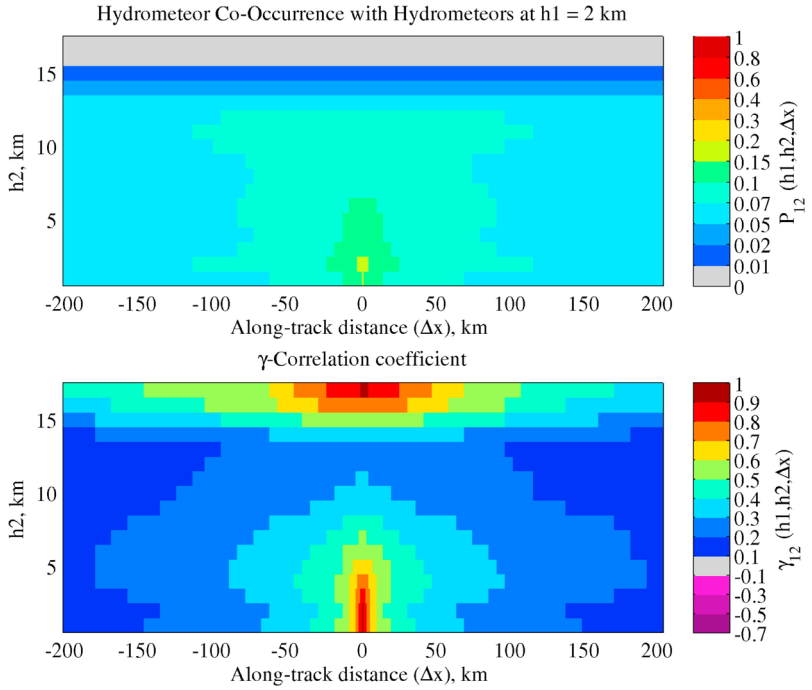

Figure 3. The same as Figure 2, but for the tropical western Pacific.

140°W) and tropical western Pacific (TWP; 10°S to 10°N, 130°E to 170°W) as a prelude to examining zonal mean structures in section 5. We begin with an examination of hydrometeor co-occurrence followed by an examination of reflectivity distributions, and finally surface (column) precipitation occurrence and rain rate.

4.1. Hydrometeor Co-occurrence

[15] Figures 2 (top) and 3 (top) show the two-point probability of hydrometeor co-occurrence (P_{12}) derived from CloudSat retrievals for the case where one point is located at an altitude of 2 km (that is, $h1 = 2$ km) in the NP and TWP, respectively. Not surprisingly, the baroclinic NP and barotropic TWP atmospheres show dramatically differing spatial structures. The probability of hydrometeor co-occurrence for a hydrometeor at $h1 = 2$ km in the NP is larger at altitudes $h2 < 7$ km and smaller at altitudes $h2 > 9$ km than in the TWP. In some respects this result is trivial in that such follows from the simple fact that hydrometeors fill a larger fraction of the atmosphere in the NP below 7 km and a smaller fraction above 9 km, than in the TWP. For this reason, it is more insightful to examine the associated gamma correlation (Figures 2, bottom, and 3, bottom), which indicates how much more likely one is to find a hydrometeor at the second point than one would expect of uncorrelated data. The gamma correlation function makes clear that presence of hydrometeors at $h1 = 2$ km in the NP is strongly correlated with the presence of hydrometeors at this same altitude and in adjacent layers over long horizontal distances, much longer than in the TWP. The correlation in the NP is somewhat stronger with the layer just above ($h2 = 3$ km) than below it ($h2 = 1$ km), while the opposite is true in the TWP. As one would expect, the TWP has stronger correlation in the vertical ($\Delta x = 0$) between hydrometeors at 2 km and those above 4 km. In particular, there appear to be oddly large correlation values at $h2 > 14$ km. To a large degree, these large values are a result of the CloudSat sensitivity limit. The probability of co-occurrence shown in Figure 3 (top) has small values at these same altitudes. Ground-based radar and lidar observations, as well as lidar observations from the NASA CALIPSO satellite, show there are a lot of hydrometeors at these altitudes that are not detected by CloudSat because of the low reflectivity of this condensate. When CloudSat does detect hydrometeors at these altitudes, they are usually part of a deep convective system and hence are strongly correlated with detections lower in the atmosphere. So while the CloudSat detections at 15 km are real, the correlation is artificially high relative to what a radar with better sensitivity would likely produce. Nonetheless, we expect that a radar simulator using a −30 dBZe threshold when used in conjunction with a cloud-resolving model should be able to reproduce this effect.

[16] The results in Figures 2 and 3 are for the case where $h1 = 2$ km and the coordinates $h2$ and Δx are allowed to vary. In Figures 4 and 5, we examine gamma for the case $h1 = h2$, which we will hereafter term horizontal correlation, for the NP and TWP, respectively. As a result of the horizontal invariance assumption used in constructing these data, both the probability of co-occurrence and gamma correlation coefficient must be symmetric about $\Delta x = 0$, and so we display the data only for positive values of Δx. In both regions, we find the gamma correlation is strongest in the middle troposphere (a result we find to be true globally, as discussed further in section 5). We further find that one can fit an equation of the form

$$\gamma(h, \Delta x) = \exp\left[-(|\Delta x|/a(h))^{b(h)}\right] \quad (3)$$

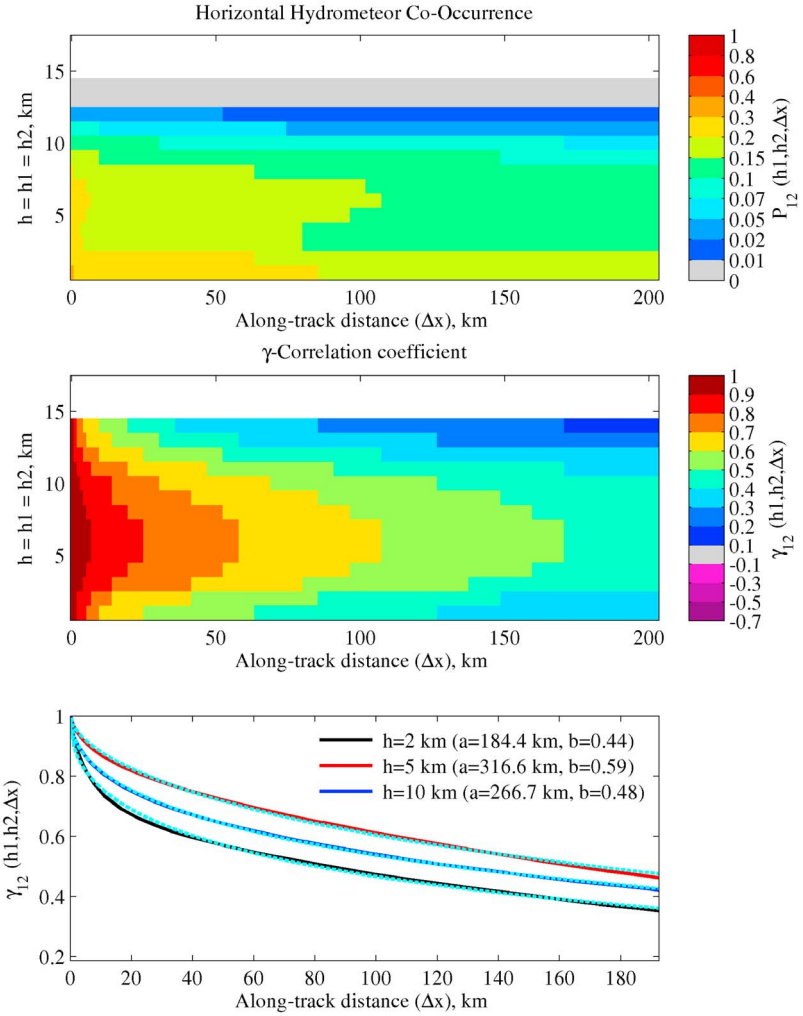

Figure 4. North Pacific. (top) Two-point probability of hydrometeor co-occurrence, P_{12} ($h1$, $h2$, Δx), where $h1 = h2$. (middle) Gamma correlation function. (bottom) Plot of gamma correlation at the altitudes listed in the legend. The dashed cyan lines are a fit using the parameters a (correlation length) and b (shape parameter); see text.

to the observed data rather well, as shown for a few altitudes in Figures 4 (bottom) and 5 (bottom). Here a and b are coefficients that vary with height. In Figures 4 and 5, the dashed cyan lines show the fit with the resulting coefficients a and b listed in the legend. *Gebremichael and Krajewski* [2004] and *Ciach and Krajewski* [2006] successfully used this form in modeling rainfall spatial structures using radar and rain gauge data. Following these authors, we call the "a" parameter the correlation length (which defines the distance at which the correlation drops to 1/e) and the "b" parameter the shape factor or shape parameter. Figure 4 shows the horizontal correlation length in the NP exceeds 300 km in the midtroposphere while values in the Tropical Western Pacific are less than 150 km. The shape parameter has a strong effect on the correlation function. Values less than one (such as those found here) denote a correlation function that decays faster than the exponential function at distances less than the correlation length and more slowly at distances greater than the correlation length. (In Appendix B, we discuss the shape parameter and its relationship to the associated spatial power spectrum). In section 5, we use the correlation length and shape parameter to characterize zonal mean hydrometeor structure.

[17] In many regions hydrometeors tend to be preferentially located at certain altitudes, which are revealed by distinct peaks in the hydrometeor occurrence profile. Most commonly there are distinct peaks in the lower and upper troposphere (often called the low-cloud or high-cloud peak for obvious reasons). We note that the location of these peaks does not necessarily align with the largest correlation length. In the TWP, for example, Figure 5 (top) shows a peak in hydrometeor occurrence near 11 km, while the correlation length (Figure 5, bottom) is longest near 8 km. Physically this means that while clouds (and precipitation) are less frequently found at 8 km, when they occur they tend to be more spatially compact (less spread out or less broken).

[18] Figure 6 compares the vertical gamma correlation of the NP and TWP (that is, $\Delta x = 0$ is fixed while $h1$ and $h2$ are allowed to vary). By construction the diagonal of these plots has a value of 1 (for the simple reason that if there are hydrometeors $h1 = 2$ km then the probability there are hydrometeors at the same point, $h2 = 2$ km, is 100%) and they

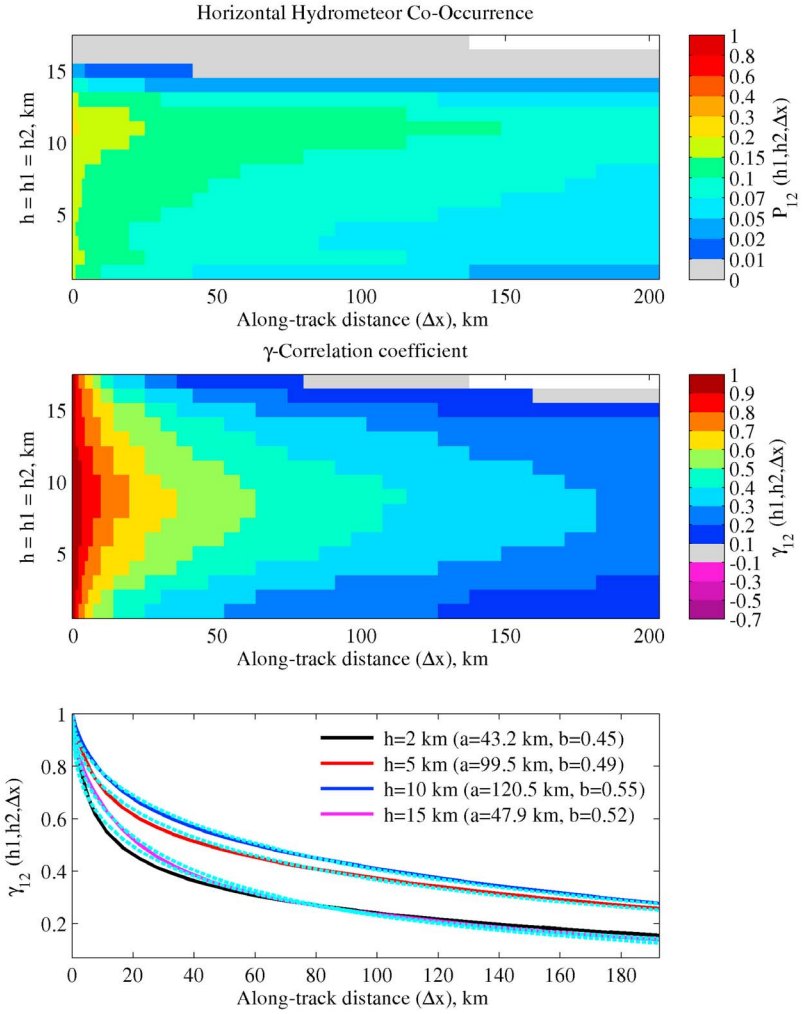

Figure 5. The same as Figure 4, but for the tropical western Pacific.

are symmetric about the diagonal. Figure 6 shows that the correlation in the vertical is generally stronger in the TWP than in the NP, as was noted earlier for the case of $h1 = 2$ km. In the NP, the strength of the vertical correlation increases with altitude, visible in Figure 6 as an increasing width of the distribution about the diagonal for larger values of h. The artificially high correlation between detections above 14 km with those at lower altitudes in the TWP (also discussed in section 3) is likewise prominent, as one would expect on the basis of the previous discussion. Unfortunately, the exponential form introduced in equation (3) does not generally fit the observed change in correlation with vertical distance (Δh) very well, as the correlation tends to reach a minimum and then increase or remain steady as $h2$ approaches the tropopause. For small values of Δh ($\sim$1 to 4 km) and where $h2$ is far from the tropopause, the fit may be satisfactory (not shown). Even with this restriction, however, the correlation length parameters vary significantly with $h1$ and we would caution against using a single value for all altitudes. *Oreopoulos and Norris* [2011] similarly found that for the same separation distance the overlap (correlation) parameters are significantly different for various atmospheric layers.

4.2. Hydrometeor Reflectivity

[19] As discussed in section 3, the two-point probability density function (pdf) of radar reflectivity is a five-dimensional variable. Here we examine this function by fixing the position variables ($h1$, $h2$, and Δx) at select values and display the resulting joint distribution of reflectivity as a color image. Our goal in this section is not to provide an exhaustive discussion of the pdf, but rather to briefly examine some primary characteristics and in particular to gauge the degree to which the traditional linear correlation coefficient is a useful way to characterize the joint distributions. We begin by examining how the joint reflectivity distribution varies in the horizontal (i.e., $h1 = h2$). Figure 7 displays CloudSat results for the TWP where the altitude is fixed at 8 km and Δx has values ranging from 1.1 km to 44 km (depending on the panel). For the case of $\Delta x = 1.1$ km, Figure 7a shows a concentration of high values along the diagonal. This indicates that when a given value of $Z1$ is observed (e.g., 10 dBZe), there is a high probability that $Z2$ will have a value close to this same value ($\sim$10 dBZe). We note that the width of the distribution (about the diagonal) tends to be somewhat smaller at large reflectivity values than

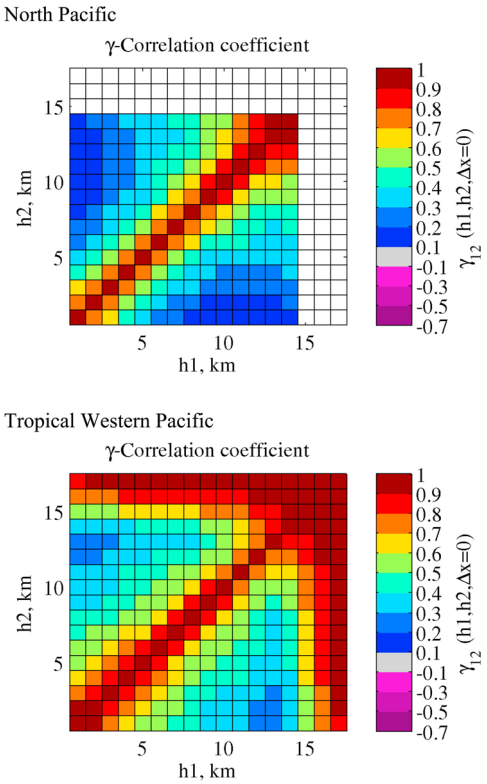

Figure 6. Gamma correlation coefficient in the vertical (i.e., $\Delta x = 0$). (top) North Pacific. (bottom) Tropical western Pacific.

at small reflectivity values, which means the spatial correlation is stronger for larger reflectivities. Calculated over all reflectivity values, the linear correlation coefficient (equation (1) given in section 3) for this case is quite large, 0.95. Figures 7a–7d show that as Δx increases, the values along the diagonal progressively decrease and the width of the distribution (around the diagonal) increases. Accordingly the linear correlation coefficient decreases and by Δx = 44 km has dropped to a value of 0.32, which is below the threshold used to define the correlation length. The details of the (single point) reflectivity distribution change with altitude and location. For example, we see here the most likely reflectivity value that one will encounter in the TWP at 8 km is near 5 dBZe, which is not true elsewhere. Nonetheless, we observe the same general pattern of smoothly going from a well defined linear relationship to the broad (or diffuse) pattern one would expect of independent data is universal in the horizontal.

[20] As was the case for hydrometeor occurrence, the linear correlation in the horizontal smoothly decreases as Δx increases and is reasonably well fit by the simple two-parameter model given by equation (3), as shown in Figure 8. As one might expect, the horizontal correlation length of reflectivity tends to be longer in the NP than in the TWP; and the correlation length of hydrometeor reflectivity is much shorter than the correlation length of hydrometeor occurrence at all altitudes. For example, comparing Figure 5 (bottom) with the second panel in Figure 8 (bottom), we see the horizontal correlation length of hydrometeor occurrence is 43 km at an altitude of 2 km, while the horizontal correlation length

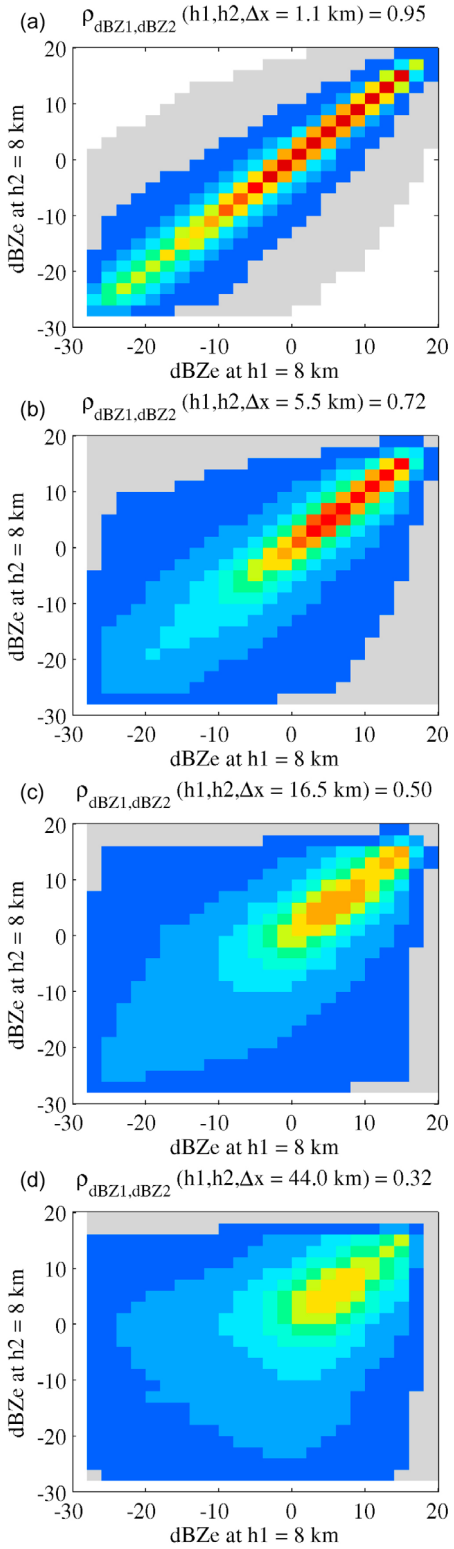

Figure 7. Tropical western Pacific. Joint histograms of radar reflectivity at two points both located 8 km above sea level ($h1 = h2 = 8$ km). Shown are histograms for points separated by along-track distances (Δx) of (a) 1.1, (b) 5.5, (c) 16.5, and (d) 44 km. Colors are logarithmic, with the maximum (joint) probability in red and decreasing probability (by a factor of 10) going to orange, yellow, green, light blue, dark blue, and gray.

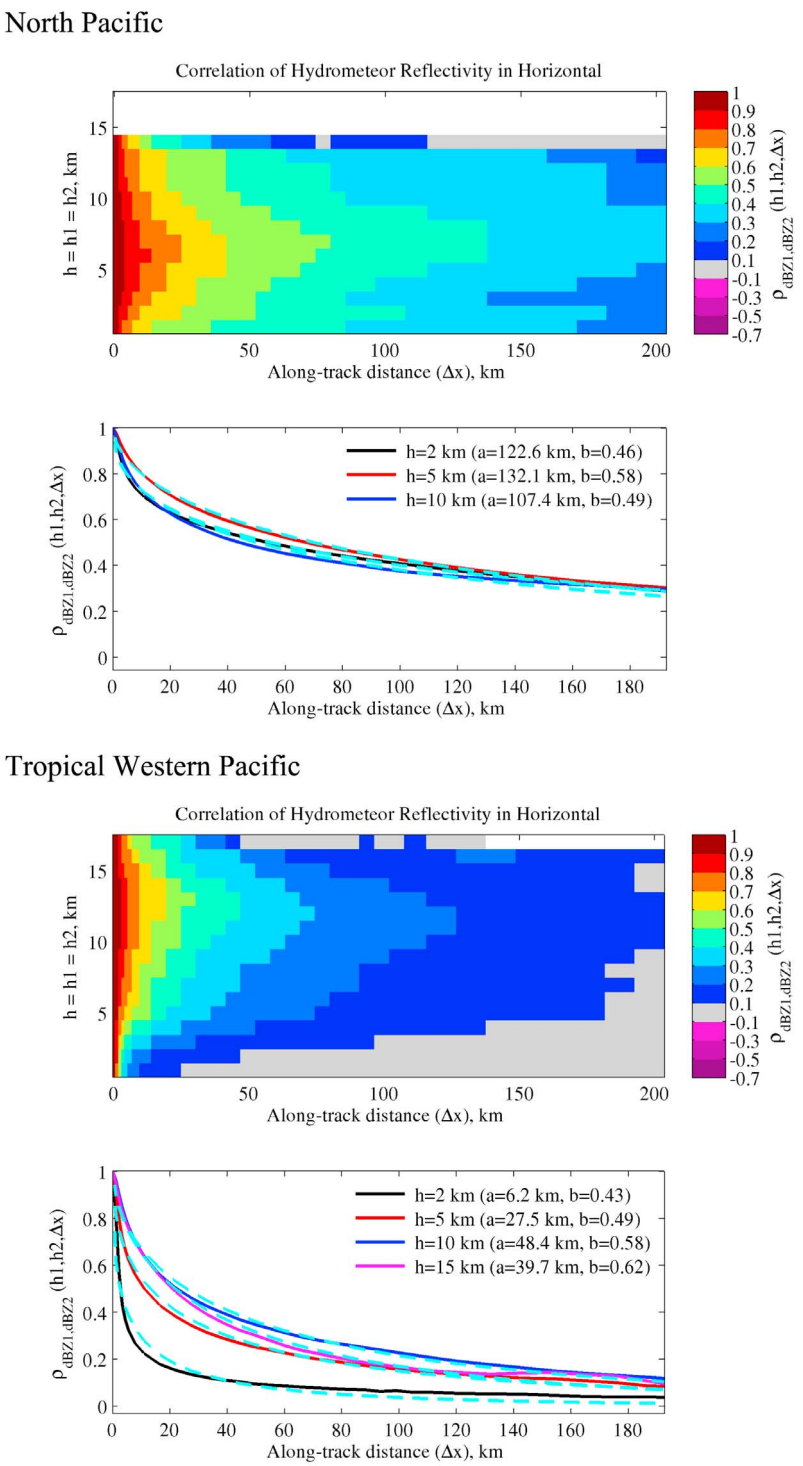

Figure 8. Linear correlation of radar reflectivity along the horizontal for (top) the North Pacific and (bottom) tropical western Pacific.

of reflectivity is 6.2 km at this same altitude. The shape parameters for occurrence and reflectivity, on the other hand, have relatively similar values generally between about 0.4 and 0.6, indicating that the correlation of both quantities initially decreases rapidly at scales less than the correlation length. See Appendix B for further discussion.

[21] The two-point pdf of radar reflectivity in the vertical displays more complicated behavior than in the horizontal. Figure 9 shows CloudSat results for the TWP where the altitude of the first point is fixed at $h1 = 8$ km and $\Delta x = 0$. The altitude of the second point, $h2$, varies in each panel and is listed along the vertical axis. For the case of $h2 = 7$ km (the

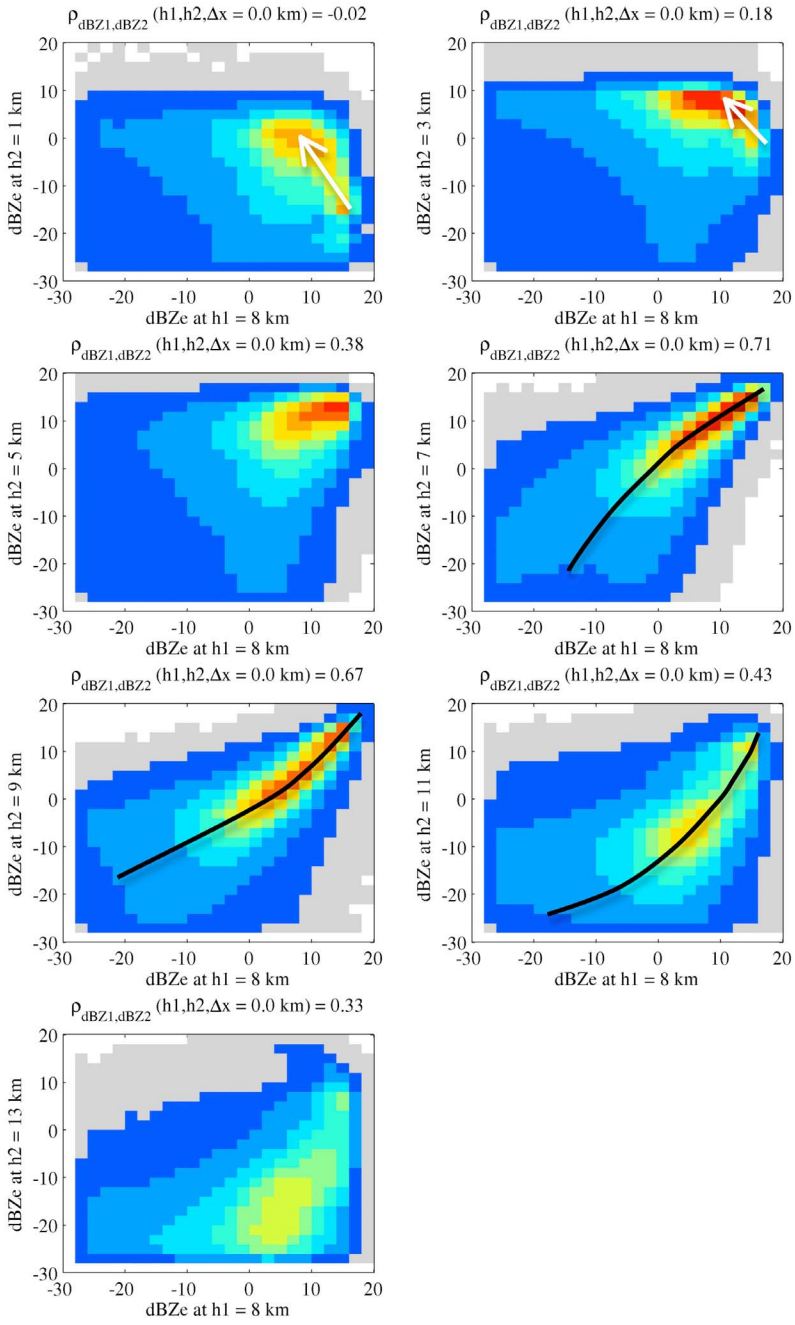

Figure 9. Tropical western Pacific. Joint histograms of radar reflectivity at two points, where one point is located 8 km above sea level ($h1 = 8$ km) and the second point is located at several different altitudes in the same radar column ($\Delta x = 0$). Shown are histograms for the second altitude ($h2$) located at 1, 3, 5, 7, 9, 11, and 13 km. White arrows highlight attenuation, while black lines ($h2 = 7$, 9, and 11 km) highlight nonlinearity in nearby layers (see text).

layer just below $h1$) and $h2 = 9$ km (the layer just above $h1$) there is a strong relationship. However, the relationship does not peak along the diagonal but rather follows a slightly curved path (highlighted in the each panel by a black line) with the reflectivity at the higher altitude being smaller. When viewing individual events, one frequently observes a decrease in reflectivity values with increasing altitude above the freezing level. This occurs because ice water contents and particle sizes tend to be smaller at higher altitudes, both of which result in lower radar reflectivities. These trends are due to a variety of factors such as a reduction in the water vapor available to be condensed at higher altitudes and ice particles growing in size through aggregation as they fall. As was the case in the horizontal, the width of the distribution is generally narrower for high reflectivity values. When $h2$ is more than 1 km above $h1$ (represented in Figure 9 by the panels with $h2 = 11$ km and $h2 = 13$ km), the peak in the pdf shifts

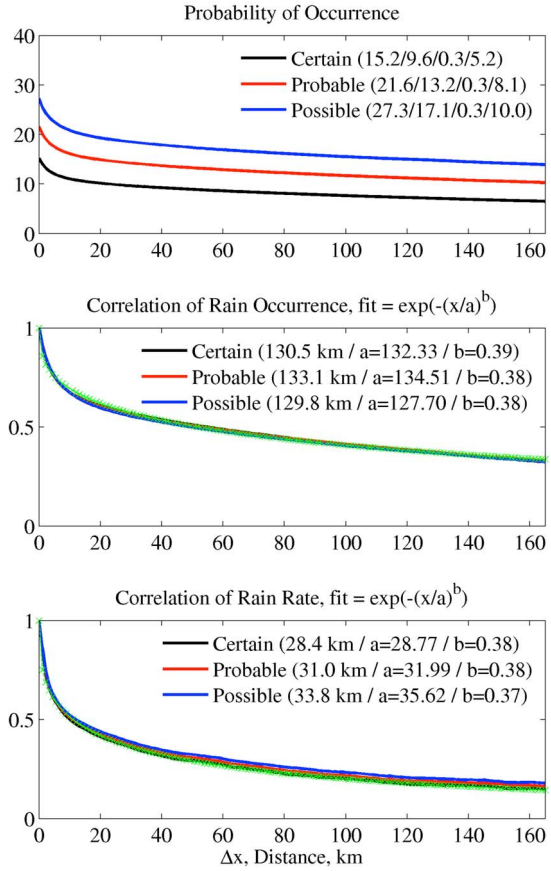

Figure 10. North Pacific. (top) Two-point probability of rain co-occurrence based on the CloudSat column precipitation product using rain-certain, rain-probable, and rain-possible confidence thresholds. Legend lists single-point probability of rain detection; see text. (middle) Correlation of rain occurrence. (bottom) Linear correlation of rain rate. Green line shows two-parameter fit to curve with rain-certain threshold.

farther from the diagonal and the shape of the distribution becomes broader.

[22] At the freezing level (about 5 km in the TWP) there is frequently a change in the observed reflectivity as ice particles begin to melt. While precipitation radars often show an increase in reflectivity values (including, at times, a bright band due to large ice particles coated by water), 94 GHz radars such as CloudSat more often show a decrease in reflectivity. This occurs because attenuation is much larger at 94 GHz than at lower frequencies, especially for liquid water. From the downward viewing perspective of CloudSat, the attenuation in any layer affects reflectivity values for all lower layers. While at times a "peak" in the CloudSat radar reflectivity is observed at the freezing level, this peak is not a bright band in the traditional sense, meaning a large reflectivity enhancement due to ice particles coated by water. Rather increasing attenuation below the top of the melting layer can create a relative maximum in the CloudSat reflectivity near the top of the melting layer [*Sassen et al.*, 2007]. As a result of the large microphysical change that occurs and strong liquid water attenuation, joint reflectivity distributions where one point is above the freezing level and the other is at or below look characteristically different than the case where both points are above the freezing level. This situation is captured in the panels of Figure 9 where $h2$ = 1, 3, and 5 km. In the case where $h2$ = 3 km, we see that when the reflectivity at $h1$ = 8 km is near 15 dBZe (the value along the x axis), the most likely reflectivity observed at 3 km is near 0 dBZe, and as the reflectivity at 8 km decreases (toward 5 dBZe), the most frequently occurring value of reflectivity at 3 km actually increases to near 10 dBZe. This trend is highlight by a white arrow. This anticorrelation occurs because high dBZe values at 8 km are associated with larger amount of condensed water which generate more attenuation. When the reflectivity at 8 km is below about 5 dBZe the affect of attenuation on reflectivities at lower altitudes becomes modest in the TWP; however, Figure 9 also shows there is essentially no correlation between the reflectivity at 8 km and at 3 km when this is the case.

4.3. Rain Rate and Rain Occurrence

[23] Figures 10 (top) and 11 (top) show the probability of surface precipitation co-occurrence for the NP and TWP, respectively, determined from the CloudSat column precipitation product. In Figures 10 and 11, three lines are shown that correspond to precipitation occurrence identified on the basis of the precipitation-certain, precipitation-probable, and precipitation-possible thresholds (see section 2). For each category, the legend provides four values that indicate (from

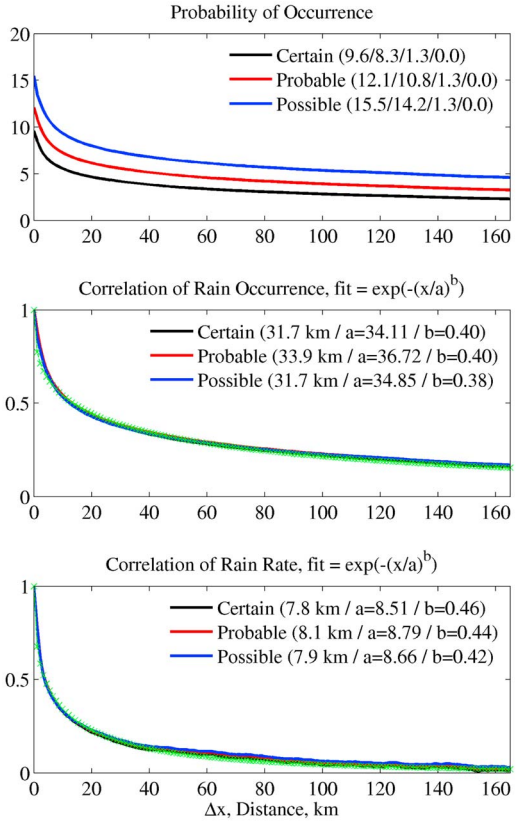

Figure 11. Same as Figure 10, but for the tropical western Pacific.

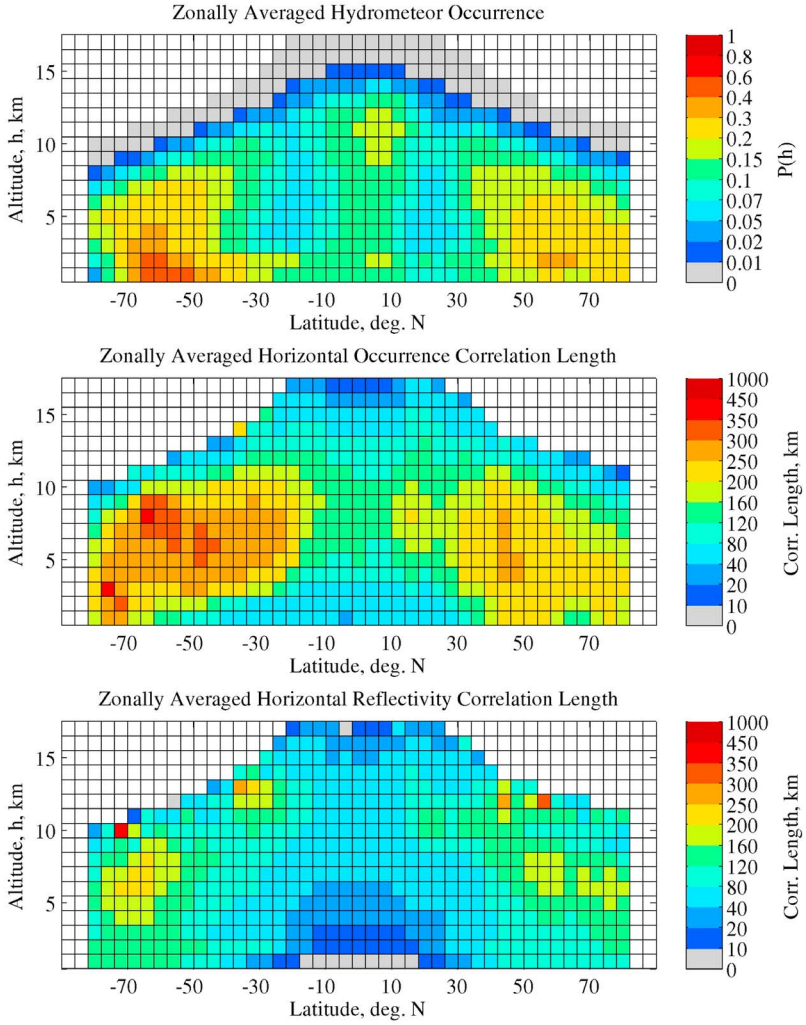

Figure 12. Zonally averaged parameters. (top) Probability of hydrometeor occurrence with altitude. (middle) Hydrometeor occurrence horizontal correlation length. (bottom) Hydrometeor reflectivity horizontal correlation length.

left to right): (1) the probability of precipitation for the region (which naturally is also the value plotted at $\Delta x = 0$), (2) the probability of a successful rain rate retrieval, (3) the probability of an unsuccessful rain rate retrieval due to the radar pulses becoming attenuated (PIA limit was reached), and (4) the probability of an unsuccessful rain rate retrieval due to the presence of mixed-phase precipitation or snow at the surface. The sum of the last three numbers is equal to the first number (the total probability of precipitation). In the NP, there is no rain rate retrieval for a significant fraction of the detected precipitation events because of a mixed phase or snow present at the surface (about one third missing) and significant caution should be exercised in using the CloudSat rain rate retrievals for this region. On the other hand, we see that there is only a small fraction of the time during which the CloudSat radar is attenuated to the point where the path integrated attenuation (and rain rate) cannot be determined (0.3% for the NP and 1.3% for the TWP). This is not to suggest that these heavier rains are unimportant. The rain rate tends to be exponentially distributed such that the area covered by rain is dominated by light rains while the volume of rain is dominated by heavy rains [*Berg et al.*, 2010]. It is also because of the exponential rain rate distribution that we expect to see a significantly larger likelihood of probable or possible precipitation than certain precipitation in the CloudSat product, although there may well be a significant number of false positives in precipitation possible counts, as well. In spite of this notable change in the probability of occurrence (or equivalently fractional area-covered by precipitation) with threshold, we find that the precipitation occurrence correlation functions (Figures 10, middle, and 11, middle) and the rain rate correlation functions (Figures 10, bottom, and 11, bottom) change little with these thresholds. This is true globally, except in subtropical stratocumulus zones (none shown) where we find that lighter rains do appear to have longer correlation lengths.

[24] As was the case for hydrometeor horizontal occurrence and reflectivity, the precipitation occurrence and rain rate correlation functions are reasonably well fit by the simple two-parameter exponential model given by equation (3). As one might expect, Figures 10 and 11 show that the correlation length of precipitation occurrence is longer than that

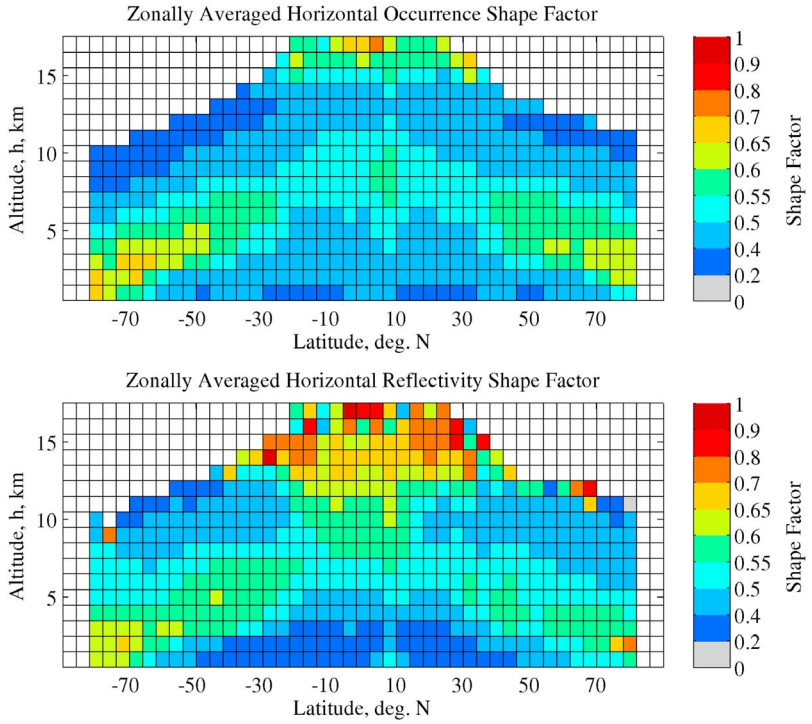

Figure 13. Zonally averaged shape parameters. (top) Hydrometeor occurrence in the horizontal. (bottom) Hydrometeor reflectivity in the horizontal.

of rain rate; and correlation lengths are longer in the NP than in the TWP.

[25] Not surprisingly, the correlation length of hydrometeor occurrence near the surface (e.g., at 1 km or 2 km above sea level) is longer than the correlation length of surface precipitation occurrence, as can be seen by comparing Figures 4 and 5 with Figures 10 and 11. Perhaps somewhat less intuitive is that the correlation length of horizontal hydrometeor reflectivity in the North Pacific is about 130 km at an altitude of 2 km (Figure 8), which is much longer than the correlation length of rain rate (about 30 km, Figure 10). The difference in these correlation lengths is possible because the CloudSat rain rate retrieval is not driven by the measured reflectivity (as would be the case for rain rate retrievals from a precipitation radar), but rather by the path integrated attenuation; and the CloudSat rain rate retrieval is able to capture variability in the rain rate that is independent of variability in the observed reflectivity. In the TWP, we find the correlation length of hydrometeor reflectivity in the horizontal is about 9 km (at an altitude of 2 km) which is very close to that of the column rain rate, about 8 km. This similarity may be due to the fact that most of the near-surface hydrometeors detected by CloudSat (in this region) are precipitation (or a mixture of cloud and precipitation), because CloudSat is unable to detect most small cumulus.

[26] Comparison of CloudSat results with those from other sources, such as ground-based radars or rain gauge networks, is complicated by differences in the spatial and temporal scales involved [*Ciach and Krajewski*, 2006]. Shape parameter in particular appears to have a strong scale dependence. Nonetheless, a rain rate correlation length of 30 km for the North Pacific and 8 km for the tropical western Pacific is broadly consistent with expectations from ground-based radar systems in other midlatitude and tropical regions. For example, *Overeem et al.* [2009] found a rain rate correlation distance of 36 km on the basis of a 10 year data set of scanning precipitation radar measurements from the Netherlands, while *Gebremichael and Krajewski* [2004] found a correlation length of 4 km to 5 km (with an estimated uncertainty of about 2 km) for central Florida in August and September (during the TEFLUN-B experiment), as well as in Brazil (during the TRMM Large-Scale Biosphere–Atmosphere field experiment).

5. Zonal Averages

[27] In this section we examine the zonally averaged spatial correlation of hydrometeor occurrence, reflectivity, and rain rate, beginning with a discussion of hydrometeor correlation in the horizontal (i.e., $h1 = h2$), followed by correlation in the vertical ($\Delta x = 0$), and ending with precipitation occurrence and column rain rates. These data are zonally averaged in the sense that all CloudSat observations in each latitude band are used to construct the relative frequencies and variances needed in equations (1) and (2) to determine the correlation coefficients. Figure 12 (top), for example, shows the zonally averaged (single-point) probability of hydrometeor occurrence with altitude used in equation (2). We do not calculate the correlation coefficients on a fixed latitude and longitude grid and then average these data in longitude, because the CloudSat sampling is sufficiently sparse that calculation of the spatial correlation tends to be noisy unless measurements from many overpasses are combined. The results shown in this section are on a 4° latitude grid. We discuss sampling issues further in section 6.

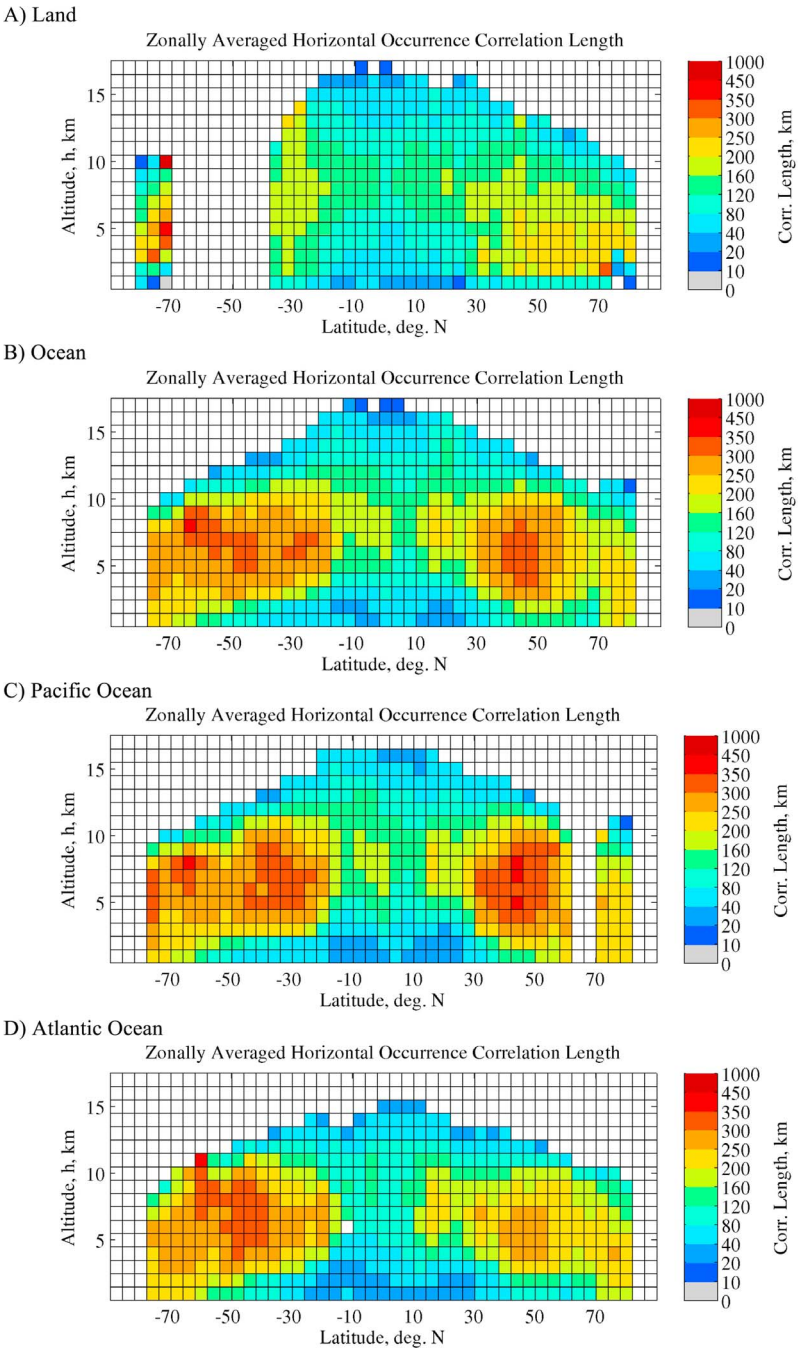

Figure 14. Zonally averaged hydrometeor occurrence correlation length restricted to (a) land, (b) ocean, (c) Pacific Ocean (120°W–145°E), and (d) Atlantic Ocean (0°–50°W).

5.1. Horizontal Correlation

[28] Following the same approach used in section 4, we fit the simple two-parameter model given by equation (3) to the correlation function of hydrometeor occurrence and reflectivity in the horizontal. The resulting correlation lengths are shown in Figures 12 (middle) and 12 (bottom), and the shape parameters are shown in Figure 13. (We note that the handful of points near the top of the troposphere with particularly large or small values are due to noise in the fitting procedure due to the very low frequency of hydrometeor detection in these regions, as denoted by the gray area in Figure 12, top).

[29] Overall, correlation lengths tend to be longer in middle and high latitudes than in the tropics and subtropics. While hydrometeor occurrence also tends to be high in middle and high latitudes (Figure 12, top), the pattern of correlation length (in hydrometeor occurrence or reflectivity) is not aligned with the pattern of hydrometeor occurrence. Rather, correlation lengths peak in the midtroposphere at most latitudes, including in the subtropics, but with the correlation length of hydrometeor reflectivity (Figure 12, bottom) peaking higher in the atmosphere than for hydrometeor occurrence (Figure 12, middle) at most latitudes.

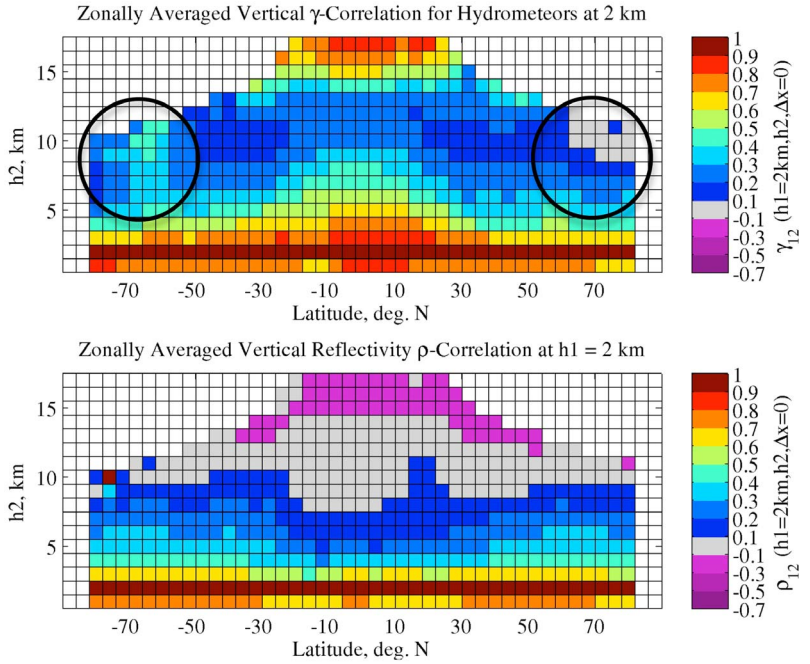

Figure 15. Zonally averaged vertical correlation of (top) hydrometeor occurrence and (bottom) hydrometeor reflectivity for hydrometeors at 2 km above mean sea level with hydrometeors at other altitudes. Black circles in Figure 15 (top) highlight the stronger correlation in occurrence in the Southern Hemisphere high latitudes as compared with the Northern Hemisphere.

These results are consistent with the more detailed discussion presented in section 4 for the NP and TWP.

[30] The correlation length of hydrometeor occurrence is not symmetric between the Northern and Southern Hemispheres. It is generally longer in the southern middle and high latitudes (latitudes less than about −20°) than in the Northern Hemisphere middle and high latitudes (latitudes greater than about 20°). As shown in Figures 14a and 14b, some of this difference is related to differences in hydrometeor structure between land and ocean. At least in the zonal mean, hydrometeors over land have shorter correlation lengths and are longest below 5 km, whereas correlation lengths over ocean tend to peak higher in the atmosphere. Nonetheless, restricting the calculation of hydrometeor occurrence correlation length to oceanic regions (Figure 14b), we find the correlation lengths remain notably different between the two hemispheres. There are also striking differences between hydrometeor structures over the Pacific and Atlantic Oceans (Figures 14c and 14d), including a peak in occurrence (but not reflectivity) correlation length in the Northern Hemisphere subtropics (near latitude 16°) in the altitude range between 5 and 10 km which is not present (or is notably weaker) in the southern subtropics.

[31] This analysis incorporates all observations, regardless of synoptic conditions, which makes the sources of these differences unclear. Nonetheless, we speculate that the differences between the southern and northern midlatitudes are due primarily to differences in the spatial structure of extratropical cyclones. Recent research by *Barnes and Hartmann* [2011] on the spatial structure of cyclone vorticity fields also shows significant differences between the Northern and Southern Hemispheres, as well as between the Pacific and Atlantic sectors.

[32] At middle and high latitudes, the shape parameters for hydrometeor occurrence (Figure 13, top) and reflectivity (Figure 13, bottom) have a similar pattern and one that is more (though not entirely) symmetric with respect to hemisphere than the correlation lengths. In the tropics, the shape parameter for reflectivity (Figure 13, bottom) is smaller near the surface and larger in the upper troposphere than the shape parameter for occurrence (Figure 13, top). Near the surface both the reflectivity correlation length and shape parameters indicate that the correlation of hydrometeor reflectivity reduces much more quickly than hydrometeor occurrence. In the tropical upper tropopause, on the other hand, we find the correlation length of occurrence and reflectivity are about the same (about 50 km; see Figure 12), while the shape factor is larger for reflectivity correlation. Physically this means that while reflectivity and occurrence become decorrelated at about the same distance, the correlation of reflectivity decays more slowly than hydrometeor occurrence, indicating that hydrometeors in the upper troposphere tend to have similar reflectivity values even as cloud (condensate) fields become broken and the existence of a cloud in one radar column is not a strong predictor of clouds in nearby columns.

5.2. Vertical Correlation

[33] As was discussed in section 4, the vertical spatial correlation of neither hydrometeor occurrence nor reflectivity is well fit by the two-parameter exponential model. As a result, it is more difficult to summarize vertical correlation than was the case in the horizontal and we opt here to concentrate on the case where $h1 = 2$ km (shown in Figure 15).

[34] The Figure 15 (top) shows the zonally averaged (gamma) correlation of hydrometeor occurrence for a hydrometeor at an altitude of 2 km, with hydrometeors at

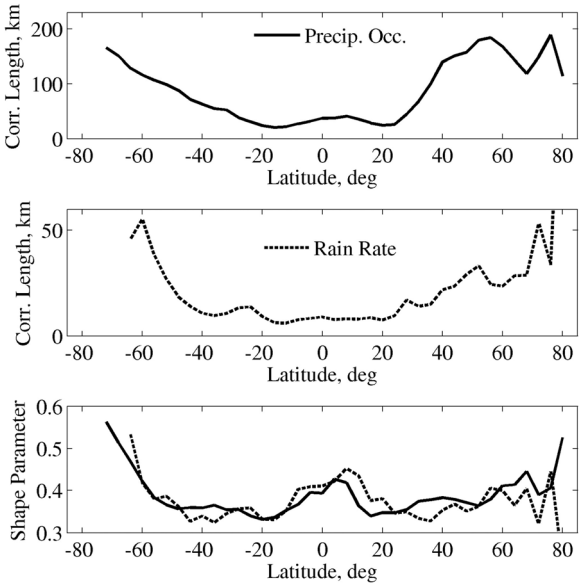

Figure 16. (top) Zonally averaged horizontal correlation length of precipitation occurrence and (middle) rain rate. (bottom) Associated shape parameter for precipitation occurrence (solid line) and rain rate (dashed line).

other altitudes. By definition, hydrometeor occurrence at 2 km is perfectly correlated with itself, and as one would expect from the discussion in section 4 there is a strong correlation with hydrometeors in neighboring (∼1 km) layers at all latitudes. In the tropics (and to a lesser degree in the subtropics), hydrometeor occurrence at altitudes above 4 km is more strongly correlated with hydrometeor occurrence at 2 km than at middle and high latitudes, which is characteristic of convection. As was the case in the TWP, hydrometeors at 2 km are strongly (gamma) correlated with hydrometeors at 14 km to 16 km because CloudSat tends to detect hydrometeors above 14 km only as part of deep convective systems. The vertical correlation structure of the Northern and Southern Hemisphere extratropics is largely similar, with the notable exception of stronger vertical correlation in occurrence between hydrometeors at 2 km and those above 8 km in the southern high latitudes as compared with the Northern Hemisphere (see black circles in Figure 15, top). These characteristics (strong correlation with neighboring layers, stronger vertical correlation in the tropics, and stronger correlation in the Southern Hemisphere high latitudes than in the northern are true for all values of $h1$ (not shown).

[35] The zonally averaged vertical correlation of reflectivity is shown in Figure 15 (bottom). Here we see the correlation of reflectivity is generally much weaker than for hydrometeor occurrence. The tropics, in particular, show weaker correlation in reflectivity between hydrometeors at 2 km and above 4 km than was the case with hydrometeor occurrence. These low correlation values are due to the nonlinear relationship between reflectivities at different altitudes, as discussed in section 4.2. The weak negative correlation between reflectivity at 2 km and reflectivity values near the top of the troposphere at some latitudes is not noise, but rather is the result of radar attenuation (following the discussion in section 4.2). The strong correlation with neighboring layers and fairly symmetric results in the Northern and Southern Hemispheres shown here are characteristic for hydrometeors at all values of $h1$ (not shown).

5.3. Precipitation Occurrence and Rain Rate

[36] Figure 16 compares the zonally averaged precipitation occurrence and rain rate correlation lengths and shape parameters. As shown in section 4.3, the precipitation occurrence correlation parameters are similar for precipitation-certain, precipitation-probable, and precipitation-possible thresholds and so results only for the precipitation-probable threshold are shown. The correlation length of both precipitation occurrence and rain rate is longer in middle and high latitudes than in the subtropics and tropics (between latitudes 20°N and 20°S). However, the rain rate correlation length remains uniformly short throughout the subtropics and tropics (with a value near/below 10 km), while the precipitation occurrence correlation length shows a minor peak between latitudes 0° to 10°N, associated with the more extensive precipitation along the Intertropical Convergence Zone (ITCZ). Correlation lengths appear to be somewhat longer in the Northern Hemisphere midlatitudes than in southern midlatitudes (the opposite of what we find for hydrometeor occurrence at most altitudes). Both rain rates and precipitation occurrence retrievals are only run over ocean surfaces and are affected by mixed phase and snow, so the retrieval values at higher latitudes should be viewed with caution. In particular, the strange dip in the precipitation correlation length at 70°N is likely an artifact of these limitations. The shape factor for precipitation occurrence and rain rate are quite similar, with a peak in the tropics (along the ITCZ).

6. Discussion and Conclusions

[37] CloudSat observations are an attractive source for characterizing hydrometeor spatial structure because CloudSat provides a globally consistent data set in regards to spatial resolution, viewing geometry and calibration, as well as good (∼−28 dBZe) sensitivity to cloud and light precipitation. Nonetheless, the observations have limitations related to sensitivity and sampling. Even with its good sensitivity, CloudSat is unable to detect all hydrometeors, and in particular, many hydrometeors near the tropopause and in the boundary layer go undetected because of their inherently low radar reflectivity and in the boundary layer also because of surface clutter. In the future we plan to analyze hydrometeor occurrence using a combination of CloudSat radar and Calipso lidar observations, which are sensitive to most clouds (hydrometeors) that go undetected by the radar, when they are not obscured by higher altitude optically thick clouds.

[38] Our primary motivation for undertaking this study was to assess the ability of simple correlation statistics to capture regional differences that can be used in quantitative model evaluations, as well as provide insights into the nature of cloud and precipitation structure. The analysis demonstrates that the traditional linear (or Pearson) correlation coefficient provides a useful measure of the strength of the relationship between hydrometeor reflectivity at two points in the horizontal (that is, two points at the same altitude). The dependence of hydrometeor occurrence and reflectivity correlation (as well as precipitation occurrence and rain rate)

with horizontal offset (Δx) can be reasonably well fit with a simple two parameter exponential model. While the observed correlations in the vertical direction are also potentially useful, the relationship between reflectivity values at different altitudes is not linear (as revealed by the two point reflectivity distributions, Figure 9). Consequently, the linear correlation coefficient in the vertical is less easily interpreted and there is clearly room for developing additional statistics that better characterize the two-point probability distributions. Nonetheless, it is clear that the simple correlation statistics examined here are able to capture some regional differences and provide some insights into the nature of cloud and precipitation structure. Therefore, we expect these spatial statistics will prove useful in the evaluation of models, most prominently those models with horizontal resolutions of a few km or less. In particular, we anticipate comparing these statistics with Multiscale Modeling Framework (MMF) climate model simulations. In the MMF, a 2D or small 3D cloud resolving model is embedded into each grid cell of a GCM in place of traditional cloud parameterizations. It should be stressed that the observed spatial correlation statistics are dependent on the spatial resolution of the observations and our intent is to use a radar simulator to produce simulated radar fields and simulated correlation statistics with the same sensitivity and spatial resolution as the observations.

[39] The correlation length and shape parameters may also prove useful in the parameterization of subgrid hydrometeor structure for coarse resolution models [e.g., *Shonk et al.*, 2010]. Perhaps the largest impediments in directly using the correlation parameters derived here for subgrid parameterizations are (1) the radar reflectivity is a strong function of particle size and water content with contributions from both cloud and precipitating sized hydrometeors, whereas models likely need a way to parameterize separately subgrid cloud and precipitation water content variability and the overlap between these two classes of hydrometeors and (2) there is considerable seasonal variability in the correlation parameters (not shown in this paper), reflecting a strong dependence of the correlation parameters on dynamical and thermodynamical conditions, suggesting that using a single mean or even seasonal values may not be sufficient.

[40] The correlation parameters also have significant potential value in assessing uncertainty for regional averages, for example in rainfall accumulation [*North and Nakamoto*, 1989; *Steiner et al.*, 2003] or mean hydrometeor (cloud) fractions [*Key*, 1994; *Astin et al.*, 2001; *Settle and Van de Poll*, 2007]. With regard to sampling uncertainty in this study, there is a trade-off between the size of the region examined (and where we implicitly assume that cloud and precipitation properties are horizontally invariant) and the length of time it takes CloudSat to gather enough observations. We have not yet undertaken a quantitative uncertainty analysis of the derived correlation parameters, a task we leave to future research. Nonetheless, the results shown here demonstrate that correlation lengths of hydrometeor occurrence are from tens to a few hundred kilometers long; thus, a given CloudSat overpass provides only one or at most a few independent samples for a region that has a length of 1° in latitude. In all of the statistics presented in this paper, we required a minimum of 100 overpasses where hydrometeors were present at both points (specified by $h1$, $h2$, and Δx) before calculating any two-point statistics for that combination of $h1$, $h2$, and Δx – guaranteeing at least 100 independent samples. Since we expect the uncertainty drops with one over the square root of the number of independent samples (which is true for any random variable whose cumulative distribution is at least fourth-order differentiable), this suggests a relative error (standard deviation) of 10% or less. The length of time required to obtain 100 independent samples where there are hydrometeor detected at both points depends on the frequency of co-occurrence and the number of samples obtained in each overpass. For events that occur 5% of the time (a reasonable threshold for hydrometeor co-occurrence except perhaps near the top of the troposphere where CloudSat detections are sparse), that implies 500 to 2000 total overpasses are needed (depending on whether we assume one to four independent samples per overpass). For zonally averaged results, this means collecting data for a period of about 111 days to 27 days, so a seasonal to monthly scale is possible. A $10° \times 10°$ latitude by longitude region would require about 4 times as long, so seasonal to yearly. Obviously, if we want a smaller area or to ensure a smaller relative error or to capture the structure when hydrometeor co-occurrence is less than 5%, then a longer time period is needed. The correlation lengths of precipitation occurrence and rain rate are sometimes shorter than those of hydrometeor occurrence (giving more independent samples) but of course rain generally occurs less frequently than clouds and so the time period over which one wants to aggregate data tends to be about the same. Of course, this simple estimate makes a number of assumptions, including that the statistics are stationary over time. This is not entirely true as there is a significant seasonal cycle in most regions, and we leave examination of diurnal, seasonal and interannual variability as a topic for future research.

[41] Another important aspect of the CloudSat sampling is that the CloudSat transects occur along specific azimuths (that is, only along the orbital ground track). Ground-based radar observations have long shown that while precipitation correlation functions are nearly isotropic on a scale of 10 km, on larger scales midlatitude precipitating systems sometimes (perhaps often) have elliptical correlation functions [*Zawadzki*, 1973; *Heymsfield*, 1976; *Lee et al.*, 2009]. This suggests that the correlation parameters may depend on the CloudSat azimuth. CloudSat flies in a polar orbit with ascending node (daytime) overpasses having an azimuth that is about 10° counterclockwise from north (except at high latitudes) and descending node (nighttime) overpasses about 10° clockwise over most of the Earth. We therefore expect that the results presented in sections 4 and 5 (composited using all overpasses) are representative of hydrometeor spatial structure in a north–south direction.

[42] The results shown in sections 4 and 5 also represent a composite over all events. Such composites tend to be dominated by the most frequently occurring type of events. *Masunaga et al.* [2005] used Tropical Rainfall Measurement Mission precipitation radar measurements to derive a rain correlation scale length, which is conceptually similar to the precipitation occurrence correlation length examined here. These authors found that deep stratiform and deep convective systems (based on their classification via radar echo top and infrared brightness temperatures) had correlation lengths of

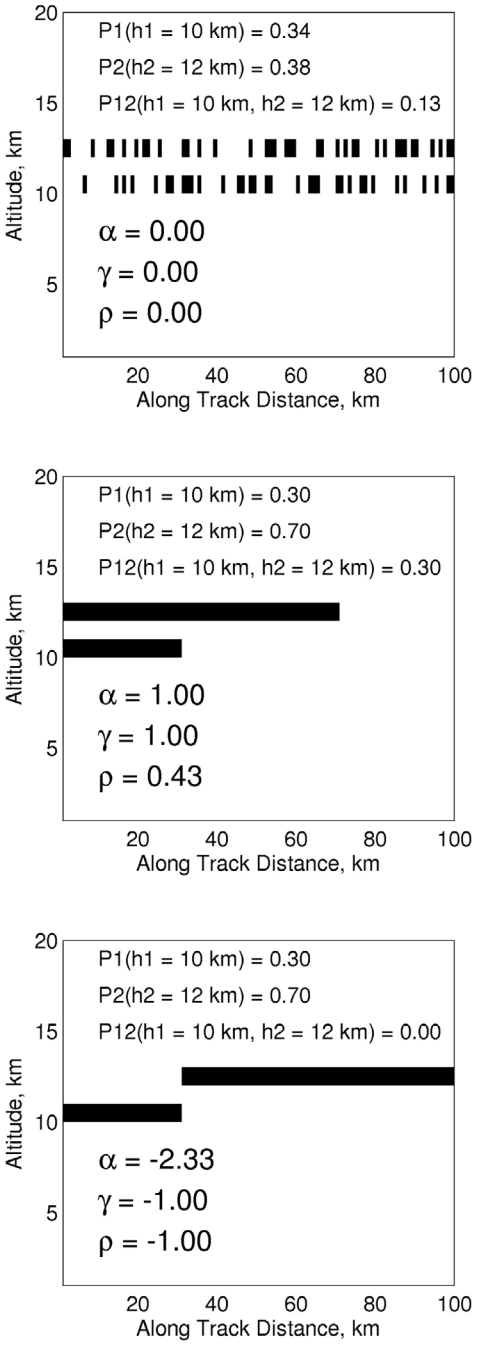

Figure A1. Example of alpha, gamma, and rho (linear) correlation coefficients for cases with (top) random cloud overlap, (middle) maximum cloud overlap, and (bottom) minimum cloud overlap.

126 and 156 km, respectively, while shallow convection and cumulus congestus had correlation lengths of only 8–19 km. Because the number of shallow convective events is much larger than the number of large/deep convective events, our statistics (in tropical regions) are largely dominated by the structure of shallow convective events and hence we find short correlation lengths. In the longer term, it is our intent to generate these spatial statistics while aggregating observations (and simulations) by atmospheric state following the approach used by *Marchand et al.* [2009b]. This approach will allow us to examine spatial structure for different types of events, for example, warm fronts versus cold fronts or isolated convection versus monsoon convection.

Appendix A: Discussion of Correlation Coefficients and the Outer Scale

[43] In principle, one can apply the linear (or Pearson) correlation coefficient (equation (1); section 2) to binary data (such as hydrometeor occurrence). If one substitutes occurrence data for reflectivity data in the definition (that is, let $\langle Z1\rangle = P1 = P(h1)$, $\langle Z2\rangle = P2 = P(h2)$, $\langle Z1Z2\rangle = P_{12}(h1, h2, \Delta x)$), one obtains

$$\rho = \frac{P_{12}(h1, h2, \Delta x) - P(h1)P(h2)}{\sqrt{(P(h1) - P(h1)^2)(P(h2) - P(h2)^2)}}.$$

The numerator of the gamma correlation coefficient (equation (2); section 2) is identical to that for ρ (above). However, the denominators differ. The reason for this change is that the above linear correlation coefficient does not vary between 1 and -1 (except in the limiting case of $P(h1) = P(h2)$), but occupies a smaller range. An example is shown in Figure A1. In Figure A1 the black areas indicate regions occupied by clouds and we examine the simple case of cloud overlap in two layers (that is, $\Delta x = 0$ and $h1$ differs from $h2$). In the case of randomly overlapped clouds (Figure A1, top) the value of P_{12} equals the product $P(h1)P(h2)$, the value expected for uncorrelated data, so both ρ and gamma yield a value of zero. When however the two cloud layers are maximally overlapped (Figure A1, middle) ρ yields a value less than 1. This limitation of the linear correlation coefficient has long been recognized in the statistics community where one sometimes finds an alternative definition for the correlation of binary data defined as

$$\alpha(h1, h2, \Delta x) = \frac{P_{12}(h1, h2, \Delta x) - \text{Uncorrelated Value}}{\text{Max Value Possible} - \text{Uncorrelated Value}} = \frac{P_{12}(h1, h2, \Delta x) - P(h1)P(h2)}{\min(P(h1), P(h2)) - P(h1)P(h2)},$$

where Max Value Possible means the maximum value that P_{12} could possibly have, which is of course simply the smaller of the two values P_1 and P_2 (since there cannot be a detection at both points 1 and 2 more often than there is a detection at P_1 or P_2). Correlation defined in this way is sometimes called phi-over-phi max correlation. Here we call this alpha correlation, because this definition is a more generalized form of the alpha overlap parameter that has been used in a variety of cloud (hydrometeor) overlap studies [e.g., *Mace and Benson-Troth*, 2002; *Hogan and Illingworth*, 2000; *Naud et al.*, 2008]. Alpha-correlation works well in the case of positive correlation, when the value of P_{12} is greater than the product $P(h1)P(h2)$, however this coefficient can yield values that are less than -1, for smaller values of P_{12} as depicted in Figure A1 (bottom). We have therefore extended the concept of phi-over-phi max correlation to include phi-over-phi min for negative correlations, which we

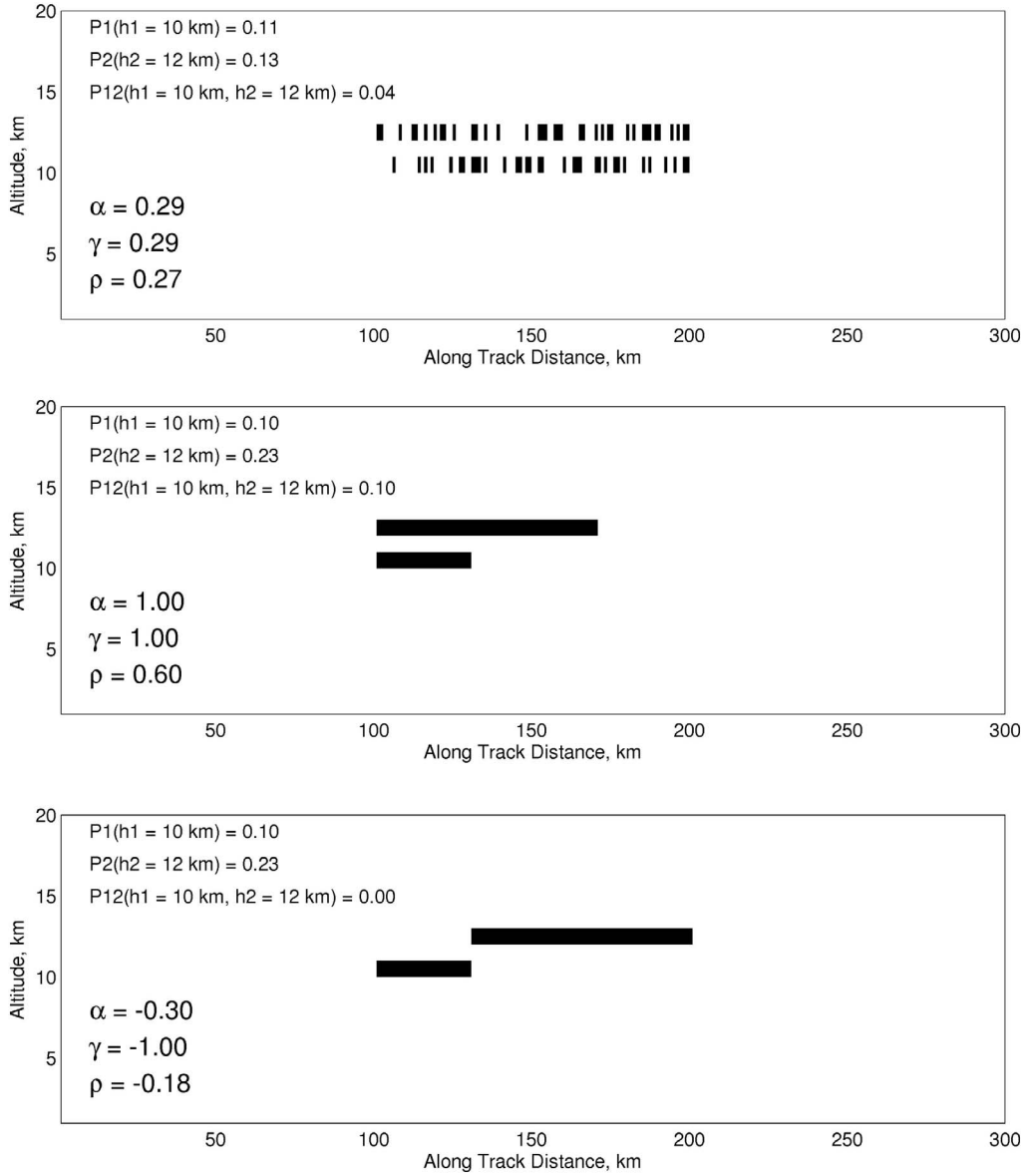

Figure A2. The same as Figure A1, but assuming neighboring regions devoid of cloud.

are calling gamma correlation. That is,If $P_{12}(h1, h2, \Delta x) \geq P(h1)P(h2)$,

$$\gamma(h1, h2, \Delta x) = \frac{P_{12}(h1, h2, \Delta x) - \text{Uncorrelated Value}}{\text{Max Value Possible} - \text{Uncorrelated Value}};$$

otherwise,

$$\gamma(h1, h2, \Delta x) = -\frac{P_{12}(h1, h2, \Delta x) - \text{Uncorrelated Value}}{\text{Min Value Possible} - \text{Uncorrelated Value}}.$$

Min Value Possible is the smallest value that P_{12} can possible have, which is 0 if $P_1 + P_2$ is less than 1 and $P_1 + P_2 - 1$ if $P_1 + P_2$ is greater than 1. In practice, the difference between alpha and gamma is not very important to the results presented in this study because we rarely find negative correlations in our statistics. In fact, on the basis of previous studies of cloud overlap, we initially expected to find more regions with negative correlations than we did. Our incorrect expectation was a result of not fully appreciating the role of conditional sampling in previous cloud overlap studies. That is to say, that in these previous studies the authors calculate the overlap between cloud layers only when clouds are present in both layers. Thus there is an explicit outer scale over which the existence of a layer (and cloud or hydrometeor fraction for each layer) is determined. As noted by Mace and Benson-Troth, the size of the outer scale matters to their retrieved value and comparisons between model output and observations should be careful to match the outer scale. This is demonstrated in Figure A2 (top), which shows the same cloud field examined in Figure A1 (top), except that we have surrounded the cloud field by neighboring regions devoid of cloud. As a result of this change, the correlation coefficient (by any of the definitions we examined) takes on a moderate positive value. This occurs because while on a small (inner) scale the cloudy elements are randomly positioned with respect to each other, on the larger (outer scale) the cloud

layers occur in the same part of the domain and their occurrence is therefore correlated. *Oreopoulos and Norris* [2011] have demonstrated this affect for data collected at the ARM SGP site.

[44] In this study we calculate P_{12} from the CloudSat 2B-GeoProf data simply by counting how often the CloudSat hydrometeor mask indicates a hydrometeor is detected at both points (specified by $h1$, $h2$ and Δx) using every pair of radar profiles (separated by Δx) over a given region of interest and dividing this by a count of the number of times it could have occurred. That is, we construct our counts using all of the data (from many orbits) before applying the equations in section 2, so extensive areas devoid of hydrometeors are not removed from the statistics and there is no conditional sampling. In effect our outer scale is commensurate with the length of the data set (equivalent to many thousands of km). The results of this study show the correlation length of hydrometeor occurrence is as large as a few hundred kilometers (in midlatitudes) and so overlap statistics calculated using layers defined on scales less than this may change significantly with the outer length scale.

Appendix B: Power Spectrum

[45] In this paper, we describe hydrometeor spatial structure using simple correlation functions. An alternative approach would be to concentrate on the (spatial) frequency structure via an examination of the power spectra. The power spectrum can be obtained from the correlation function (and vice versa) by means of the Fourier Transform. An analysis of the power spectrum typically involves displaying the spectrum on a log-log plot and fitting a line (where possible) to the results in order to determine the slope of the line or equivalently fitting the spectrum to a power law function of the form $S(k) = ck^d$, where d would define the slope on the log-log plot (and is sometimes called the scaling constant), c is a constant, S is the spectrum and k the (spatial) frequency.

[46] One can show that the Fourier Transform of a power law is itself a power law, specifically,

$$F\left\{a|x|^b\right\} = \int_{-\infty}^{\infty} a|x|^b e^{-jkx} dx = 2a\Gamma(b+1)\sin(b\pi/2)k^{-(1+b)} = ck^d,$$

where F denotes the Fourier Transform, Γ is the gamma function, b is greater than -1, and $j = \sqrt{-1}$. The scaling constant is therefore given by,

$$d = -(1+b).$$

So, a value of $b = 0.5$ for the correlation function is equivalent to a $d = -1.5$ scaling constant for the power spectrum.

[47] In sections 4 and 5, the horizontal correlation function was fit to an equation of the form $C(x) = \exp[-(|x|/a)^b]$. (We have added the absolute value to reflect the symmetry of the correlation function about zero, which as stated in section 4 is guaranteed to be true by the assumption of horizontal invariance.) One can expand this exponential function in a power series as

$$\exp\left[-\left(\frac{|x|}{a}\right)^b\right] = 1 - \left(\frac{|x|}{a}\right)^b + \frac{1}{2!}\left(\frac{|x|}{a}\right)^{2b} - \frac{1}{3!}\left(\frac{|x|}{a}\right)^{3b} + \ldots.$$

The power spectrum for the exponential can therefore be written as a sum of power laws and as such would not have a single scaling constant. However, for $x \ll (2^{1/b})a$ the exponential can be approximated as,

$$\exp\left[-\left(\frac{|x|}{a}\right)^b\right] \approx 1 - \left(\frac{|x|}{a}\right)^b.$$

Thus at small distances the exponential function is essentially a power law and the power spectrum has a scaling factor of $-(1+b)$ for spatial frequencies $k \gg 1/(2^{1/b}a)$.

[48] An important caveat here is that we fit this functional form to the CloudSat observations over the range of x = 0–200 km, which is generally larger than the correlation length. In practice the exponential form is not a perfect fit and we find that if we restrict the fitting process to a much smaller range (e.g., $x < a/10$), we frequently obtain a somewhat larger value for b, especially for the horizontal reflectivity correlation. An examination of the power spectra (not shown) indicates there is generally a scale break (below the correlation length scale) where the spectrum behaves as $\sim k^{-3}$ to $k^{-1.5}$ depending on location, while at scales larger than the correlation length the spectrum goes as $\sim k^{-1.5}$ to $k^{-1.2}$. We noted in section 1 that *Fliflet and Manheimer* [2006] used a high-power scanning millimeter-wavelength (94 GHz, W band) radar to study the spatial correlation of reflectivity from cirrus clouds and that they found the reflectivity autocorrelation function decreased with scale as $r^{2/3}$ (that is, b = 2/3 and equivalent to a wave number spectrum of $k^{-5/3}$) for scale lengths ranging from 30 m to 10 km. Restricting the range from 1 to 10 km, we too find $b = \sim 2/3$ for midlatitude clouds in the upper troposphere (and $b \sim 0.5$ when fitting over the range 1–200 km).

[49] In effect, the shape parameter reported here is a compromise between the value that best represents the rapid decay in the correlation at small scales (much less than correlation length) and the slow decay in the correlation at larger scales. In some sense it is remarkable that the simple exponential form is able to capture the transition as well as it does. Nonetheless, these results suggest that there may be some benefit in investigating a 3 (or more) parameter fit and undertaking a more thorough analysis of the correlation function (and spectrum) at small spatial scales.

[50] **Acknowledgments.** The author would like to express his thanks to Phil Partain, Don Reinke, and the entire staff of the CloudSat Data Processing Center at CIRA, Colorado State University, for gathering, processing, and providing CloudSat data. Also, thanks go to John Haynes, Tristan L'Ecuyer, and Wesley Berg for their efforts in developing the CloudSat precipitation product and answering many questions. This research was supported by the NASA JPL CloudSat project (under contract 1303809).

References

Astin, I., L. Di Girolamo, and H. M. Van de Poll (2001), Baysian confidence intervals for true fractional coverage from finite transect measurements: Implications for cloud studies from space, *J. Geophys. Res.*, *106*, 17,303–17,310, doi:10.1029/2001JD900168.

Austin, P. M., and R. A. Houze (1972), Analysis of the structure of precipitation patterns in New England, *J. Appl. Meteorol.*, *11*, 926–935, doi:10.1175/1520-0450(1972)011<0926:AOTSOP>2.0.CO;2.

Barnes, E. A., and D. L. Hartmann (2011), The global distribution of eddy-length scales, *J. Clim.*, doi:10.1175/JCLI-D-11-00331.1, in press.

Berg, W., T. L'Ecuyer, and J. M. Haynes (2010), The distribution of rainfall over oceans from spaceborne radars, *J. Appl. Meteorol. Climatol.*, *49*(3), 535–543, doi:10.1175/2009JAMC2330.1.

Bodas-Salcedo, A., M. J. Webb, M. E. Brooks, M. A. Ringer, K. D. Williams, S. F. Milton, and D. R. Wilson (2008), Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities, *J. Geophys. Res.*, *113*, D00A13, doi:10.1029/2007JD009620.

Bodas-Salcedo, A., et al. (2011), COSP: Satellite simulation software for model assessment, *Bull. Am. Meteorol. Soc.*, *92*, 1023–1043, doi:10.1175/2011BAMS2856.1.

Ciach, G. J., and W. F. Krajewski (2006), Analysis and modeling of spatial correlation structure in small-scale rainfall in central Oklahoma, *Adv. Water Resour.*, *29*, 1450–1463, doi:10.1016/j.advwatres.2005.11.003.

Crane, R. K. (1990), Space-time structure of rain rate fields, *J. Geophys. Res.*, *95*, 2011–2020, doi:10.1029/JD095iD03p02011.

Ellis, T. D., T. L'Ecuyer, J. M. Haynes, and G. L. Stephens (2009), How often does it rain over the global oceans? The perspective from CloudSat, *Geophys. Res. Lett.*, *36*, L03815, doi:10.1029/2008GL036728.

Fliflet, A. W., and W. M. Manheimer (2006), Measurement of correlation functions and power spectra in clouds using the NRL WARLOC radar, *IEEE Trans. Geosci. Remote Sens.*, *44*(11), 3247–3261, doi:10.1109/TGRS.2006.879114.

Gebremichael, M., and W. F. Krajewski (2004), Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets, *J. Appl. Meteorol.*, *43*, 1180–1199, doi:10.1175/1520-0450(2004)043<1180:AOTSCO>2.0.CO;2.

Gupta, V. K., and E. C. Waymire (1993), A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–267, doi:10.1175/1520-0450(1993)032<0251:ASAOMR>2.0.CO;2.

Haynes, J. M., T. S. L'Ecuyer, G. L. Stephens, S. D. Miller, C. Mitrescu, N. B. Wood, and S. Tanelli (2009), Rainfall retrieval over the ocean with spaceborne W-band radar, *J. Geophys. Res.*, *114*, D00A22, doi:10.1029/2008JD009973.

Heymsfield, G. M. (1976), Statistical objective analysis of dual-Doppler radar data from a tornadic storm, *J. Appl. Meteorol.*, *15*, 59–68, doi:10.1175/1520-0450(1976)015<0059:SOAODD>2.0.CO;2.

Hogan, R. J., and A. J. Illingworth (2000), Deriving cloud overlap statistics from radar, *Q. J. R. Meteorol. Soc.*, *126*, 2903–2909, doi:10.1002/qj.49712656914.

Key, J. (1994), The area coverage of geophysical fields as a function of sensor field-of-view, *Remote Sens. Environ.*, *48*, 339–346, doi:10.1016/0034-4257(94)90008-6.

Lee, C. K., G. W. Lee, I. Zawadzki, and K. Kim (2009), A preliminary analysis of spatial variability of raindrop size distributions during stratiform rain events, *J. Appl. Meteorol. Climatol.*, *48*, 270–283, doi:10.1175/2008JAMC1877.1.

Lohou, F., A. Druilhet, and B. Campistron (1998), Spatial and temporal characteristics of horizontal rolls and cells in the atmospheric boundary layer based on radar and in situ observations, *Boundary Layer Meteorol.*, *89*(3), 407–444, doi:10.1023/A:1001791408470.

Lothon, M., D. H. Lenschow, D. Leon, and G. Vali (2005), Turbulence measurements in marine stratocumulus with airborne Doppler radar, *Q. J. R. Meteorol. Soc.*, *131*, 2063–2080, doi:10.1256/qj.04.131.

Lovejoy, S., et al. (2008), The remarkable wide range scaling of TRMM precipitation, *Atmos. Res.*, *90*, 10–32, doi:10.1016/j.atmosres.2008.02.016.

Mace, G. G., and S. Benson-Troth (2002), Cloud-layer overlap characteristics derived from long-term cloud radar data, *J. Clim.*, *15*, 2505–2515, doi:10.1175/1520-0442(2002)015<2505:CLOCDF>2.0.CO;2.

Mandapaka, P. V., P. Lewandowski, W. E. Eichinger, and W. F. Krajewski (2009), Multiscaling analysis of high resolution space-time lidar-rainfall, *Nonlinear Processes Geophys.*, *16*, 579–586.

Marchand, R., G. G. Mace, T. Ackerman, and G. Stephens (2008), Hydrometeor detection using Cloudsat: An Earth-orbiting 94-GHz cloud radar, *J. Atmos. Oceanic Technol.*, *25*, 519–533.

Marchand, R., J. Haynes, G. G. Mace, T. Ackerman, and G. Stephens (2009a), A comparison of simulated cloud radar output from the multiscale modeling framework global climate model with CloudSat cloud radar observations, *J. Geophys. Res.*, *114*, D00A20, doi:10.1029/2008JD009790.

Marchand, R. T., N. Beagley, and T. Ackerman (2009b), Evaluation of hydrometeor occurrence profiles in the multiscale modeling framework climate model using atmospheric classification, *J. Clim.*, *22*, 4557–4573.

Masunaga, H., T. S. L'Ecuyer, and C. D. Kummerow (2005), Variability in the characteristics of precipitation systems in the tropical Pacific. Part I: Spatial structure, *J. Clim.*, *18*, 823–840, doi:10.1175/JCLI-3304.1.

Moreau, E. J. T., and E. L. Bouar (2009), Rainfall spatial variability observed by X-band weather radar and its implication for the accuracy of rainfall estimates, *Adv. Water Resour.*, *32*, 1011–1019, doi:10.1016/j.advwatres.2008.11.007.

Naud, C. M., A. Del Genio, G. G. Mace, S. Benson, E. E. Clothiaux, and P. Kollias (2008), Impact of dynamics and atmospheric state on cloud vertical overlap, *J. Clim.*, *21*, 1758–1770, doi:10.1175/2007JCLI1828.1.

North, G. R., and S. Nakamoto (1989), Formalism for comparing rain estimation designs, *J. Atmos. Oceanic Technol.*, *6*, 985–992, doi:10.1175/1520-0426(1989)006<0985:FFCRED>2.0.CO;2.

Oreopoulos, L., and P. M. Norris (2011), An analysis of cloud overlap at a midlatitude atmospheric observation facility, *Atmos. Chem. Phys.*, *11*, 5557–5567, doi:10.5194/acp-11-5557-2011.

Overeem, A., I. Holleman, and A. Buishand (2009), Derivation of a 10-year radar-based climatology of rainfall, *J. Appl. Meteorol. Climatol.*, *48*, 1448–1463, doi:10.1175/2009JAMC1954.1.

Pollard, B. D., S. Khanna, S. J. Frasier, J. C. Wyngaard, D. W. Thomson, and R. E. McIntosh (2000), Local structure of the convective boundary layer from a volume-imaging radar, *J. Atmos. Sci.*, *57*, 2281–2296, doi:10.1175/1520-0469(2000)057<2281:LSOTCB>2.0.CO;2.

Rickenbach, T. M., and S. A. Rutledge (1998), Convection in TOGA COARE: Horizontal scale, morphology, and rainfall production, *J. Atmos. Sci.*, *55*, 2715–2729, doi:10.1175/1520-0469(1998)055<2715:CITCHS>2.0.CO;2.

Sassen, K., S. Matrosov, and J. Campbell (2007), CloudSat spaceborne 94 GHz radar bright bands in the melting layer: An attenuation-driven upside-down lidar analog, *Geophys. Res. Lett.*, *34*, L16818, doi:10.1029/2007GL030291.

Settle, J. J., and H. M. Van de Poll (2007), On the Bayesian estimation of cloud fraction from lidar transects, *J. Geophys. Res.*, *112*, D09211, doi:10.1029/2006JD007251.

Shonk, J. K. P., R. J. Hogan, J. M. Edwards, and G. G. Mace (2010), Effect of improving representation of horizontal and vertical cloud structure on the Earth's global radiation budget. Part I: Review and parameterization, *Q. J. R. Meteorol. Soc.*, *136*, 1191–1204, doi:10.1002/qj.647.

Smythe, G. R., and D. S. Zrnic (1983), Correlation analysis of Doppler radar data and retrieval of the horizontal wind, *J. Clim. Appl. Meteorol.*, *22*, 297–311, doi:10.1175/1520-0450(1983)022<0297:CAODRD>2.0.CO;2.

Steiner, M., T. L. Bell, Y. Zhang, and E. F. Wood (2003), Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar dataset, *J. Clim.*, *16*(22), 3759–3778, doi:10.1175/1520-0442(2003)016<3759:COTMFE>2.0.CO;2.

Stephens, G. L., et al. (2008), CloudSat mission: Performance and early science after the first year of operation, *J. Geophys. Res.*, *113*, D00A18, doi:10.1029/2008JD009982.

Tanelli, S., S. L. Durden, E. Im, K. S. Pak, D. G. Reinke, P. Partain, J. M. Haynes, and R. T. Marchand (2008), CloudSat's cloud profiling radar after two years in orbit: Performance, calibration, and processing, *IEEE Geosci. Remote Sens.*, *46*(11), 3560–3573, doi:10.1109/TGRS.2008.2002030.

Zawadzki, I. I. (1973), Statistical properties of precipitation patterns, *J. Appl. Meteorol.*, *12*, 459–472, doi:10.1175/1520-0450(1973)012<0459:SPOPP>2.0.CO;2.

R. Marchand, Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA 98195-5672, USA. (rojmarch@u.washington.edu)