



HAL
open science

Assessment of uncertainty in cloud radiative effects and heating rates through retrieval algorithm differences: Analysis using 3 years of ARM data at Darwin, Australia

Jennifer M. Comstock, Alain Protat, Sally A. Mcfarlane, Julien Delanoë, Min Deng

► To cite this version:

Jennifer M. Comstock, Alain Protat, Sally A. Mcfarlane, Julien Delanoë, Min Deng. Assessment of uncertainty in cloud radiative effects and heating rates through retrieval algorithm differences: Analysis using 3 years of ARM data at Darwin, Australia. *Journal of Geophysical Research: Atmospheres*, 2013, 118 (10), pp.4549-4571. 10.1002/jgrd.50404 . hal-00812908

HAL Id: hal-00812908

<https://hal.science/hal-00812908>

Submitted on 28 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Assessment of uncertainty in cloud radiative effects and heating rates through retrieval algorithm differences: Analysis using 3 years of ARM data at Darwin, Australia

Jennifer M. Comstock,¹ Alain Protat,² Sally A. McFarlane,³ Julien Delanoë,⁴ and Min Deng⁵

Received 1 October 2012; revised 20 March 2013; accepted 8 April 2013; published 22 May 2013.

[1] Ground-based radar and lidar observations obtained at the Department of Energy's Atmospheric Radiation Measurement Program's Tropical Western Pacific site located in Darwin, Australia, are used to retrieve ice cloud properties in anvil and cirrus clouds. Cloud microphysical properties derived from four different retrieval algorithms (two radar-lidar and two radar-only algorithms) are compared by examining mean profiles and probability density functions of effective radius (R_e), ice water content (IWC), visible extinction coefficient, ice number concentration, ice crystal fall speed, and vertical air velocity. Retrieval algorithm uncertainty is quantified using radiative flux closure exercises. The effect of uncertainty in retrieved quantities on the cloud radiative effect and radiative heating rates is presented. Our analysis shows that IWC compares well among algorithms, but R_e shows significant discrepancies, which are attributed primarily to assumptions of particle shape. Uncertainty in R_e and IWC translates into sometimes large differences in cloud shortwave radiative effect (CRE) though the majority of cases have a CRE difference of roughly 10 W m^{-2} on average. These differences, which we believe are primarily driven by the uncertainty in R_e , can cause up to 2 K/d difference in the radiative heating rates between algorithms.

Citation: Comstock, J. M., A. Protat, S. A. McFarlane, J. Delanoë, and M. Deng (2013), Assessment of uncertainty in cloud radiative effects and heating rates through retrieval algorithm differences: Analysis using 3 years of ARM data at Darwin, Australia, *J. Geophys. Res. Atmos.*, 118, 4549–4571, doi:10.1002/jgrd.50404.

1. Introduction

[2] A number of algorithms are available to retrieve the microphysical properties of clouds from remote sensing measurements. These properties are then used to determine cloud radiative effects and heating rate profiles and to evaluate model simulations. Extensive research has been performed to improve and evaluate these algorithms through direct algorithm comparisons [e.g., Turner *et al.*, 2007; Comstock *et al.*, 2007], comparisons with aircraft in situ measurements [e.g., Heymsfield *et al.*, 2008], and in some instances through surface or top-of-atmosphere (TOA) closure studies [Mather *et al.*, 2007]. However, less has been done to quantify the uncertainties in cloud proper-

ties and understand the impact of these uncertainties on our knowledge of the cloud radiative effects and heating rate profiles, particularly for ice clouds. Previous work by Vogelmann and Ackerman [1995] suggests that an error of $\pm 12\%$ in extinction optical depth τ would allow the net surface fluxes to be computed within $\pm 5\%$ (holding all other scattering calculations constant). Over a decade later, we ask the question: to what uncertainty can we estimate the radiative effect of clouds, and is it good enough to evaluate the radiative budget from large scale?

[3] Retrieval algorithm classes include those that use active remote sensors [Intrieri *et al.*, 1993; Wang and Sassen, 2002; Donovan and van Lammeren, 2001; Matrosov *et al.*, 2002; Mace *et al.*, 2002; etc.], algorithms that use passive remote sensors (Turner, 2005) and those that use some combination of both [Matrosov *et al.*, 1994; Mace *et al.*, 1998; Delanoë and Hogan, 2008]. Here we focus on algorithms that use active remote sensors (e.g., lidar and/or radar) to retrieve vertical profiles of cloud particle size and water content in ice-only clouds. The rationale for focusing on retrievals using radar and lidar is that they are the only instruments capable of characterizing the vertical distribution of cloud properties. Also, cloud radars and lidars are complementary, allowing for a very large percentage of clouds covering a broad range of physical and optical depths, to be characterized. Ground-based cloud radars will penetrate most cloud layers but will miss a portion of optically

¹Pacific Northwest National Laboratory, Richland, Washington, USA.

²Centre for Australian Weather and Climate Research, Melbourne, Victoria, Australia.

³Pacific Northwest National Laboratory, Richland, Washington, USA.

⁴Laboratoire Atmosphères Milieux et Observations Spatiales (LATMOS), Institute Pierre Simon Laplace, Université de Versailles St Quentin, Guyancourt, France.

⁵Department of Atmospheric Science, University of Wyoming, Laramie, Wyoming, USA.

Corresponding author: J. M. Comstock, Pacific Northwest National Laboratory, Richland, WA 99352, USA. (jennifer.comstock@pnl.gov)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50404

thin cirrus clouds [Comstock et al., 2002; Mace et al., 2006; Protat et al., 2006]. Conversely ground-based lidars will detect these thin cirrus clouds, but the backscatter signals will often be extinguished by supercooled liquid cloud layers in mixed-phase clouds or clouds of optical depth larger than 2 to 3 [Sassen and Cho, 1992; Protat et al., 2006]. There is an overlap in radar and lidar optical depth retrieval range for which radar-lidar observations can be used simultaneously to derive accurate retrievals of cloud properties [Donovan and van Lammeren, 2001, Wang and Sassen, 2002; Okamoto et al., 2003; Tinel et al., 2005; Delanoë and Hogan, 2008]. Within this class, we will examine algorithms that are applied to clouds detected only by radar, only by lidar, and by both radar and lidar. Details of these algorithms will be discussed in section 2.

[4] We apply several active remote sensing algorithms to ice clouds observed at the Tropical Western Pacific (TWP) site located in Darwin, Australia, which is funded by the U. S. Department of Energy Atmospheric Radiation Measurement (ARM) program [Ackerman and Stokes, 2003]. This ARM TWP site provides a unique opportunity to examine algorithm differences under diverse cloud scenes. Depending on the time of year, the Darwin site observes high optically thin cirrus, thick anvil, precipitating stratiform clouds, deep convection, boundary layer cumulus, and midlevel stratiform clouds. We will focus our comparison during periods when lower level clouds and precipitation do not obscure cirrus and anvil clouds, focusing on the optical depth range between 0.01 and 50.

[5] Our goal is to examine the uncertainty in retrieved ice cloud properties from ground-based remote sensors and how this uncertainty impacts our estimates of the cloud radiative forcing and heating rates in tropical ice clouds. As retrieved cloud properties (from ground and space-based instruments) become more extensively used for model evaluation studies, understanding uncertainties in these cloud properties is critical. Our approach is to first compare the microphysical properties (ice crystal size (R_e) and ice water content (IWC)) derived from four different algorithms. Second, we compute and examine radiative fluxes and heating rates using each set of cloud properties as input to a radiative transfer model. Third, we statistically compare the surface and top-of-atmosphere (TOA) radiative fluxes to the measured ones. Through this analysis, we ultimately quantify the current uncertainty in our ground-based estimates of cloud radiative effects and heating in the atmosphere using the ARM data.

2. Cloud Properties Retrieval Algorithms

[6] We examine retrieval algorithms that require millimeter wave radar (94 or 35 GHz frequency) and/or visible wavelength lidar (i.e., 532 nm wavelength) as input measurements. The measured quantities used as input are the radar reflectivity (Z_e), Doppler velocity (V_d), spectral width (σ_d), and lidar backscatter coefficient. Ancillary measurements such as temperature (T) and pressure from radiosonde profiles are also used. The microphysical and dynamical retrieved quantities discussed in this paper are visible extinction coefficient (α), effective radius (R_e), ice water content (IWC), reflectivity weighted ice terminal fall velocity (V_f ; which will be referred to for convenience as “terminal fall speed” throughout this paper), and the vertical

air velocity (W). The effective radius definition used is that of Stephens et al. [1990]:

$$R_s = \frac{3IWC}{2\rho_i\alpha} \quad (1)$$

where ρ_i is the density of solid ice.

2.1. Combined Lidar-Radar Algorithms

[7] Two of the algorithms use combined radar and lidar measurements to retrieve cloud properties based on whether radar, lidar, or both have the sensitivity to detect clouds. The first algorithm we describe is the variational synergistic scheme (Varcloud) developed by Delanoë and Hogan [2008]. The version of Varcloud used in this study does not include infrared radiance measurements. An iterative process is used to adjust the state vector containing α , normalized concentration (N' defined below) and particulate extinction-to-backscatter ratio S_p to minimize the difference between the forward modeled reflectivity and backscatter and the observed quantities [Rodgers, 2000]. After minimization of the cost function, the optimal state vector and look-up tables are used to derive the other cloud properties (IWC, R_e , total number concentration [N_t], and V_f). As the retrieval technique uses a variational framework, it includes a rigorous treatment of measurements and forward model errors. The forward model contains an assumed microphysical model describing the shape of the normalized particle size distribution (a two-parameter modified gamma distribution), following Delanoë et al. [2005], as well as relationships between particle mass, cross-sectional area, and maximum size. The particle size distribution is therefore prescribed by the combination of the number concentration parameter N_0^* and the mean volume weighted diameter D_m . The assumed shape is oblate spheroid with an aspect ratio of 0.6. [Hogan et al., 2012].

[8] The ice particle mass is assumed to follow the Brown and Francis [1995] mass-maximum diameter relationship derived from aircraft data in ice aggregates. The corresponding cross-sectional area-maximum size relationship is taken from Francis et al. [1998]. These two relationships are only used for crystals with sizes larger than 300 μm . Below 300 μm , the area-density-diameter relationships have been taken from Mitchell [1996] for hexagonal columns. The lidar forward model accounts for multiple scattering and attenuation using the model of Hogan [2006]. It is also important to note that the S_p is retrieved but is assumed to be constant within the layer. The radar forward model is built using the T -matrix approach and assuming an aspect ratio of 0.6 [Hogan et al., 2012].

[9] For each cloudy gate, we retrieve α and normalized concentration N' ($N' = N_0^*/\alpha^{0.6}$). Additionally, the S_p is retrieved when the gate is observed by the lidar. When both instruments observe the same pixel and the number of observations is sufficient to retrieve two moments of the PSD, then the other moments of the PSD can be retrieved (N_t , IWC, R_e). However, when only one instrument is available, a priori information, such as temperature (typically from radiosonde or model simulations), is used to constrain the normalized concentration. Single instrument retrievals are therefore similar to IWC- Z - T relationships for the radar [Liu and Illingworth, 2000; Hogan et al.,

2006; Protat et al., 2007] and IWC- α - T relationship for the lidar [Heymsfield et al., 2005].

[10] One interesting feature of this algorithm is that cloud properties are retrieved seamlessly between regions of the cloud detected by both radar and lidar and regions detected by just one of these two instruments. This is done by propagating the radar-lidar information within the closest region observed by both instruments to the region (typically several hundred meters) to where only one instrument can detect the cloud. This is possible using the a priori error covariance matrix for spreading of normalized number concentration information in height. In a very simplistic way, we give less weight to the $N'(T)$ relationship a priori in these areas by minimizing the difference between the observed and simulated parameters, and the errors in each instrument and a priori data [Delanoë and Hogan, 2008].

[11] The second radar-lidar algorithm uses a conditional approach, which selects algorithms based on available measurements. Since this algorithm combines methods from several published studies, we label this technique the “Combined Retrieval” (CombRet). When both radar and lidar detect cloud, we apply the algorithm of Wang and Sassen [2002], which requires the radar Z_e and lidar α as inputs. The lidar α is derived from the lidar backscatter profile using the method described in Comstock and Sassen [2001]. The particulate extinction-to-backscatter ratio is estimated independently for each lidar profile by varying S_p until the average above cloud backscatter coefficient minus the molecular (Rayleigh) backscatter is approximately zero. This is equivalent to a so-called “Beers law” approach and requires that the lidar signal penetrate through the cloud top. When the lidar signal is fully attenuated, we apply the radar-only approach (described below). A multiple scattering correction is applied to the lidar equation assuming a value of 0.8 [Platt, 1973; Comstock and Sassen, 2001]. When only lidar detects cloud, we apply the method adopted by Heymsfield et al. [2005] for use with satellite-based lidar. This approach essentially utilizes the relationship between the ratio IWC/ α and temperature, which are well correlated. The IWC is solved for using the measured radiosonde T and lidar α . The “generalized particle effective diameter” D_{ge} is then computed using equation (3.3) in Fu [1996], which is also used in the Wang and Sassen [2002] lidar-radar algorithm, supplying some consistency between methods. This generalized effective diameter D_{ge} [Wang and Sassen, 2002] can then be related to the effective radius using $R_e = (3(3)^{1/2})D_{ge}/8$ [Fu, 1996, equation 3.12]. For radar-only clouds, we have developed a set of tuned regressions [Matrosov, 1999; Hogan et al., 2006] relating Z_e , IWC, and T using the microphysical quantities derived from the Wang and Sassen [2002] lidar-radar algorithm in regions of lidar-radar overlap. The tuned regressions (between Z_e , IWC, and T derived from the lidar-radar algorithm) are derived per cloud scene (i.e., over a single day of observations), but also compiled using the entire TWP Darwin data set. If there are not sufficient data points to derive the regressions on a given day, regressions derived from the entire data set are used instead. Essentially the entire retrieval is run twice, first to create the “climatology” regressions first, then a second time to apply the regressions as appropriate. Analogous to the lidar-only method, we then use the IWC/ α relationship

to derive D_{ge} . Similar to the Wang and Sassen [2002] radar-lidar technique, the hexagonal column mass-maximum diameter and area-maximum diameter relationships are assumed. Note that this assumption is fully consistent with the assumptions in the radiative transfer model used in section 5.

2.2. Radar Doppler Moments Algorithms

[12] A number of algorithms exist that utilize radar Doppler moments (e.g., Z_e , Doppler velocity (V_d), and/or spectral width (σ_d)) to derive cloud microphysical properties [Matrosov et al., 2002; Mace et al., 2002; Deng and Mace, 2006; Delanoë et al., 2007]. These algorithms are applied only to radar measurements and so can be compared to both the empirical radar-only methods and the radar-lidar methods.

[13] We use two algorithms based on the Doppler moments method. First, we use the approach presented in Delanoë et al. [2007] and Plana-Fattori et al. [2010], which we refer to as the RadOn (radar-only) method. The assumption of a normalized particle size distribution shape [Delanoë et al., 2005] is the same as in Varcloud. The unique feature of this method is that the particle mass-maximum diameter and cross-sectional area-maximum diameter relationships can vary from one cloud to another, unlike the other methods. By considering a range of possible mass-diameter relationships (assuming $m(D) = aD^b$ and varying a and b over a reasonable range) and five possible area-diameter relationships [Mitchell, 1996], statistical relationships between the reflectivity weighted terminal fall velocity V_f and the equivalent reflectivity Z_e , and relationships relating these two radar parameters to the microphysical properties, are computed at 35 GHz using an extensive airborne in situ microphysical data set [Delanoë et al., 2005]. The Mitchell [1996] area-diameter relationships include various ice crystal habits (solid spheres, hexagonal plates, hexagonal columns, nonspherical aggregates, and assemblages of planar polycrystals in cirrus clouds), and the in situ data set includes ice clouds from both midlatitude and tropical data sets. For each cloud, we retain the mass-maximum diameter and area-maximum diameter combination that produces the V_f - Z_e relationship closest in the least squares sense to the V_f - Z_e relationship derived directly from the radar observations. Once these statistical relationships are retrieved the microphysical properties are directly derived from precalculated look-up tables [Plana-Fattori et al., 2010, Delanoë et al., 2007]. The method does not always provide a solution, which occurs essentially when the radar-derived V_f - Z_e relationship does not match any V_f - Z_e relationship in the microphysical database. Our experience is that it happens when updrafts associated with large reflectivities are too large to be filtered out by the method, thereby producing negative exponents of the V_f - Z_e relationship (which is not physical). Negative exponents are retrieved fairly frequently in small clouds because this statistical approach requires at least 1 h of continuous cloud measurements to work properly.

[14] The second Doppler moments algorithm uses the method described in Deng and Mace [2006] and is referred to as Rad3mom (Radar 3 moments). The algorithm indeed utilizes the first three moments of the Doppler spectrum, Z_e , V_d , and σ_d , to retrieve the ice crystal size distribution, from which the microphysical properties are computed. Assuming a first-order gamma distribution for the particle

size distribution and an exponential function for the turbulence probability density function, the set of equations describing the Doppler spectrum moments are inverted using optimal estimation theory to derive the particle size distribution, the mean vertical velocity of the air in the sample volume, and objectively derived retrieval errors [Deng and Mace, 2006]. To avoid an ill-conditioned problem, the turbulence distribution width is considered as a parameter in the algorithm and is predetermined from the Doppler spectrum width and radar reflectivity based on the observation that the spread of the particle size distribution in the velocity domain dominates the Doppler spectrum width measurement for most cirrus. The mass-maximum diameter and area-maximum diameter relationships of Yang [2000] for idealized ice crystals are used to derive the terminal fall velocity of individual ice crystals of maximum diameter D using drag theory described in Mitchell [1996]. Therefore, the microphysical model is consistent in terms of mass-maximum diameter, area-maximum diameter, and particle fall velocity-maximum diameter relationships. However, the particle habit is predetermined as hexagonal columns in this application.

[15] The ideal strategy of this paper at this point would be to highlight similarities and differences between methods and hypothesize how these differences impact agreement between microphysical properties derived from these methods. However, algorithm assumptions vary widely (shape of the particle size distribution, statistical relationships between crystal mass, size, and fall speed) so it is fair to say that these four methods, despite a few similarities, are strikingly different from one another, although they represent the state-of-the-art in ice cloud microphysics retrieval techniques. The remainder of the paper will attempt to quantify how different the microphysical and radiative properties are given these large differences between algorithms.

3. Data Sets and Methodology

[16] We use ground-based measurements obtained at the U.S. Department of Energy ARM site located in Darwin, Australia, to compile common input files so that each algorithm participant uses identical input data on a common height-time grid. The primary instruments are the ARM Millimeter Cloud Radar (MMCR), which operates at 35 GHz, and the Micropulse Lidar (MPL), which operates at 532 nm. Our input files include the CloudNet-processed MMCR data set [Illingworth *et al.*, 2007], the ARM-produced Merged Sounding Value Added Product for thermodynamic profiles [Trojan, 2010], and MPL-normalized backscatter profiles. Details about the CloudNet and ARM processed data sets can be obtained at the websites <http://www.cloud-net.org> and <http://www.arm.gov>, respectively. Each measurement was averaged 2 min temporally and 300 m vertically. We also applied water vapor and cloud water attenuation corrections [Liebe, 1985] to the radar reflectivity measurements, as well as overlap, range, and deadtime corrections to the MPL backscatter profiles. From these individual inputs, a common cloud mask was produced using both radar and lidar cloud detections [Wang and Sassen, 2001]. Points where both radar and lidar masks detect cloud are identified as radar-lidar points.

[17] Once cloudy points are identified in each profile, we assign a phase classification using the Shupe [2007] approach. This algorithm uses radar Doppler moments (Z_e , V_d , and σ_d), lidar backscatter, microwave radiometer, and temperature profiles to identify clouds (ice, liquid, or mixed), drizzle, rain, or aerosol. We do not use lidar depolarization ratio in our case because polarization-sensitive lidar was not available for the entire time period. Since this phase classifier algorithm was developed for Arctic clouds, we made some adjustments to the cutoff parameters, though it is notable that the algorithm works well for tropical clouds with minimal changes. One additional condition that we have added to this algorithm is that we do not allow water to exist at temperatures colder than -12°C because it is reported by Stith *et al.* [2002] that liquid water is rarely observed in tropical stratiform clouds observed by aircraft. Since we are interested in only ice clouds without underlying precipitation or dense boundary layer clouds (for flux closure experiments), we are confident that this phase algorithm works sufficiently. More attention may be required to identify tropical mixed-phase clouds or precipitation with accuracy. It is worth emphasizing that building this common data set significantly reduced uncertainties associated with resolution, cloud detection, and definitions of cloud that can complicate the interpretation of the intercomparison results.

4. Microphysical Properties

[18] The common radar and lidar ground-based observation data set obtained at Darwin is compiled for July 2005 to December 2009. Participants in this intercomparison applied their retrieval algorithm using this common data set. Here we examine the retrieved IWC, R_e , and α from all algorithms. In addition, some algorithms also derive total number concentration (N_t), terminal fall velocity (V_f), and vertical air velocity (W). For the analysis, the entire time period is subdivided into subsamples in order to compare similar retrieval types:

[19] 1. Radar-lidar subsample (called rali subsample) includes all data points when both radar and lidar instruments detect cloud. For these points, both radar-lidar algorithms (Varcloud and CombRet) and both radar-only Doppler moments algorithms (RadOn and Rad3mom) are applied. The purpose of applying the radar Doppler moments algorithms to the rali subsample is to examine how the two algorithm classes compare. The expectation here is that the radar-lidar methods should be more accurate than the Doppler moments methods, owing to a better extinction retrieval using the lidar measurements.

[20] 2. Radar subsample includes all regions where only radar measurements are available for the retrieval of the cloud properties. For these data points, the Doppler radar methods are applied, as well as the radar-only components of the Varcloud and CombRet algorithms. The latter algorithms tend to be more empirically based than the Doppler moments methods. This subsample allows for a more direct comparison of these two classes of radar-only methods.

[21] 3. Lidar subsample includes all regions where only lidar measurements are available, allowing for comparisons of the lidar-only part of the radar-lidar methods.

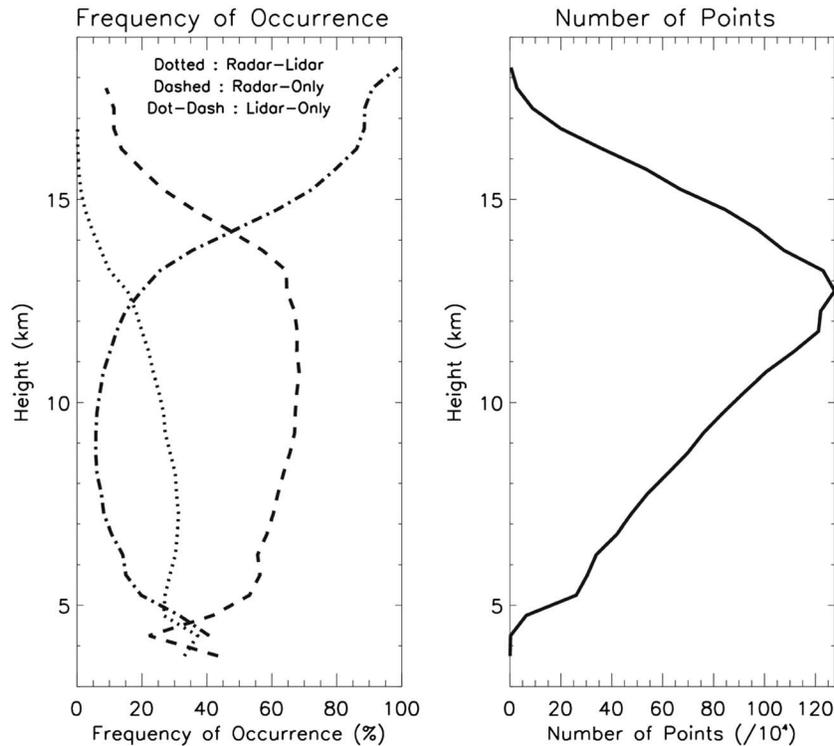


Figure 1. (left) Frequency of occurrence (%) of each retrieval combination as a percent of the (right) total number of retrieval points at each altitude: radar-lidar (dotted), radar only (dashed), and lidar only (dash-dotted).

[22] The relative frequency of radar, lidar, and rali subsamples is given as a vertical profile in Figure 1. Overall, the important features of this vertical profile are that the radar subsample largely exceeds the lidar and rali subsamples up to 13 km, while the lidar subsample dominates above 14 km. It is noteworthy that the rali subsample, for which microphysical retrievals are presumably most accurate, represents at best 20–30% of the total sample (from 5 to 12 km height, see Figure 1). This important result highlights the fact that for ground-based remote-sensing measurements, the radiative effect of clouds is actually estimated most of the time from a single-instrument retrieval (lidar only for thin cirrus above 14 km height and radar only below 14 km height). This result may change in different climatic regimes where the tropopause height is lower and clouds are not as optically thin on average (i.e., midlatitudes) and when using satellite-borne radar-lidar instruments due to different viewing geometry.

[23] Using these different subsamples, we examine the microphysical properties derived using the various retrieval methods to highlight the main discrepancies between algorithms. All differences between the microphysical retrievals will then be evaluated in terms of the differences with surface fluxes in the next section. The underlying question we address here is do the microphysics differences produced by these state-of-the-art retrieval methods correspond to *large* differences in terms of cloud radiative effect? The hope here is that the methods are able to provide statistical estimates within 5 W m^{-2} (shortwave) to provide a reference for model evaluation and space-borne radiative budget estimates.

4.1. Rali Subsample

[24] Figure 2 shows the probability density functions (PDFs) and height normalized PDFs (HPDFs) [Protat *et al.*, 2009] of IWC, α , and R_e retrieved by all algorithms for the rali subsample. Table 1 tabulates the first three moments of the PDFs displayed in Figure 2 (first row; mean, variance, and skewness) as well as the same comparisons for three selected heights of the HPDFs of Figure 2 (7, 11, and 15 km). Looking at the composite PDFs (Figure 2, first row), the radar-lidar methods produce very similar distributions for IWC and α , but very different PDFs of R_e (see also values in Table 1). CombRet is characterized by a much larger variance in the R_e distribution than the other methods (variance of 852 for the total PDF as compared with 167 for Varcloud; Table 1). RadOn is skewed toward smaller sizes, especially at 11 km (Table 1) where the mean value is half that of CombRet). Varcloud and Rad3mom have very similar distributions of R_e (as judged by the PDF moments). Given the larger positive skewness of the Varcloud distribution when compared with CombRet, the mean values obtained from the two radar-lidar methods are $10 \mu\text{m}$ apart (Table 1), although the distribution peak for the two methods is the same value of about $40 \mu\text{m}$. All algorithms exhibit a decrease in R_e with altitude but RadOn clearly has the most altitude-dependent distribution and produces much smaller R_e ($<10 \mu\text{m}$) at the highest altitudes. Rad3mom produces microphysical properties very similar to Varcloud (Table 1), although the Rad3mom distributions are systematically slightly broader (more frequent occurrence of smaller values for IWC and α , see first row in Figure 2 and Table 1). One possible reason for this

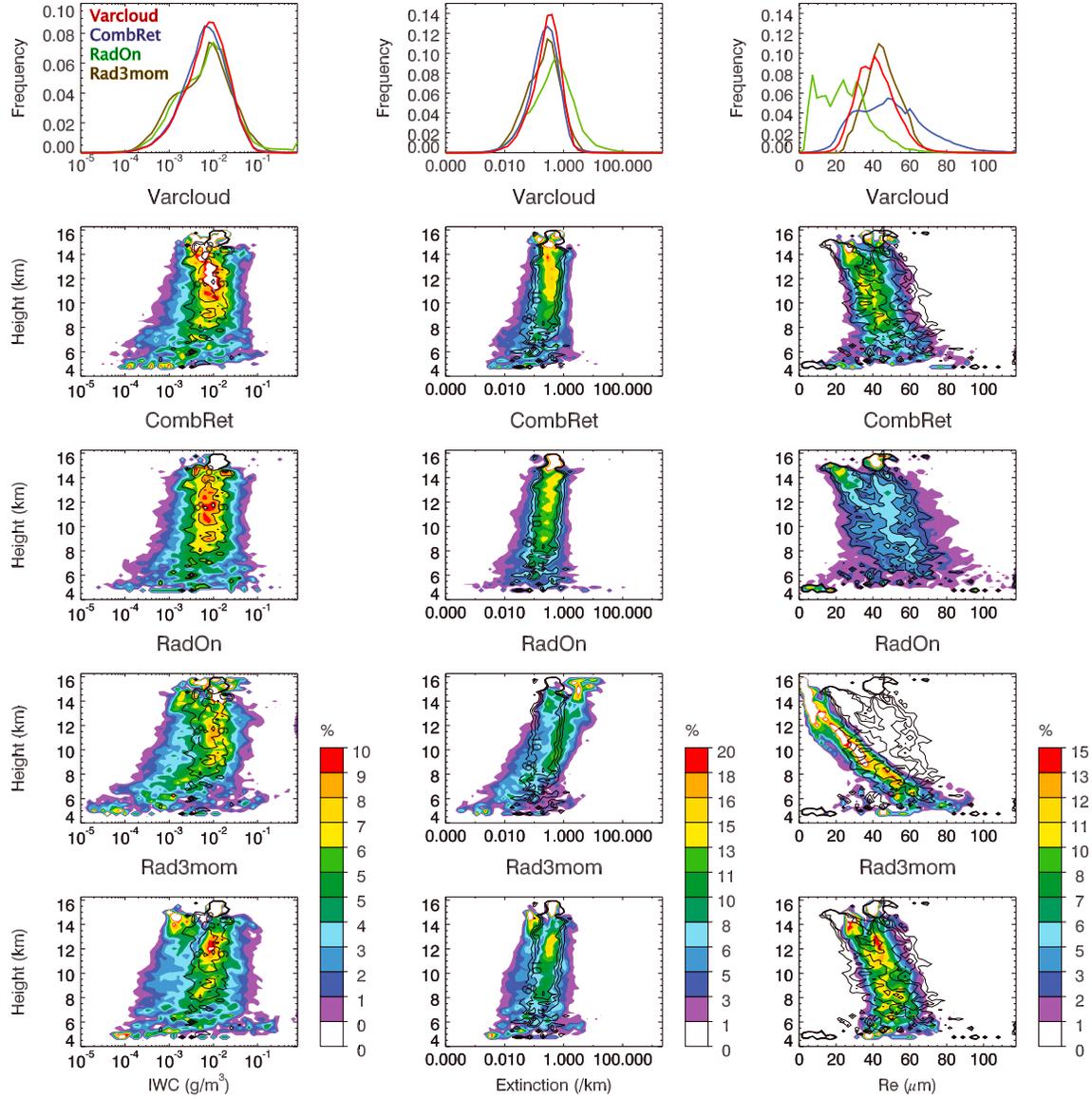


Figure 2. Microphysical properties comparison in the radar-lidar cloud detection. (first row) PDFs of (left column) IWC, (middle column) α , and (right column) R_e , respectively, from Varcloud (red), CombRet (blue), RadOn (green), and Rad3mom (yellow). Color contours of height normalized PDFs (HPDFs) of (second row) Varcloud, (third row) CombRet, (fourth row) RadOn, and (fifth row) Rad3mom, respectively. Overplotted in thin black lines are the corresponding results from CombRet for reference.

Table 1. Moments of the PDFs of $\log(\text{IWC})$, $\log(\alpha)$, and R_e for Each Retrieval Technique^a

Radar-Lidar		$\log(\text{IWC})$				$\log(\alpha)$				R_e			
		Varcloud	CombRet	RadOn	Rad3mom	Varcloud	CombRet	RadOn	Rad3mom	Varcloud	CombRet	RadOn	Rad3mom
Total PDF	Mean	-2.12	-2.15	-2.14	-2.19	-0.52	-0.61	-0.31	-0.63	43	53	26	47
	Variance	0.26	0.27	0.42	0.38	0.26	0.29	0.55	0.33	167	852	256	137
	Skewness	-0.5	-0.4	+0.1	-0.2	-1.0	-0.6	-0.1	-0.3	+7.2	+5.3	+4.8	+8.8
PDF at 7 km	Mean	-2.41	-2.34	-2.71	-2.36	-0.94	-0.88	-1.34	-0.88	75	81	81	77
	Variance	0.77	0.79	1.12	0.85	1.55	0.98	4.37	0.90	5359	5321	4958	5049
	Skewness	-0.2	-0.3	+0.4	+0.1	-0.8	-0.7	-0.1	-0.5	+2.8	+2.6	+2.9	+3.0
PDF at 11 km	Mean	-2.09	-2.16	-2.07	-2.16	-0.50	-0.65	-0.27	-0.62	46	59	28	51
	Variance	0.25	0.25	0.77	0.39	0.25	0.31	0.45	0.46	619	1742	655	1053
	Skewness	-0.7	-0.4	-0.6	-0.3	-1.5	-1.0	-0.5	-1.3	+9.0	+4.8	+10.0	+7.5
PDF at 15 km	Mean	-2.13	-2.08	-2.17	-2.44	-0.57	-0.40	+0.10	-1.15	74	76	54	76
	Variance	0.72	0.49	0.48	3.82	1.77	0.88	0.94	6.45	8476	8764	10078	8333
	Skewness	-1.0	-1.1	-0.9	-0.1	-1.7	-2.1	-3.1	-0.3	+2.1	+2.0	+2.1	+2.2

^aTotal PDF (top row) and for three selected heights: 7, 11, and 15 km. This table is for the radar-lidar subsample.

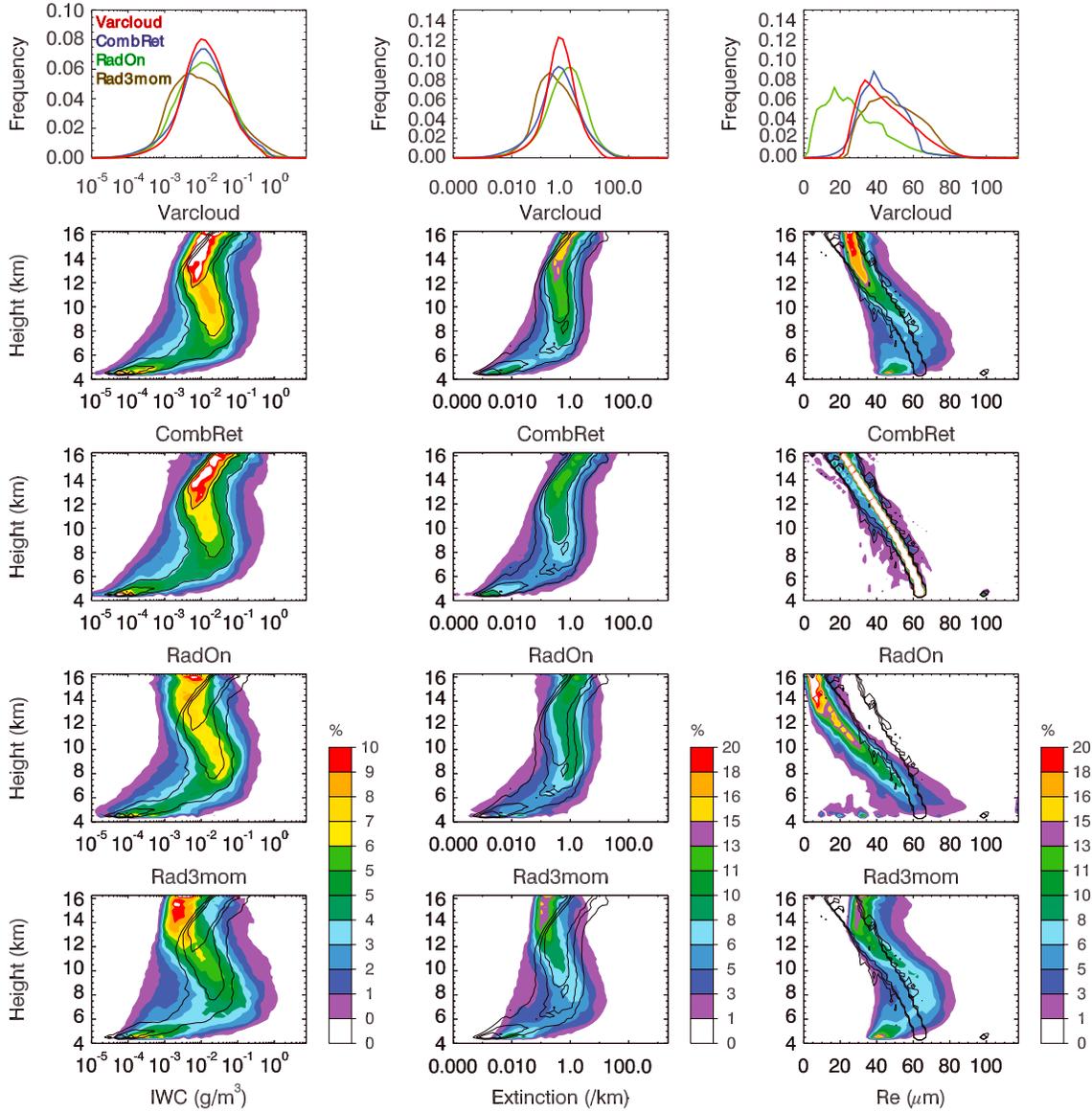


Figure 3. The same as Figure 2 except for results in the radar-only cloud detection category.

general agreement is that Rad3mom and Varcloud use the same particle habit assumption for small particles (hexagonal columns).

[25] In contrast to R_e , the IWC and α HPDFs are similar among the algorithms with the exception that RadOn α increases more with altitude as expected from the previously described smaller R_e . For IWC HPDFs, variance in the distributions with altitude is similar, though RadOn has a more pronounced decrease in IWC below 8 km and a larger variance up to 11 km (Table 1). One interesting feature in the HPDFs is that several algorithms exhibit a sharp decrease in IWC, α , and R_e at ~ 15 km, which could distinguish the microphysical properties of anvil versus in situ generated cirrus. The differences exhibited by RadOn are in part due to the implicitly retrieved (and not assumed) particle habit produced by the algorithm (through a variable mass-maximum diameter and five possible cross-sectional area-maximum diameter relationships). The implications

of such large differences in terms of the radiative effect of clouds will be analyzed in section 5.

4.2. Radar Subsample

[26] The radar subsample, as shown in Figure 1, dominates the total sample at most heights. Recall that the radar-only part of the radar-lidar methods and the Doppler radar methods is actually compared here. Presumably, the use of an additional constraint (V_d) in the Doppler radar methods should be an advantage over the radar-lidar methods that apply more empirical approaches to retrieve cloud properties when only radar detects cloud. It must be noted that the Doppler radar methods (and Varcloud through the retrieval of the particle size distribution parameters) also provide additional information that can be compared: W , V_f , and N_r .

[27] Despite having different approaches to deriving microphysical properties when only radar detects cloud,

Table 2. Same as Table 1 but for the Radar-Only Subsample

Radar-Only		log(IWC)				log(α)				R_e			
		Varcloud	CombRet	RadOn	Rad3mom	Varcloud	CombRet	RadOn	Rad3mom	Varcloud	CombRet	RadOn	Rad3mom
Total PDF	Mean	-1.84	-1.86	-1.88	-1.90	-0.28	-0.31	-0.10	-0.37	46	45	29	51
	Variance	0.41	0.54	0.53	0.64	0.33	0.58	0.53	0.50	197	425	297	202
	Skewness	-0.2	-0.2	-0.2	+0.3	-0.4	-0.3	-0.5	+0.2	+1.0	+8.2	+1.2	+0.9
PDF at 7 km	Mean	-2.16	-2.29	-2.21	-1.83	-0.72	-0.88	-0.78	-0.39	60	61	57	61
	Variance	0.67	0.78	0.88	0.99	0.54	0.79	0.84	0.83	201	155	322	192
	Skewness	-0.3	-0.2	-0.2	-0.1	-0.4	-0.2	-0.3	-0.1	+2.6	+5.0	+1.5	+5.3
PDF at 11 km	Mean	-1.78	-1.81	-1.70	-1.81	-0.23	-0.29	+0.03	-0.31	48	57	29	54
	Variance	0.34	0.46	0.44	0.64	0.25	0.46	0.40	0.48	113	2526	109	122
	Skewness	+0.1	+0.1	0.0	0.0	0.0	+0.1	-0.2	+0.2	+2.9	+4.8	+12.2	+2.3
PDF at 15 km	Mean	-1.71	-1.60	-2.10	-2.16	+0.01	+0.15	+0.08	-0.51	32	42	12	38
	Variance	0.28	0.44	0.37	0.37	0.21	0.46	0.38	0.26	78	3533	92	131
	Skewness	+0.6	-0.2	+0.6	+0.9	+0.3	0.0	+0.3	+0.8	+14.8	+4.2	+14.8	+6.9

Varcloud and CombRet produce very similar statistics for all microphysical quantities (Figure 3 and Table 2), including R_e , which is quite different from the results for the radi cloud detections with these two retrievals (Figure 2, right column). The most notable difference is that the variance of the PDF produced by CombRet is systematically larger than that of Varcloud (Table 2), especially for the R_e distribution. This general good agreement between the radar-lidar methods occurs because when only radar data are available, the two retrievals default to similar algorithms using radar reflectivity and temperature as inputs to the IWC and α retrieval.

[28] Comparisons of IWC produced by the four methods show that the PDFs produced by Varcloud, CombRet, and RadOn are similar, but corresponding HPDFs reveal different vertical distributions. The three methods agree fairly well up to 13 km, but do not agree at all above that height, where both radar-lidar methods produce an increasing IWC with height and both Doppler radar methods produce a constant IWC with height (Table 2 and Figure 3, left column). For the radar-lidar methods, this increase is caused by an increase in Z_e with height above 13 km (not shown). Therefore, a retrieval method relying on radar reflectivity only must produce an increase in IWC and α by construction, while the Z_e - V_d retrieval techniques rely on the characteristics of two or three Doppler moments. However, this result should be kept in perspective since the number of radar detections largely decreases above 13 km (Figure 1). Discrepancies between lidar and radar detections have been noted previously [Comstock *et al.*, 2002] and can have significant impacts on derived TOA IR fluxes [Borg *et al.*, 2011], which we will explore further in section 5. The IWC PDF produced by Rad3mom is characterized by a larger variance and peaks at smaller IWC than the three other methods. The HPDFs indicate that larger IWC values are produced by Rad3mom below 10 km height (see larger mean value and variance at 7 km, Table 2), while lower values are produced predominantly above 10 km height when compared with the other methods.

[29] PDFs of α show that RadOn produces larger extinction than the radar-lidar methods, primarily between 8 and 13 km (see mean values at 11 km, Table 2), for this radar subsample, while Rad3mom overall produces smaller α than the radar-lidar methods (Table 2 and Figure 3, middle column), which results from a compensation between larger values below 10 km and much smaller values above 11 km (Table 2). RadOn also has larger extinction values than the other

methods above ~ 10 km in the radi subsample (Figure 2 and Table 1). The resulting comparison of R_e (which is proportional to IWC/α , see (1)) shows that owing to compensating effects of IWC and α , the R_e produced by Rad3mom is slightly larger than Varcloud and CombRet at all heights above 6 km, with maximum differences around 8 km height and above 14 km height (Figure 3, right column). The larger extinctions produced by RadOn translate into much smaller R_e compared to the other methods above 8 km (largest differences are found above 12 km height, see also Table 2). The HPDFs show that the R_e distribution from CombRet is much narrower than the other methods due primarily to the temperature dependence of the R_e retrieval used by CombRet for “radar only” clouds. The fact that the variance is actually much larger than other methods is due to the fact that the distribution is far from normal; hence, the variance calculation is more difficult to interpret in that case. Differences in R_e between the two Doppler radar methods are very large, though the source of the discrepancies varies at different heights. Below 12 km, larger R_e values produced by Rad3mom are predominantly due to IWCs larger than those from RadOn and the other methods. Above 12 km height, smaller R_e values in RadOn are due to larger extinctions produced by RadOn (in agreement with CombRet) and IWCs similar to Rad3mom (but much smaller than Varcloud and CombRet). An assessment of the correct R_e values will be performed using the surface shortwave flux comparisons, since clouds with smaller particles should reflect more incoming shortwave radiation than those with larger particles.

[30] Additional dynamical and microphysical properties are compared for the radar subsample for three of the algorithms (Figure 4 and Table 3). PDFs of N_t produced by Varcloud and RadOn are in reasonably good agreement in terms of mean values (less than 5% difference overall, Table 3); however, the Varcloud HPDF increases more distinctly with height and the RadOn N_t distribution is much broader and much less skewed at all heights (Table 3 and Figure 4, left column). This apparent agreement in mean values of N_t between Varcloud and RadOn is somewhat surprising, but is likely caused by offsetting uncertainties that are revealed in the HPDFs (Varcloud N_t is larger above 12 km, whereas RadOn is slightly larger below 12 km). Earlier comparisons between ground-based radar-lidar retrievals of N_t (using Varcloud) and space-borne radar-only retrievals from the CloudSat radar

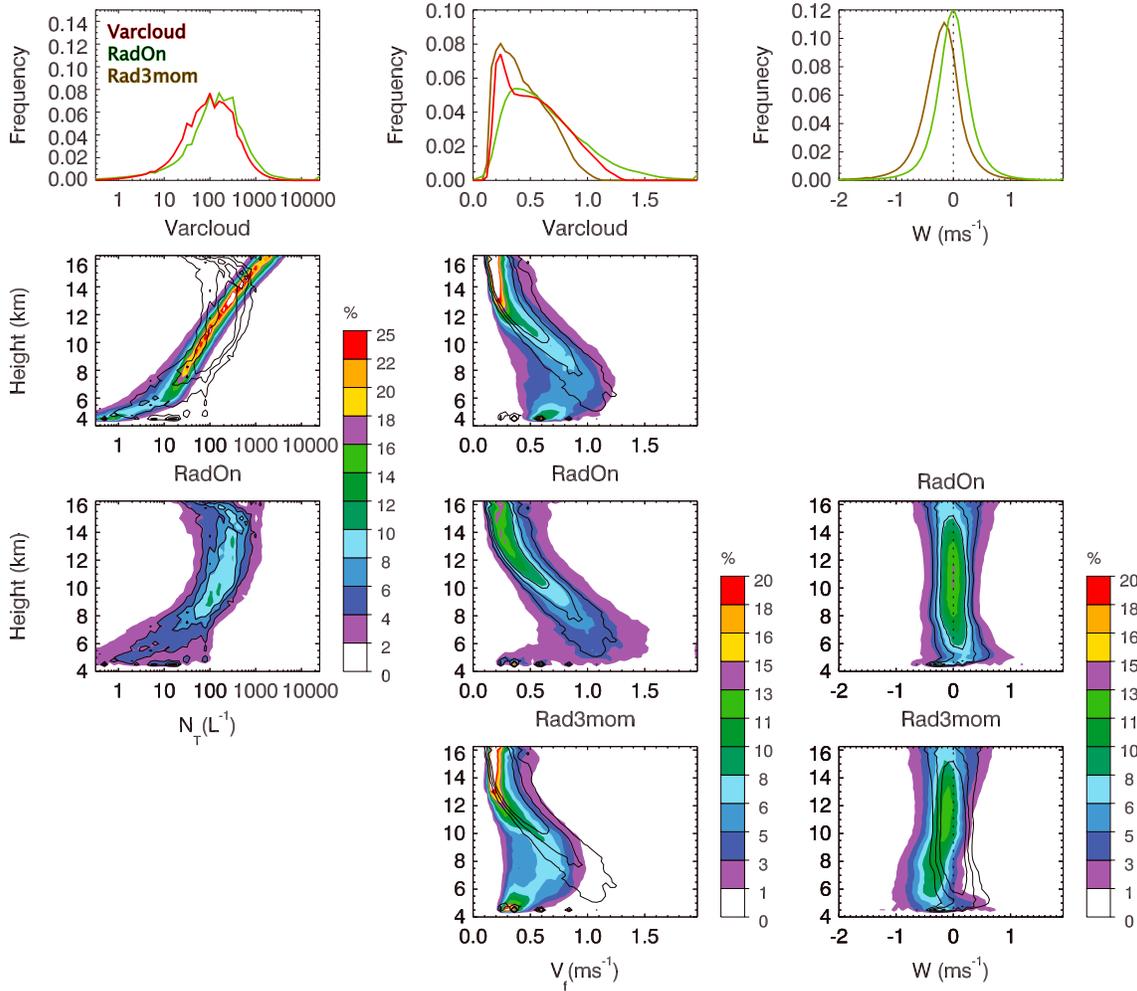


Figure 4. Comparison of total number concentration (N_t), particle fall velocity (V_f), and mean air motion (W) for radar-only cloud detections. (first row) PDFs of (left column) N_t , (middle column) V_f , and (right column) W , respectively, from Varcloud (red), RadOn (green), and Rad3mom (yellow). Color contours of HPDFs of (second row) Varcloud, (third row) RadOn, and (fourth row) Rad3mom, respectively. Overplotted in thin black lines are the corresponding results from RadOn for reference.

[Protat *et al.*, 2010] have shown that reflectivity-only retrievals of N_t could not get the order of magnitude of total concentration correct. This is because the total concentration (which is the zeroth moment of the particle size

distribution (PSD)) is indirectly related to the reflectivity measurements (the sixth moment of the PSD in the Rayleigh scattering regime), which is the main input to the radar-only methods. The differences observed between RadOn and

Table 3. Same as Table 1 but for $\log(N_t)$, V_f , and W

Radar-Only		$\log(N_t)$		V_f			W	
		Varcloud	RadOn	Varcloud	RadOn	Rad3mom	RadOn	Rad3mom
Total PDF	Mean	2.01	2.12	0.56	0.66	0.46	0.02	-0.18
	Variance	0.34	0.51	0.07	0.11	0.04	0.15	0.18
	Skewness	-0.6	-0.9	+0.5	+0.9	+0.6	-0.6	-1.0
PDF at 7 km	Mean	1.08	1.08	0.82	1.06	0.61	0.05	-0.41
	Variance	0.15	0.70	0.05	0.12	0.04	0.24	0.27
	Skewness	-0.8	-0.6	-0.0	+0.2	+0.2	-1.4	-1.5
PDF at 11 km	Mean	1.97	2.31	0.59	0.66	0.50	0.02	-0.14
	Variance	0.05	0.25	0.04	0.05	0.03	0.10	0.12
	Skewness	-1.4	+0.6	+0.5	+1.2	+0.6	-0.9	-1.5
PDF at 15 km	Mean	2.89	2.37	0.29	0.38	0.29	0.06	-0.04
	Variance	0.04	0.29	0.02	0.04	0.02	0.23	0.23
	Skewness	-3.4	-0.3	+2.3	+1.7	+1.7	+0.3	+0.1

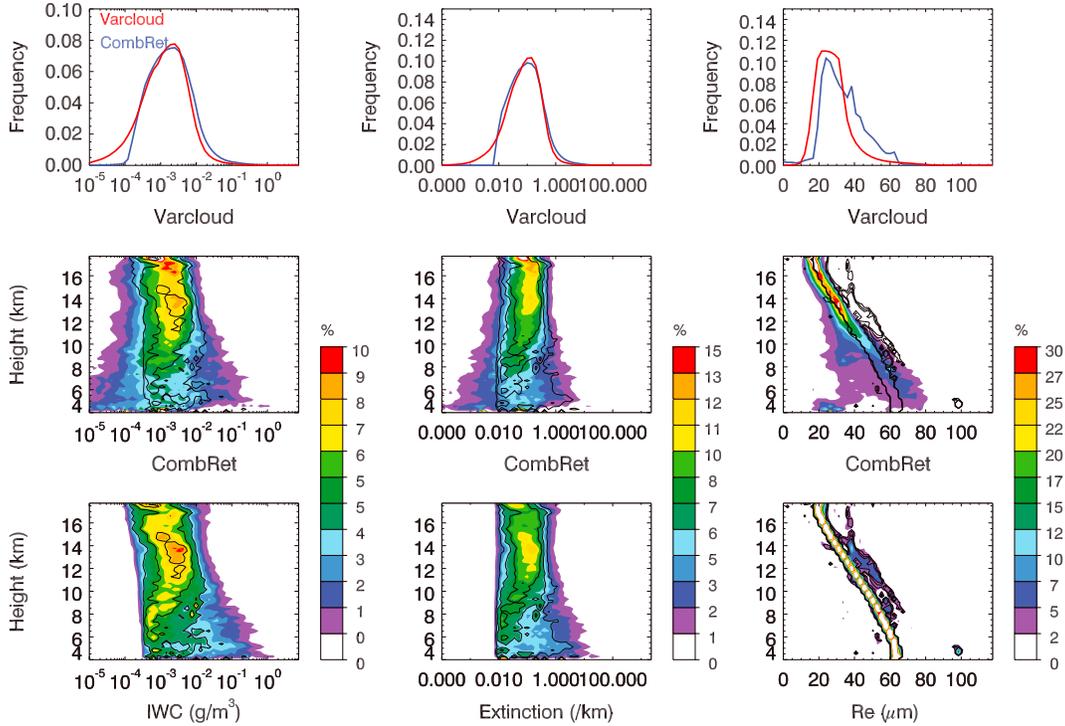


Figure 5. (top row) PDFs and HPDFs of (left column) IWC, (middle column) α , and (right column) R_e , derived by (middle row) Varcloud and (bottom row) CombRet for the lidar sample.

Varcloud are much smaller than the differences reported in Protat *et al.* [2010], at least below 12 km. Even if the two methods share the same assumption about the shape of the PSD, this comparison indicates that the two free parameters of the normalized PSD (the intercept parameter and the mean volume-weighted diameter), which are retrieved using the two methods, are in good agreement overall.

[31] The terminal fall velocity PDF shows that the RadOn method retrieves slightly larger mean values of V_f compared to Rad3mom and Varcloud (Table 3 and Figure 4, middle column), though the latter two algorithms have a sharp peak at ~ 0.25 m s⁻¹. The variance and skewness of the RadOn distribution are also larger than for the two other methods. The HPDFs and associated moments of Table 3 at three selected heights help characterize more clearly the differences in V_f . RadOn produces V_f that are almost twice as large as those retrieved by Rad3mom predominantly in the 5–10 km layer (see also mean values at 7 km in Table 3), while the agreement is better between RadOn and Rad3mom above 10 km height. Terminal fall speeds retrieved using the Varcloud algorithm fall between the two Doppler moments algorithms: Varcloud and Rad3mom agree very well in peak and width of the distributions above 10 km height, and Varcloud produces terminal fall speeds with values intermediate between RadOn and Rad3mom below 10 km height (Figure 4 and Table 3). Given the difference in R_e for the three methods (Figure 3), we can infer that the particle fall speed-maximum diameter relationship retrieved on a case-by-case basis by RadOn and the assumption by Rad3mom of hexagonal columns for all cases produce very different results. In the Doppler moments retrievals, the measured Doppler velocity is split between the vertical air velocity component (W) and

the terminal fall speed (V_f), using different methods (details can be found in Delanoë *et al.* [2007] and Deng and Mace [2006], respectively, for RadOn and Rad3mom). Varcloud uses a statistical fall speed-maximum dimension relationship approach for individual crystals by Mitchell and Heymsfield [2005]. Recent studies using multi-wavelength profiler observations over Darwin [Protat and Williams, 2011] suggest that the V_f - Z_e approach used in RadOn tends to slightly underestimate terminal fall speed in tropical ice clouds, by 5–15 cm s⁻¹ depending on height (their Figure 9). Protat and Williams [2011] also caution against using a single particle habit assumption for all clouds and showed that assuming the hexagonal columns represents relatively well small terminal fall speeds associated with low reflectivities, but will strongly underestimate the larger terminal fall speeds associated with large Z_e typically found in the lower portions of ice clouds [Protat and Williams, 2011] (Figure 5). Our comparison between RadOn and Rad3mom is fully consistent with the findings of Protat and Williams [2011]. The good agreement found between RadOn and Rad3mom above 10 km height is presumably due to the fact that hexagonal column habit assumption is relevant at these heights statistically, while it presumably underestimates terminal fall velocity below 10 km height. It also suggests that the RadOn retrieval of fall speed is reasonable, which was also a conclusion from Protat and Williams [2011].

[32] RadOn and Rad3mom also retrieve vertical air velocity, W (defined as positive upward). Retrieved PDFs by RadOn and Rad3mom are symmetric centered on mean values of +2 and -18 cm s⁻¹, respectively (Figure 4 and Table 3). The other moments of the two PDFs are similar (Table 3). The HPDFs of Figure 4 and the numbers in Table 3 show that

Table 4. Same as Table 1 but for the Lidar-Only Subsample

Lidar-Only		Log(IWC)		Log(α)		R_e	
		Varcloud	CombRet	Varcloud	CombRet	Varcloud	CombRet
Total PDF	Mean	-2.85	-2.66	-1.09	-0.97	29	35
	Variance	0.45	0.37	0.36	0.32	114	271
	Skewness	-0.4	+0.4	-0.6	+0.4	+3.0	+8.3
PDF at 7 km	Mean	-2.84	-2.23	-1.34	-0.79	67	64
	Variance	1.3	0.8	1.21	0.78	3011	777
	Skewness	-0.1	+0.4	-0.1	+0.4	+3.8	+8.2
PDF at 11 km	Mean	-2.73	-2.47	-1.11	-0.93	42	49
	Variance	0.54	0.39	0.48	0.40	539	469
	Skewness	-0.1	+0.3	-0.3	+0.2	+9.9	+11.4
PDF at 15 km	Mean	-2.87	-2.75	-1.06	-1.00	26	32
	Variance	0.36	0.28	0.30	0.27	78	598
	Skewness	-0.5	+0.4	-0.5	+0.2	+26.8	+10.4

RadOn W distributions are actually centered around 0, whereas Rad3mom is centered around a few cm s^{-1} downdraft (negative) except for below ~ 8 km where RadOn becomes more positive ($+ 5 \text{ cm s}^{-1}$) and Rad3mom more negative (mean value of -41 cm s^{-1}). This corresponds to the differences in the V_f between these two retrievals, which have been discussed previously.

4.3. Lidar Subsample

[33] Figure 5 shows the PDFs and HPDFs of IWC, α , and R_e produced by Varcloud and CombRet. The PDF comparisons show that Varcloud has a slightly larger frequency of small IWC and α compared to CombRet, which translates into smaller mean values, larger variances, and slightly negative skewness of the Varcloud distributions at all heights (Table 4). For R_e , the PDF produced by Varcloud is shifted toward slightly smaller values compared with CombRet (mean value of 29 versus 35 μm , Table 4). The HPDFs show that the R_e differences are of similar magnitude at all heights, with R_e produced by Varcloud being systematically 5 μm smaller than those produced by CombRet, with the notable exception of mean values from Varcloud being slightly larger at 7 km height (Table 4). Extinction results for CombRet show a somewhat artificial cutoff in the α PDF and HPDFs, which is likely caused by the forced max/min values for S_p , though a specific cutoff for α is not introduced into the algorithm. Recall from Figure 1 that the majority of lidar-only clouds occurs above 10 km; hence, the agreement above that altitude is somewhat constrained, particularly for α , which is primarily driven by the lidar ratio. PDFs of lidar ratio derived by the two methods exhibit significant differences for rali and lidar subsamples (Figure 6). Varcloud almost always retrieves a value of 33 sr because the a priori value of S_p is the center value, and the algorithm varies around that value. The CombRet algorithm begins the iteration at the largest allowed value of S_p rather than the center value, which results in a wider distribution, centered around 40 sr for the “lidar only” subsample. The range of allowed S_p is 10 to 66 sr. Sakai *et al.* [2003] summarizes the available measurements of S_p in different climate regimes. While smaller values (5–25 sr) have been measured in midlatitude cirrus, larger values (39–79 sr) have been observed in tropical regimes. Theoretical calculations also presented in Sakai *et al.* [2003] suggest that small crystals tend to have

large values and hexagonal crystals tend to have small values. It is interesting that for the rali subsample the PDF of S_p is very broad compared to the “lidar only” sample, which has a peak near 38 sr. This could be indicative of a shift in the type of cirrus detected when radar does not detect the cloud (i.e., optically thin cirrus versus denser anvils). The small values of S_p (< 20 sr) retrieved by CombRet likely indicate that the lidar profile is attenuation limited in some of the rali profiles, since the rali sample tends to have large optical depths than the “lidar only” sample. Despite these differences in the lidar ratio, the retrieved α agrees well, as

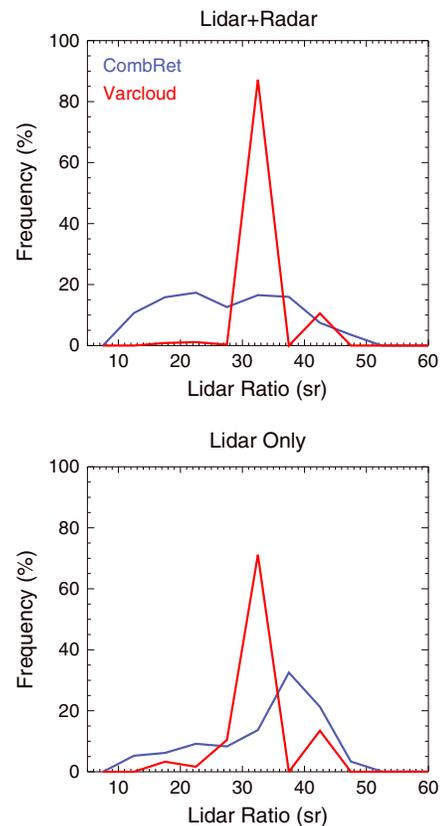


Figure 6. Frequency distributions extinction-to-backscatter ratio (lidar ratio) retrieved by the CombRet and Varcloud algorithms for (top) rali and (bottom) lidar-only subsamples.

Table 5. Surface SW Flux Comparison Statistics Including Number of Observations in Each Subsample^a

Retrieval	Num. Obs.	R^2	<10%	<20%	<50%	Mean	STD_DEV
<i>All Observations</i>							
CombRet	47033	0.93	45.8	63.5	85.2	13.1	37.3
Varcloud	36096	0.93	41.3	59.7	84.1	16.4	37.8
RadOn	23019	0.89	35.1	54.2	81.7	1.45	41.2
Rad3mom	21259	0.92	36.5	55.3	79.9	21.9	40.1
<i>Rali Observations</i>							
CombRet	1779	0.94	29.9	52.3	83.6	14.5	37.7
Varcloud	1779	0.95	34.6	52.8	87.4	14.4	38.4
RadOn	1779	0.93	28.6	47.4	85.9	-4.49	37.3
Rad3mom	1779	0.96	39.7	59.3	82.7	19.2	36.6
<i>Radar Only</i>							
CombRet	15397	0.90	38.7	58.1	83.1	11.1	38.7
Varcloud	15397	0.91	38.2	57.6	84.6	13.8	36.5
RadOn	15397	0.87	34.2	54.1	82.0	-5.0	39.1
Rad3mom	15397	0.91	35.4	54.7	80.1	21.4	38.9
<i>Lidar Only</i>							
CombRet	10516	0.93	40.2	58.4	82.3	16.4	41.3
Varcloud	10516	0.92	38.9	56.1	77.6	25.5	42.0

^a R^2 represents the correlation coefficient between the computed and observed surface SW flux. Also listed are percentage of computed fluxes that fall within 10, 20, and 50% of the observations, and the mean and standard deviation (STD_DEV) of the percent difference between the retrieved and observed flux.

shown in the HPDFs, which could be compensated for by the different multiple scattering treatments.

5. Flux Comparisons

5.1. Methodology

[34] The microphysics comparison shows obvious discrepancies between the algorithms. While direct comparisons of microphysical quantities retrieved with different algorithms are insightful, they do not provide a measure of success, nor do they provide quantified uncertainty estimates. An independent measure, such as analysis of surface and top of atmosphere (TOA) fluxes, derived from the retrieved microphysical properties, is a possible way to assess the overall uncertainty in the algorithms. In addition to providing an independent measurement, radiative fluxes are used extensively by the modeling community as an evaluation tool. We use radiative flux closure to quantify the retrieval uncertainty in terms of the derived cloud radiative effects. To do this, we compare broadband fluxes computed using the retrieved microphysical properties as input into a radiative transfer model with longwave (LW) and shortwave (SW) broadband fluxes measured by surface (or TOA) radiometers. The “best estimate” quality-controlled surface flux measurement produced by the DOE ARM program (called “QCRAD”) is used as the reference surface flux measurement [Long and Shi, 2006], and LW fluxes derived from geostationary satellites are used as the reference TOA flux measurement [Minnis *et al.*, 2008]. For the TOA comparisons, we focus on the LW fluxes because narrowband to broadband conversions of SW-reflected flux are strongly dependent on solar zenith angle and scene type [Loeb *et al.*, 2005].

[35] For the flux comparisons, the cloud mask is carefully screened to remove profiles that may contain low and middle level liquid water clouds and precipitating clouds.

We again subdivide the data set according to instrument detection; however, since the surface flux represents a hemispheric irradiance, rather than a vertical profile, each profile (rather than each point in the profile) must be classified as a single type. Therefore, the cloud mask is used to identify profiles when 80% of the detections in a single profile can be categorized as radar, rali, or lidar only. The reason that 80% is used (rather than 100%) is because the data set is so dominated by radar detections (Figure 1) that the subsample size for rali and lidar only would be extremely small (for instance, there are no 100% rali profiles in our data set).

[36] The Fu-Liou radiative transfer (RT) model [Fu and Liou, 1992; Fu, 1996] is used to compute the surface fluxes from the retrieved cloud properties. Since the input data set and retrieved quantities (including profiles of temperature, humidity, IWC, and R_e) were already on a common height-time grid, it was straightforward to compute the fluxes and heating rates. A broadband Lambertian surface albedo of 0.095 is assumed. This value represents a mix between the higher albedo of the surfaces at the Darwin ARM site and the lower albedo of the surrounding ocean. A longwave emissivity of 1 is assumed. Surface air temperature is obtained from the Merged Sounding product to represent the surface temperature. The independent pixel approach is used in the radiative transfer calculations, so the radiative heating rates and fluxes are calculated independently for each profile. The combined radar/lidar cloud mask is used to determine whether each height in the profile is clear or cloudy for the radiative transfer calculations.

[37] For each profile, we also calculate the fluxes and heating rates for a corresponding clear sky profile in which the temperature and humidity profiles are the same, but no clouds are included in the computation. By subtracting the calculated clear sky profiles from the all-sky profiles, we can examine the effect of differences in the microphysics on the cloud heating rate profiles. Aerosols are assumed to be negligible, which is generally a fair assumption for Darwin, with the exception of the dry season when agricultural burning takes place. However, the dry season is also typically less cloudy. This technique has been previously applied to other ARM tropical sites to compute radiative fluxes [Mather *et al.*, 2007], where it was shown that computed clear sky fluxes agree to the observed values within <2% in the longwave (LW) and <5% in the shortwave (SW).

[38] Though we have good confidence that the clear-sky fluxes are accurate, there are some assumptions that are made in the Fu-Liou code concerning the scattering properties of the ice crystals that are inconsistent to those made by the Varcloud and RadOn retrieval methods (a mix of ice aggregates and hexagonal columns for Varcloud, variable on a case-to-case basis for RadOn). On the other hand, the CombRet and Rad3mom use the same scattering properties as those assumed in the radiative transfer code. This range of habit assumptions is common in the retrieval and radiative transfer communities as determining the scattering properties of realistic atmospheric ice crystals across the electromagnetic spectrum is an ongoing research topic [Baran, 2012]. In future work, we hope to modify the radiative transfer code to use scattering properties more consistent with the habit assumptions made in the Varcloud and RadOn methods to quantify how much of the difference in the calculated radiative effects is related solely to the habit assumptions.

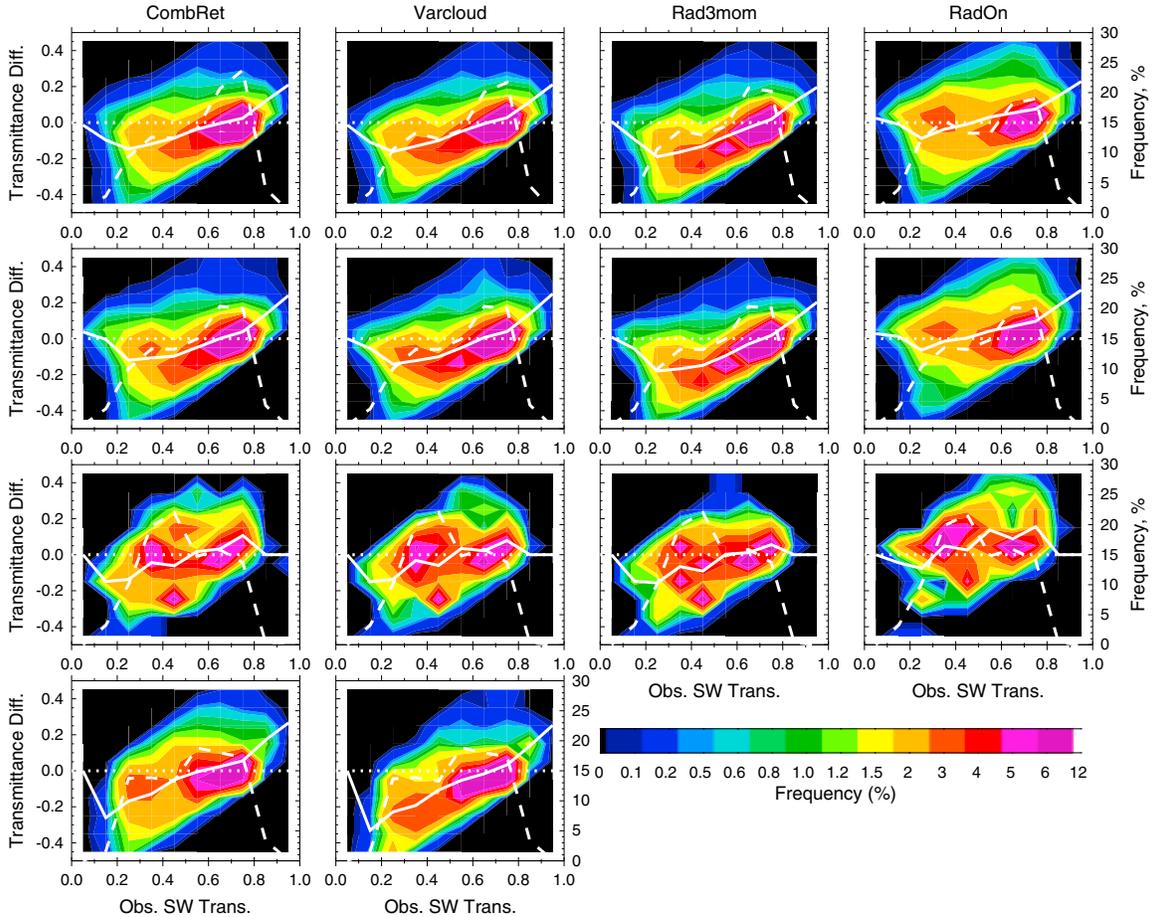


Figure 7. Frequency of SW transmittance difference (observed-calculated) as a function of the observed SW transmittance for each algorithm: (first row) all retrievals, (second row) radar only, (third row) rali, and (fourth row) lidar only. Solid white line represents the mean SW transmittance difference and dashed white line is the frequency of observations in a particular observed SW Transmittance bin.

5.2. Surface Downwelling Shortwave Comparisons

[39] First we compare the computed downwelling shortwave (DSW) flux at the surface with the measured flux (summarized in Table 5). Fluxes are computed using the retrieved cloud properties as input to the RT model. Results are compiled for radar only, rali, lidar only, and all retrievals. For each subcategory (radar only, rali, and lidar only), we only include times when all algorithms report a retrieved value, so there are the same number of points included in each PDF (per category). The exception is for the “all” retrievals category, where we include all the times when an individual algorithm retrieves cloud properties regardless of the method. For example, the Varcloud and CombRet algorithms will include all profiles that are lidar, radar, and rali. RadOn and Rad3mom can be applied to profiles identified as radar only or rali. This subset essentially provides a picture of how well the algorithm performs over each condition. The observed flux in the “all retrievals” case represents all observed flux values when a cloud was detected and any algorithm reports microphysical properties. So for some algorithms (such as RadOn and Rad3mom; see Table 5) the number of points in the “all retrievals” comparisons will be less than the total observed due to fewer cloud detections or when the algorithm fails to converge to a solution.

Compiling the results in this way allows us to compare the full set of potential cloud detections and reveals how well the PDF compares to observations if a significant number of cloud detections is not retrieved by a particular algorithm (i.e., by using only radar or only lidar).

[40] Surface DSW flux measurements occur only during the daylight hours and so are dominated by cirrus anvils produced by diurnally influenced convection [e.g., May *et al.*, 2012; Protat *et al.*, 2009]. For this reason, we expect that the radar subsample will have the largest sample size (Table 5).

[41] One drawback of comparing DSW fluxes at the surface is that the diurnal cycle dependence can mask the differences between large and small optical depths. For a more direct comparison of observed and computed surface fluxes as a function of optical depth, we compute the SW transmittance at the surface (defined as the DSW flux at the surface divided by the DSW flux at the TOA) from both computed and observed surface fluxes (Figure 7). Results are compiled for all retrievals and each subsample. Using SW transmittance, rather than SW flux, removes the diurnal cycle dependence so that performance under different optical depth conditions can be more readily examined. For the “all retrievals” case, Figure 7 demonstrates that for small transmittance values <0.5 (corresponding to large column

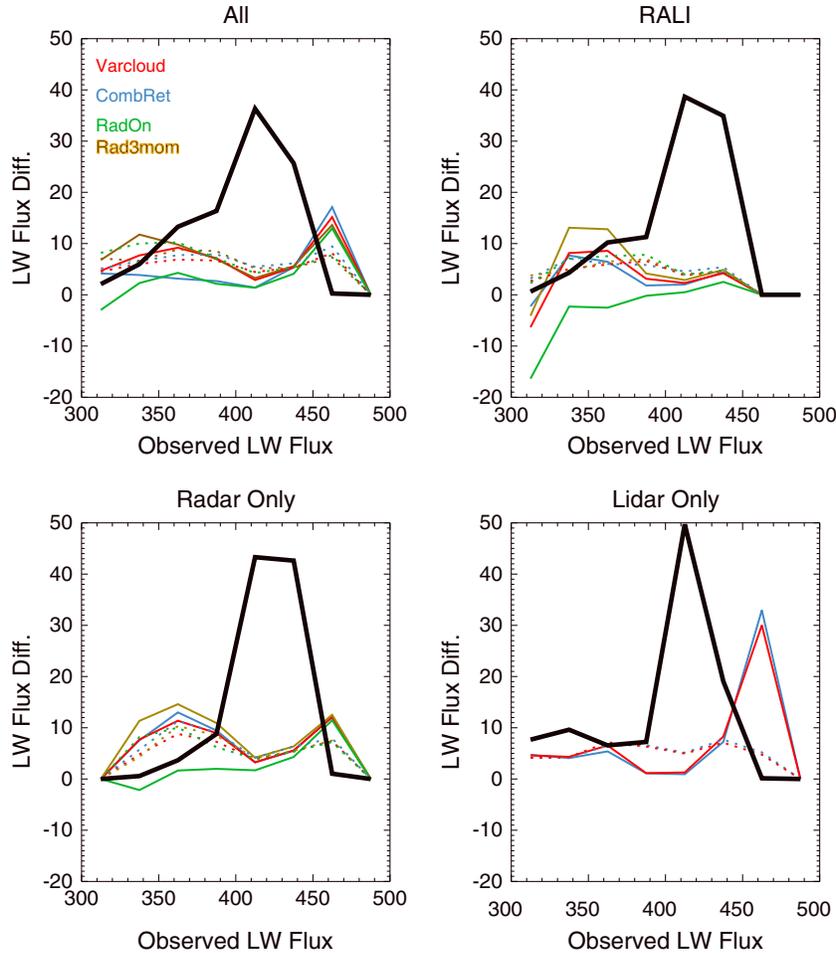


Figure 8. Mean surface LW flux difference (observed-calculated) as a function of the observed surface LW flux for each retrieval (solid lines). Dotted lines are the standard deviation of the mean LW flux difference. The thick solid black line represents the frequency of observations (in %) for each observed LW flux bin. All units are in W m^{-2} .

optical depth), the CombRet, Varcloud, and Rad3mom algorithms are on average biased low as compared to RadOn. RadOn again has a larger variance, particularly for transmittance >0.5 . The results for the radar-only sample are similar to the “all retrievals” case due to the dominance of radar samples. From these results, we can infer that for large optical depth clouds, the empirical approach used by Varcloud and CombRet tends to underestimate the cloud optical depth. Interestingly, the Rad3mom also demonstrates the same trend as Varcloud and CombRet. One of the primary differences between the two Doppler moments algorithms is that the RadOn algorithm retrieves a particle shape, whereas Rad3mom assumes hexagonal crystals, and hence, the mass-dimensional relationships are fixed. This variation on the Z_e-V_d algorithm appears to be an important component in accurately determining the extinction and hence the particle size. For the rali case (Figure 7, third row), CombRet, Varcloud, and to some extent Rad3mom show some improvement over the radar-only sample, particularly for the thin optical depth cases (transmittance >0.5) where the lidar would add the most value to the retrieval. The improved performance of Rad3mom for the rali cases could

indicate that this subset of clouds contains more hexagonal shaped crystals (more in situ generated cirrus, less anvil).

[42] While the transmittance comparisons help to put the algorithm differences in perspective without the diurnal cycle component, we also wish to compare the computed fluxes from the retrieved cloud properties with surface broadband measurements in a statistical way. This type of direct comparison or “closure exercise” helps to quantify the uncertainty in the retrieved microphysical properties. Direct comparisons of surface DSW fluxes (Table 5) indicate that for the “all observations” case, the modeled flux agrees within 20% of the observed surface SW flux over half of the time and within 50% of the observed SW flux, 80–85% of the time for all algorithms. Table 5 also lists the mean and standard deviation of the percent difference between the retrieved and observed surface SW flux. RadOn shows the smallest bias (1.45%) but the largest standard deviation (37.8%). Rad3mom has the largest bias, with CombRet and Varcloud falling in between. Note that the number of retrieved profiles varies among algorithms because cloud properties are not reported if the algorithm does not converge to a solution (for Varcloud, Rad3mom,

Table 6. LW Flux Comparison Statistics for All Observations^a

Retrieval	R^2	<10%	Mean	STD_DEV
<i>TOA LW Fluxes</i>				
CombRet	0.74	48.2	-7.3	6.3
Varcloud	0.72	43.1	-6.2	5.7
RadOn	0.77	42.3	-3.8	10.0
Rad3mom	0.77	33.2	-9.1	4.4
<i>Surface LW Fluxes</i>				
CombRet	0.97	99.9	-0.79	2.0
Varcloud	0.98	99.9	-1.23	1.4
RadOn	0.95	99.9	-0.67	1.5
Rad3mom	0.97	99.9	-1.38	1.8

^a R^2 represents the correlation coefficient between the computed and observed LW flux. Also listed is the percentage of points that are within 10% of the observations and the mean and standard deviation (STD_DEV) of the percent difference between the retrieved and observed flux. Results are tabulated for TOA and surface fluxes for the “All Retrievals.” Results are not significantly different for the various subsamples.

and RadOn). The CombRet applies some type of retrieval (i.e., empirical) for each profile, as long as a valid reflectivity and/or lidar extinction value is available and hence has the largest number of retrieved profiles.

[43] Dividing results by measurement category, rali results have ~30% and 35% of points having uncertainty <10% for CombRet and Varcloud (Table 5), respectively, for the DSW as compared with RadOn (29%) and Rad3mom (40%), which has the smallest uncertainty for the rali conditions. Overall the impact of adding the lidar during rali conditions is mixed because the two rali algorithms have only slightly smaller overall uncertainty compared to RadOn (but larger uncertainty than Rad3mom) as demonstrated in Table 5 rali results. Looking more closely at the rali results in Figure 7 confirms that RadOn is slightly more biased than the others, suggesting that RadOn reflects too much incoming radiation (observed transmittance is larger than modeled). The Doppler velocity measurement appears to be a stronger constraint on the microphysical retrievals for a subset of the observations; however, there are still some details in the algorithm that cause the standard deviation to be very large in a number of cases. Interestingly, CombRet and Varcloud have the smallest mean uncertainty under all sky conditions (Table 5), which could be caused by either their larger sample size or the smaller variance in the uncertainty (Figure 7).

[44] In contrast to the radar-lidar subsample, for the radar-only cases, the two combined retrievals have the smallest uncertainty on average, with ~38% of points agreeing with observations within 10% compared with the RadOn and Rad3mom algorithms (34% and 35%, respectively). The mean flux difference is overall smallest for the radar subsample, with mean differences ranging from 5% to 21%, though the standard deviation (SDEV) remains larger than 30% for all retrievals. It is somewhat surprising that the Z_e-V_d algorithms do not provide significant improvement over the reflectivity only methods (CombRet and Varcloud). Judging from Figure 7 (radar subsample), CombRet, Varcloud, and Rad3mom have slightly less bias and less variance for transmittances larger than 0.5, though RadOn does a better job when transmittance <0.5, which corresponds to optically thicker clouds. Since the frequency of observations is larger for the higher transmittance values, it could explain the seeming smaller uncertainty for the two

radar-lidar algorithms. CombRet and Varcloud have similar uncertainty for lidar-only cases, though the CombRet mean difference is ~9% less than for Varcloud, indicating that the extinction coefficient is better constrained in CombRet, particularly for transmittances <0.4. Despite these statistics and the large $R^2 > 0.9$ for all cases, there is still significant uncertainty in the retrieved cloud properties, as demonstrated by the large number of points with uncertainty >20%.

5.3. Longwave Radiative Flux Comparisons

[45] Longwave (LW) fluxes are primarily driven by the absorption optical depth rather than the scattering component that dominates the shortwave flux. For this reason, we expect that the ice mass and the vertical distribution of this mass will have a larger impact on the surface fluxes than the particle size. In addition, downwelling LW (DLW) fluxes measured at the surface are strongly influenced by the water vapor between the surface and the ground such that the impact of optically thin clouds on the DLW will be below the detection threshold of surface broadband LW measurements. This appears to hold true for the surface DLW flux differences (Figure 8 thin solid lines) where the smaller value of LW flux is associated with optically thin clouds. There are some unique features in Figure 8 that are worth noting. First, there are two peaks in the frequency of observations (thick solid black line): a primary peak between 400 and 450 $W m^{-2}$ and a secondary peak between 300 and 350 $W m^{-2}$. The peak between 400 and 450 $W m^{-2}$ represents the radiative effect due to anvil clouds, whereas the subpeak below 400 $W m^{-2}$ is due to thin cirrus that is detected primarily by lidar. Focusing first on the peak between 400 and 450 $W m^{-2}$ for the “all retrievals” case, the agreement is consistent for all algorithms, with CombRet and RadOn having a smaller mean difference in the 400–450 $W m^{-2}$ peak. The SDEV (dotted line) in the primary peak is <5 $W m^{-2}$ for all algorithms. For the secondary peak (300–350 $W m^{-2}$), the difference among algorithms is much larger and the SDEV is larger, particularly for RadOn. This may be indicative that the reflectivity-based algorithms are less sensitive to these thin clouds. Varcloud and Rad3mom are more biased than the other two algorithms for the secondary peak. Results are similar for the radar-only subset, except that the CombRet is more biased in the secondary peak than in the “all retrievals” case. In the rali case, RadOn and Rad3mom have large biases in secondary peak, though biased in opposite directions and the algorithms that use radar and lidar to derive cloud properties have smaller biases for thin clouds, which is as expected. All algorithms are less biased and have smaller SDEV in the primary peak, indicating that the retrieval of cloud properties in thicker anvil clouds is better constrained than for thin clouds, at least from the LW perspective. It also indicates that the location and vertical distribution of the IWC is fairly well characterized for these cases. Overall, the surface LW flux comparisons summarized in Table 6 show that the results are highly correlated ($R^2 > 0.9$) and the mean percent difference is <2%, with a comparable SDEV. Diagnostics of method performance in different flux ranges will be of great help to guide further retrieval method improvements. It is important to mention that the DLW radiative effect changes significantly when single remote sensors are used to retrieve cloud properties as apparent in the

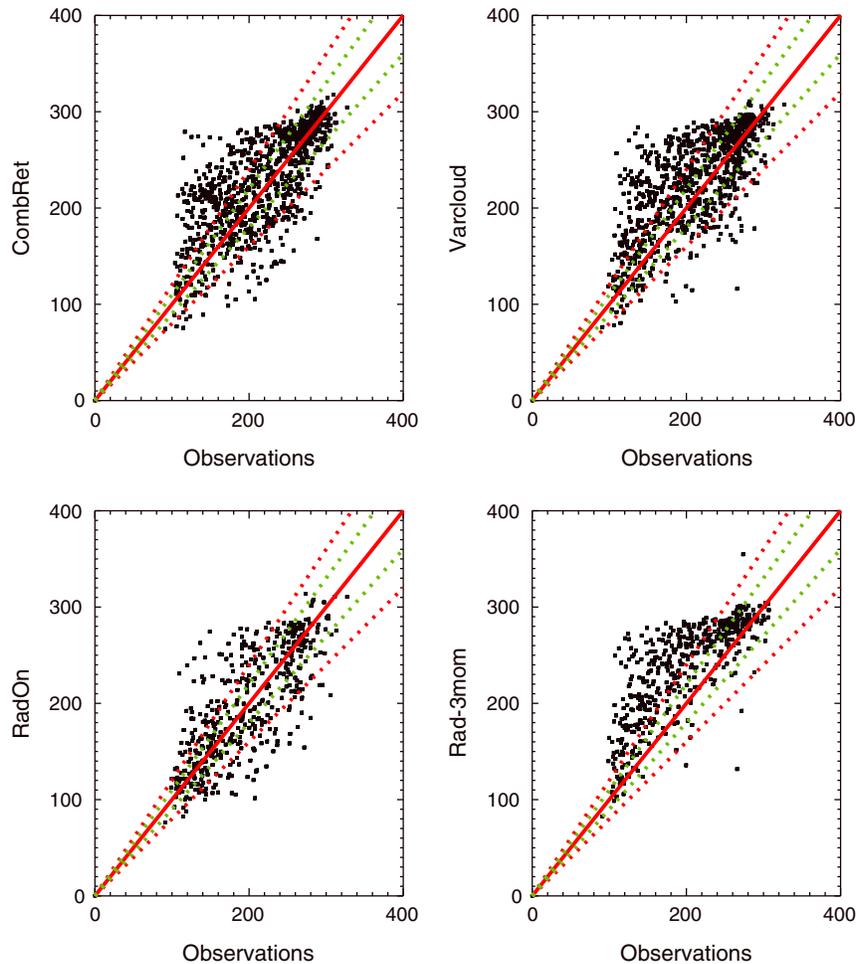


Figure 9. TOA LW flux comparisons for all retrievals. Green and red lines represent 10% and 20% uncertainty, respectively. All units are in W m^{-2} .

Varcloud and CombRet results for “all retrievals.” Direct comparisons of DLW fluxes at the surface reveal that more than 99% of points have an uncertainty $<10\%$ for all algorithms (Table 6).

[46] An additional constraint on algorithm uncertainty is shown in the direct comparison of measured and retrieved upwelling fluxes at the top-of-atmosphere (TOA). We compare the LW TOA flux measurements from the satellite-based pixel level data product VISST (Visible Infrared Solar-Infrared Split Window Technique) [Minnis *et al.*, 2008] with LW TOA fluxes computed using the retrieved cloud properties (Figure 9). Error in measured TOA fluxes is roughly $3\text{--}5 \text{ W m}^{-2}$ with biases ranging from 0.2 to 0.4 W m^{-2} [Loeb *et al.*, 2007]. VISST data are derived from MTSAT satellite observations and have $\sim 4 \text{ km}$ spatial resolution. Fluxes for the 9 pixels centered on the nearest pixel to the Darwin site are averaged to obtain the TOA flux. Given the 4 km pixel size, this 9 pixel average likely includes both ocean and land pixels. The exact geolocation of each pixel is somewhat uncertain due to the uncertainty in the satellite navigation system. VISST data are available only once per hour for a 5 month period (January–February 2006 and October–December 2007) over the Darwin site, so we only compare the closest VISST pixel to the “all retrievals” case to have sufficient numbers

of data points to compute statistics. Note that this VISST product is also a retrieval algorithm (although the technique has been “trained” with TOA flux measurements), so it cannot be fully considered as a “reference,” which was the case for the surface comparisons.

[47] Correlation coefficients (R^2) between the computed and observed upwelling TOA LW fluxes for each algorithm in Figure 9 are tabulated in Table 6. All algorithms have similar $R^2 \sim 0.7$ and between 62% and 73% of the computed values fall within 10% of the observed TOA flux depending on the algorithm and roughly 87–96% of points are within 50% of the observed flux. The two combined retrievals (CombRet and Varcloud) have very similar uncertainty when compared with the observations (mean percent difference is 6–7% and SDEV $\sim 5\text{--}6\%$), which is slightly larger than the surface measurements. RadOn agrees more frequently with observations than compared with Rad3mom (by $\sim 10\%$), which is likely due to the tendency for Rad3mom to have smaller IWC values. This somewhat lower performance of Rad3mom corresponds to the intermediate LW TOA fluxes (see biases in the $150\text{--}200 \text{ W m}^{-2}$ range in Figure 9). One contributor to the larger uncertainties seen in the TOA LW flux comparisons (over the surface LW fluxes) is due to reduced cloud detection by lidar or radar depending on the conditions.

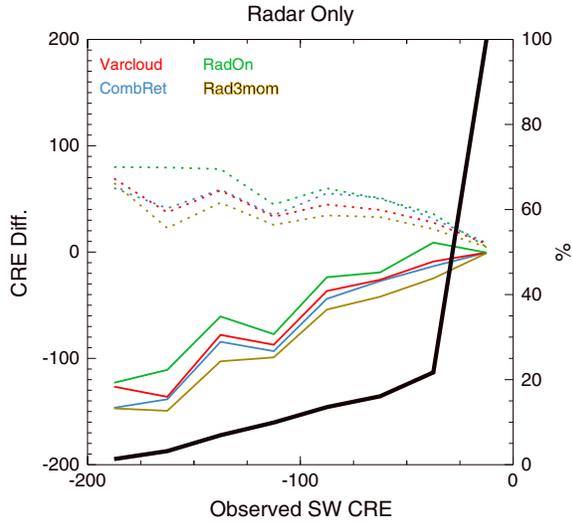


Figure 10. Mean SW CRE difference (observed-calculated) as a function of observed SW CRE for each algorithm (thin solid lines). Dotted lines are the standard deviation of the mean CRE difference. The thick solid black line represents the frequency of observations (in %, right axis) for each observed CRE bin. CRE units are in $W m^{-2}$.

As was noted by *Borg et al.* [2011], the TOA LW fluxes are significantly impacted when the radar does not detect cloud top, or likewise when the lidar does not detect thin cirrus due to poor signal-to-noise ratio, which is often the case with the

MPL. In recent work not included in this study, we find that using the new Raman lidar located at the Darwin site improves the detection of high thin cirrus over the MPL. Using these improved measurements to better understand the radiative impact of topical clouds will be the topic of future work.

[48] For the “lidar only” subsample, we have performed a sensitivity test using the CombRet where we assume a constant lidar ratio of 33 to help understand the impact of extinction uncertainty on the computed fluxes. The results indicate that on average, the mean difference in TOA LW fluxes is reduced by 1% and SDEV is reduced by 3.8% when assuming $S_p = 33$ sr using the CombRet. The impact on surface LW fluxes is opposite in that the mean difference increases by 0.6% and the SDEV increases by 0.5% when assuming $S_p = 33$ sr. While these changes are relatively small, it would be worthwhile in future work to better constrain S_p by looking at direct measurements of S_p from high spectral resolution or Raman lidar systems in different climate regimes.

6. Cloud Radiative Effect

[49] To assess the impact of the uncertainty in the retrieved microphysical properties on the radiative effect of clouds, we examine differences in the cloud radiative effect (CRE, defined as cloudy minus calculated clear sky) in terms of both the fluxes and the heating rates for each retrieval subsample. We note again that these subsamples only include profiles that contain ice clouds with no underlying liquid clouds or precipitation and thus do not represent the full radiative effect of ice clouds observed at Darwin.

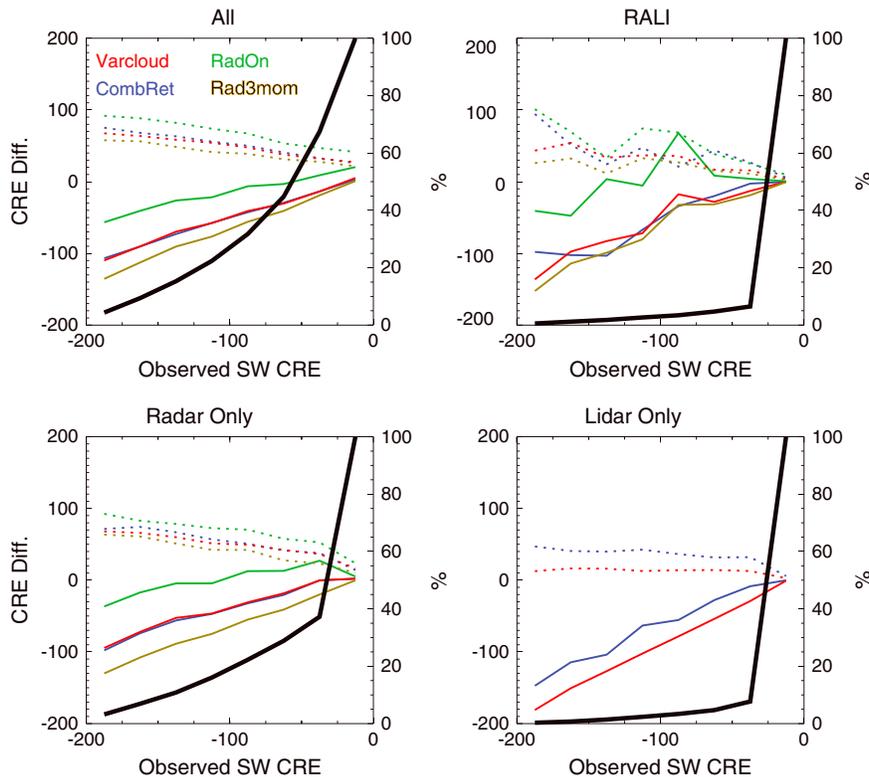


Figure 11. Same as in Fig. 10 except observations when the $fsc < 90\%$ are removed. Results are for the radar subsample only. CRE units are in $W m^{-2}$.

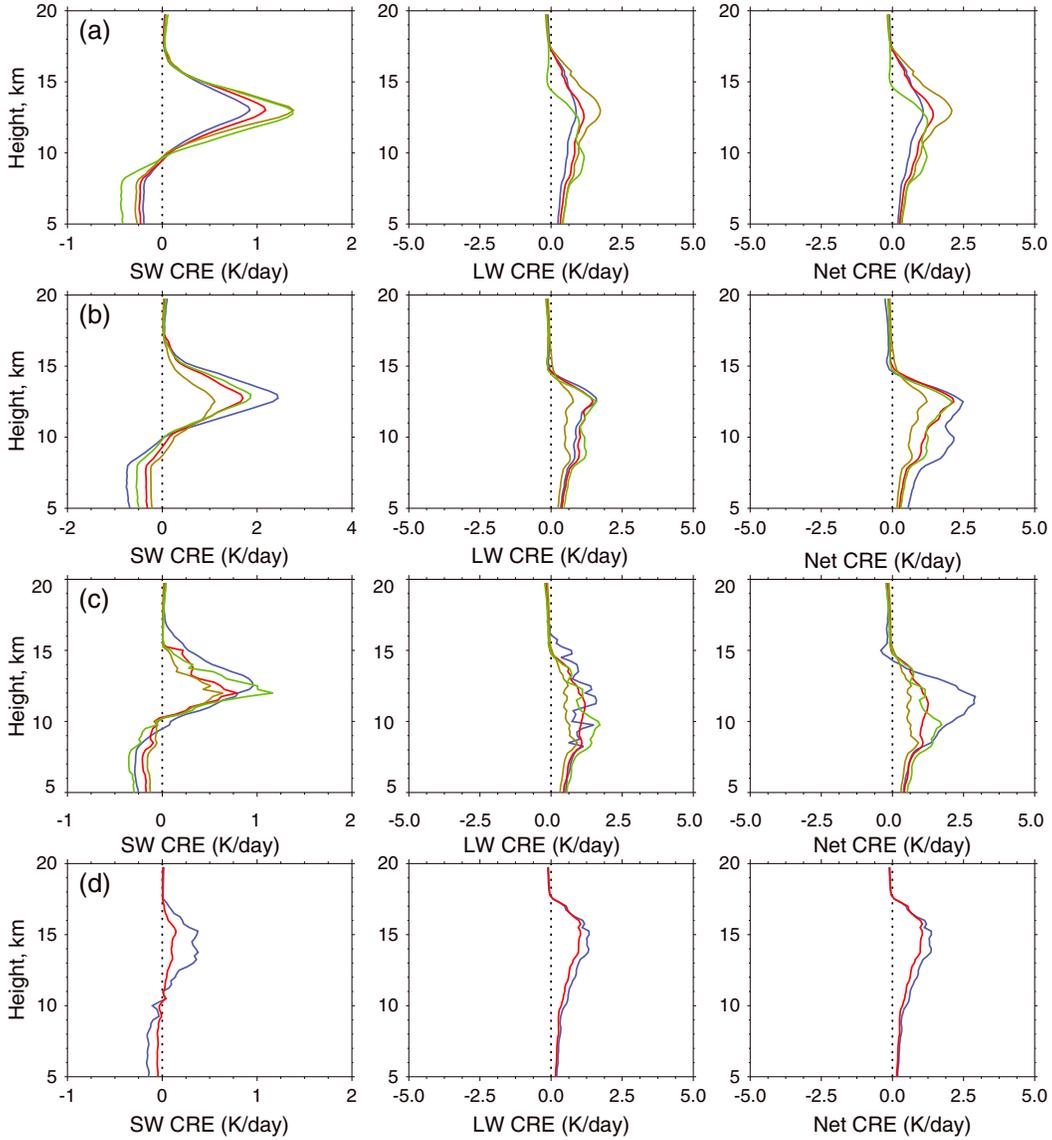


Figure 12. Mean profiles of cloud radiative effect (cloudy-clear sky) on the heating rate profile for (first row) all algorithms, (second row) radar only, (third row) rali, and (fourth row) lidar only. Individual retrievals are CombRet (blue), Varcloud (red), Rad3mom (yellow), and RadOn (green).

[50] As mentioned in section 5, the largest difference between the four algorithms is in the retrieved R_e . First we look at the SW CRE at the surface. For the “radar only” subsample, RadOn overall has the smallest mean difference for the range of observed SW flux CRE (Figure 10), but the SDEV of the CRE difference is larger than for other methods (dotted lines). A closer look indicates that for an observed CRE $< 50 \text{ W m}^{-2}$, Varcloud has a mean difference of $\sim 0 \text{ W m}^{-2}$ with SDEV up to 40 W m^{-2} , which constitutes $\sim 65\%$ of the total sample size. Although there are some offsetting uncertainties, RadOn overall outperforms the other algorithms, particularly for large CRE $> -200 \text{ W m}^{-2}$. For the “rali” subsample, RadOn agrees with the observed CRE within 10 W m^{-2} for up to 98% of the observations. The other three algorithms have larger mean differences (up to 30 W m^{-2} for observed CRE $< 50 \text{ W m}^{-2}$). The SDEV for Rad3mom and Varcloud is about half that of RadOn and CombRet. The poor agreement for all

algorithms for observed SW CRE $> 50 \text{ W m}^{-2}$ constitutes only a small percentage of the observations ($< 8\%$). Interestingly, for the rali cases, RadOn sometimes produces a good estimate of CRE between -400 and -200 W m^{-2} , but at other times (greater than -400 W m^{-2}) performs poorly. This will require further investigations. Lidar-only results show that the CombRet produces a smaller mean CRE difference than Varcloud, but the SDEV is about twice that of Varcloud for the majority of the observations ($< 50 \text{ W m}^{-2}$). It is noteworthy that for the lidar and rali subsamples, $\sim 95\%$ of all observations have CRE $< 50 \text{ W m}^{-2}$, and the mean CRE difference is between 10 – 20 W m^{-2} depending on the algorithm used.

[51] In all comparisons up to this point, we have considered the flux differences without considering the fractional sky cover (f_{sc}) during the time of the observation. Part of our reasoning for including all-sky cover situations is due to the need for retrieval algorithms to be applied for all-sky

conditions so that continuous data sets can be utilized for model evaluation and long-term studies. The uncertainty in the computed fluxes is expected to increase when the f_{sc} significantly deviates from 100%. To examine the impact of f_{sc} on the flux closure exercise, we use the radiative flux analysis data product [Long and Ackerman, 2000; Long *et al.*, 2006] to determine the f_{sc} and remove observations with $f_{sc} < 90\%$ from the comparison. The results for the radar-only subsample are included in Figure 11. The sample size for the lidar-only case did not change significantly, and for the rali case, the sample size was reduced so low that a comparison was not statistically significant, so these cases are not shown. For the radar-only case, the spread between the four algorithms collapses from $\sim 100 \text{ W m}^{-2}$ to $\sim 20 \text{ W m}^{-2}$ difference, and the SDEV decreases particularly for observed SW CRE $> 100 \text{ W m}^{-2}$. This indicates that many of the points with large CRE difference may be impacted by an inhomogeneous cloud field within the radiometer field of view, which results in a larger uncertainty of $\sim 20\%$ at an observed CRE of 100 W m^{-2} . Given this analysis, we do not believe that including some cases with smaller f_{sc} will impact the conclusions of this paper.

[52] Lastly, we compare the cloud effect on the SW, LW, and net radiative heating rate profiles as produced by the four methods and for the different subsamples (Figure 12). Note that in Figure 12, the average SW heating rate includes daytime profiles only, whereas the LW and net heating rates include both daytime and nighttime profiles. For each algorithm, we average only profiles that contain clouds and that were included in the flux analyses. For the “all retrievals” case, RadOn produces $\sim 1.5 \text{ K d}^{-1}$ SW heating, which is roughly comparable to Rad3mom. CombRet and Varcloud produce ~ 0.75 and 1 K d^{-1} of SW heating, respectively. RadOn produces LW cloud top cooling, whereas the other three algorithms do not, and RadOn has a “level of zero net heating” that is nearly 3 km lower than the other algorithms. Note that the differences in average heating rates for the “all retrievals” case, which includes all of the retrievals for a particular algorithm, are due in part to the different number of profiles retrieved by each algorithm (Table 5). In particular, the CombRet and Varcloud “all retrievals” profiles include the optically thinner lidar-only clouds, which tend to have smaller SW heating than the radar-detected clouds, and also have LW heating that occurs at higher altitudes.

[53] The average LW CRE heating rate profiles for the rali and “radar only” subsamples are fairly similar for each algorithm with the exception of Rad3mom, which is slightly smaller. The SW CRE heating rate differences are larger than the LW CRE heating rate difference, which is expected due to the differences in the R_e . LW CRE heating rate similarities are driven primarily by the ice mass and its vertical distribution. The results for the “lidar only” subsample suggest that clouds retrieved by CombRet have a larger CRE heating rate than those retrieved by Varcloud.

[54] The differences in the heating rate profiles which are due both to the different sampling of clouds by each retrieval algorithm and to the differences in retrieved properties, could be important if the retrieved heating profiles are used to evaluate or constrain models. Heating rate differences could have large impacts on cloud-scale dynamics as the vertical structure of radiative heating

can modify the stability of the cloud layer and affect the in-cloud vertical motion and maintenance of the cloud layer [Liu *et al.*, 2003]. Additionally, the level of net zero heating, which is the height of the transition between radiative cooling and heating in tropical tropopause layer (TTL), is impacted by the altitude and radiative properties of clouds, which will affect the vertical mass fluxes in this transition layer [Corti *et al.*, 2005].

7. Discussion and Summary

[55] The microphysics comparisons and radiative flux closure results provide a quantitative assessment of the uncertainty in computed fluxes as a result of differences in the retrieved microphysical properties of ice clouds. Several interesting findings are revealed through this intercomparison:

[56] 1. Although IWC PDFs are similar, differences are apparent in the vertical distribution, which is very important for the LW calculations. While IWC HPDFs are comparable between roughly 10 and 12 km for all algorithms, Rad3mom IWC is smaller at higher altitudes, whereas RadOn is smaller at lower altitudes. These differences in the vertical distribution contribute to the heating rate differences.

[57] 2. With the exception of RadOn, the retrieved R_e has a similar peak frequency ($\sim 40 \mu\text{m}$) for clouds detected by radar. RadOn produces on average a factor of two smaller R_e than the other algorithms above 8 km height. However, fall velocity estimates for RadOn are larger than Rad3mom, particularly below 10 km (i.e., Figure 4). Although this implies that small crystals fall faster, this is not necessarily the case. While Rad3mom assumes a specific relationship between R_e and V_f , RadOn does not: they are derived independently. Hence, for the same R_e , fall speed retrieved by RadOn would actually be larger than that retrieved by Rad3mom. For lidar-only clouds (Figure 5), peak frequency is $\sim 25 \mu\text{m}$, but on average, Varcloud is $\sim 5 \mu\text{m}$ smaller than CombRet.

[58] 3. The R_e differences produce discrepancies in computed DSW radiative fluxes (Figure 7), and have a profound influence on the mean radiative heating profile (Figure 12). However, in terms of mean SW flux statistics, the R_e differences result in ~ 16 – 26% difference in (Table 5, column “Mean”) for radar-only cases, and up to 6–10% difference for all retrievals for the $<10\%$ category. The SDEV of the flux differences for individual algorithms is still quite large compared to the mean difference.

[59] 4. Retrieval uncertainty (as determined in this study by error in radiative flux closure) is still large, with approximately 36–46% of points having $>20\%$ error in computed surface fluxes depending on the algorithm (all retrievals case). Error is reduced with respect to the TOA LW, where up to $\sim 70\%$ of points have error $<20\%$. At least for this tropical cloud data set, the radar-lidar algorithms do not reduce the uncertainty in retrieved properties over Doppler radar algorithms, but do appear to be less biased. As mentioned previously, the number of profiles that meet the criteria for the rali case is small due to the extensive presence of optically thick anvil clouds that the lidar does not penetrate and optically thin clouds that the radar often does not detect.

[60] Naturally, there are several caveats worth mentioning that will impact the results and contribute to the variance as

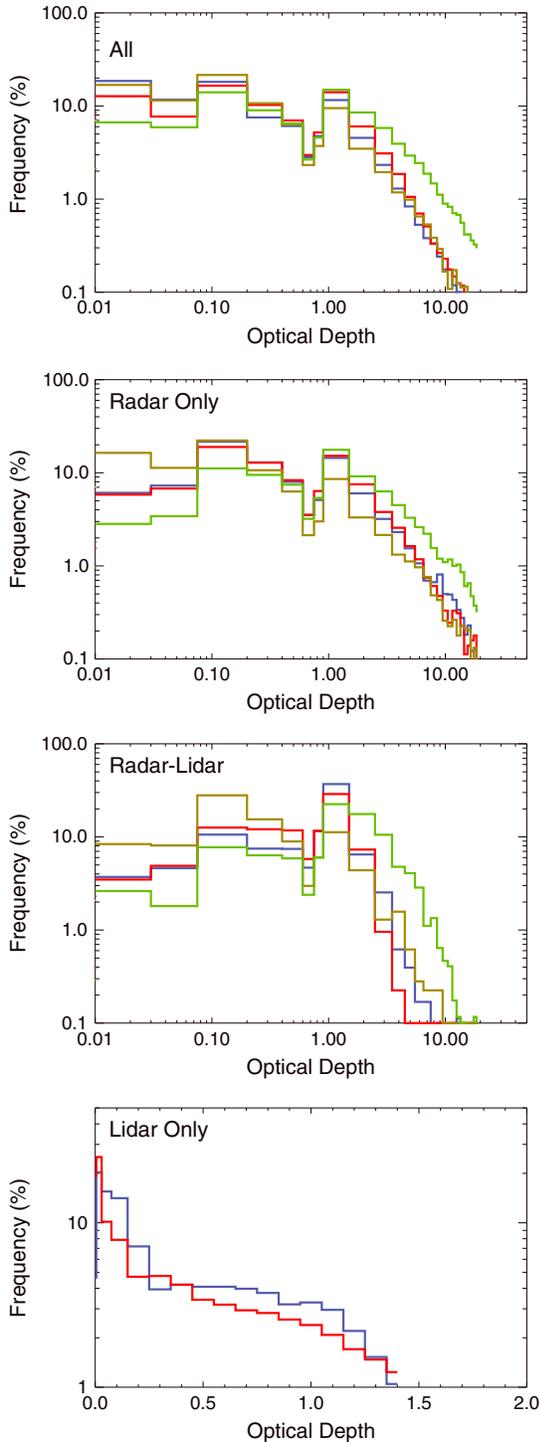


Figure 13. Optical depth distribution for each subsample and algorithm type. Individual retrievals are CombRet (blue), Varcloud (red), Rad3mom (yellow), and RadOn (green).

well. First, total sky cover was not a criterion used to screen the data, which means we did not compare just overcast skies. There is a mismatch in viewing angles between the profiling active remote sensors and the hemispherical view of the surface radiometers. Non-overcast skies can introduce errors in the one-dimensional RT calculations due to neglecting multiple scattering contributions, though cloud

fraction will likely have a more significant impact on the computed downwelling fluxes. Computed uncertainties could potentially be less due to these effects. In future work, versions of these retrievals could be applied to scanning cloud radar measurements recently installed at the ARM sites to reduce these potential 3-D effects. Second, assumptions concerning the particle shape and mass-dimensional relationships are not consistent between the retrieval algorithms and the RT model with the exception of *Wang and Sassen* [2002], which assumes the *Fu* [1996] parameterizations to relate extinction and particle size. This discrepancy could increase the uncertainty for the RadOn and Varcloud algorithms, and the use of a single particle habit for all clouds in microphysical retrievals and RT calculations is a source of uncertainties as well. Third, the uncertainty in computed fluxes varies depending on the total column optical depth, that is, algorithms sometimes perform better or worse for small or large optical depths. This is evident in the transmittance comparisons (Figure 8), but can also be evident by looking at the distribution of retrieved optical depth (Figure 13) and transmittance difference as a function of optical depth (Figure 14). In Figure 13, the first notable feature is the bimodal distribution of retrieved optical depths. We speculate that the peak observed at the smallest τ is due to thin in situ generated cirrus, whereas the second peak at $\tau \sim 1$ is due to anvil generated cirrus. Note that in previous studies of tropical cirrus over Nauru Island [*Comstock et al.*, 2002], this bimodal distribution was not observed because the cloud observations were dominated by in situ cirrus due to somewhat suppressed convective activity over the duration of that study. Here we demonstrate the bimodality of cirrus optical depth PDF for a regime influenced by both convective activity and suppressed conditions. It is also apparent from Figure 13 that both peaks are seen in all algorithm subsamples, including the lidar-only case. Note for the rali case that the overlap of radar and lidar data appears to drop off significantly after $\tau \sim 2$ for most algorithms (though RadOn has a much larger frequency of $\tau > 2.0$ due to the smaller R_e). This demonstrates the range of optical depth where the rali algorithms are applicable.

[61] Given the optical depth distributions, we divide the SW transmittance comparisons as a function of optical depth (Figure 14). For small optical depth ($\tau < 1.0$), all four algorithms agree well with observed transmittance larger than ~ 0.6 , and the rali algorithms (Varcloud and CombRet) actually produce better agreement than the radar-only algorithms for this optical depth subset (that is, when transmittance is larger than 0.6, which has the largest sample size). These larger transmittance cases correspond with the thinnest τ . We suspect that the large disagreement in the small transmittance values demonstrates the viewing mismatch between the hemispheric surface SW flux measurement and the profiling instruments (e.g., there could have been boundary layer clouds present out of the viewing angle of the active sensors). These results for $\tau < 1$ do suggest that the rali algorithms provide some improvement over the radar-only algorithms in the small optical depth regime. For $\tau > 1$, smaller transmittance values (representing the larger optical depth cases) indicate that RadOn transmittance is good for transmittances lower than 0.3, then biased low (observed transmittance is more

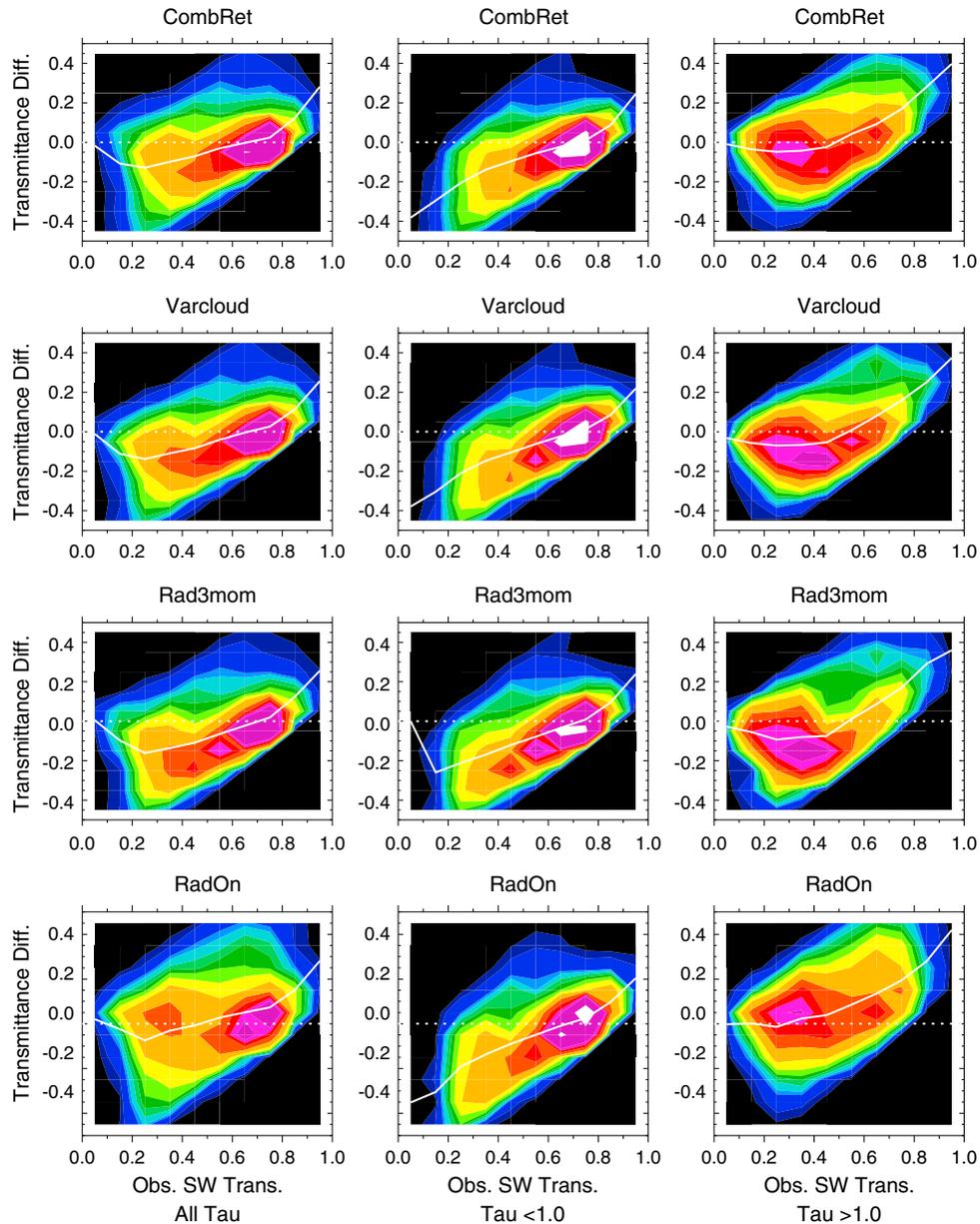


Figure 14. Observed SW transmittance as a function of transmittance difference (observed-calculated) shown for all optical depths, and divided into groups of small $\tau < 1.0$ and large $\tau > 1.0$. Results are for all retrievals.

often positive) suggesting an overestimation of the optical depth, whereas the other three algorithms retrieve a transmittance that is for the most part too large, suggesting an underestimation of the optical depth. Since the IWC for all algorithms compared better than the R_e , this result suggests that RadOn underestimates the R_e , whereas the three other methods overestimate the R_e .

[62] Given these results, we can reflect back on the Vogelmann and Ackerman study that determined that an error of $\pm 12\%$ in extinction optical depth τ would allow the net surface fluxes to be computed within $\pm 5\%$. Unfortunately, we are somewhat restricted in our ability to show how close we have come to achieving this uncertainty because we cannot compare directly the optical depth from the SW flux observation and that from the “pencil-beam”

remote sensors. Instead, Figure 14 suggests that the mean transmittance difference is roughly ± 0.1 for either large or small optical depth provided that the appropriate algorithm is applied depending on the sky conditions. To improve the SW closure study, the view-angle mismatch could be alleviated by utilizing narrow-beam zenith-pointing short-wave measurements, such as the zenith-pointing version of the shortwave array spectroradiometer (SAS) now deployed at the ARM Southern Great Plains site or for use with the ARM Mobile Facility. This new instrument will provide a better reference for SW flux closure studies. Likewise, the narrow field-of-view AERI would also provide some constraints on the surface LW comparisons.

[63] Clearly, there is more work to be done to fully understand the differences between algorithms and guide

improvements of these algorithms using radiative flux closure. The results presented here suggest that R_e differences between RadOn and other algorithms could be a function of particle shape assumptions. Pinning down the source of the R_e difference will be an important focus in future work.

[64] **Acknowledgments.** This research was supported by the DOE Atmospheric System Research Program and data used in the study were obtained by the DOE Atmospheric Radiation Measurement Program. The PNNL authors were also partly funded by the NASA Energy and Water Cycle Study (NEWS). Julien Delanoë's research is partly funded by CNES (Centre National d'Études Spatiales). We wish to thank Mandy Khaiyer for assistance in obtaining and utilizing the satellite flux data and Chuck Long for providing his radiative flux analysis data set. We also wish to thank three anonymous reviewers for their insightful comments that have helped to improve the presentation of the results.

References

- Ackerman, T. P., and G. M. Stokes (2003), The Atmospheric Radiation Measurement Program, *Phys. Today*, *56*, doi:10.1063/1.1554135.
- Baran, A. (2012), From the single-scattering properties of ice crystals to climate prediction: A way forward, *Atmos. Research*, *112*, 45–69.
- Borg, L. A., R. E. Holz, and D. D. Turner (2011), Investigating cloud radar sensitivity to optically thin cirrus using collocated Raman lidar observations, *Geophys. Res. Lett.*, *38*, L05807, doi:10.1029/2010GL046365.
- Brown, P. R. A. and P. N. Francis (1995), Improved measurements of the ice water content in cirrus using a total-water probe, *J. Atmos. Oceanic Technol.*, *12*, 410–414.
- Comstock, J. M., and K. Sassen (2001), Retrieval of cirrus cloud radiative and backscattering properties using combined lidar and infrared radiometer (LIRAD) measurements, *J. Atmos. Oceanic Technol.*, *18*, 1658–1673.
- Comstock, J. M., T. P. Ackerman, and G. G. Mace (2002), Ground-based lidar and radar remote sensing of tropical cirrus clouds at Nauru Island: Cloud statistics and radiative impacts, *J. Geophys. Res.*, *107*(D23), 4714, doi:10.1029/2002JD002203.
- Comstock, J. M., et al. (2007), An intercomparison of microphysical retrieval algorithms for upper-tropospheric ice clouds, *Bull. Amer. Meteor. Soc.*, *88*, 191–204.
- Corti, T., B. P. Luo, T. Peter, H. Vömel, and Q. Fu (2005), Mean radiative energy balance and vertical mass fluxes in the equatorial upper troposphere and lower stratosphere, *Geophys. Res. Lett.*, *32*, L06802, doi:10.1029/2004GL021889.
- Delanoë, J., A. Protat, J. Testud, D. Bouniol, A. J. Heymsfield, A. Bansemmer, P. R. A. Brown, and R. M. Forbes (2005), Statistical properties of the normalized ice particle size distribution, *J. Geophys. Res.*, *110*, D10201, doi:10.1029/2004JD005405.
- Delanoë, J., A. Protat, D. Bouniol, A. Heymsfield, A. Bansemmer, and P. Brown (2007), The characterization of ice cloud properties from Doppler radar measurements, *J. Appl. Meteor. Climatol.*, *46*, 1682–1698.
- Delanoë, J., and R. J. Hogan (2008), A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer, *J. Geophys. Res.*, *113*, D07204, doi:10.1029/2007JD009000.
- Deng, M., and G. G. Mace (2006), Cirrus microphysical properties and air motion statistics using cloud radar Doppler moments. Part I: Algorithm description, *J. Appl. Meteor. Climatol.*, *45*, 1690–1709.
- Donovan, D. P., and A. C. A. P. van Lammeren (2001), Cloud effective particle size and water content profile retrievals using combined lidar and radar observations 1. Theory and examples, *J. Geophys. Res.*, *106*(D21), 27,425–27,448, doi:10.1029/2001JD900243.
- Francis, P. N., P. Hignett, and A. Macke (1998), The retrieval of cirrus cloud properties from aircraft multi-spectral reflectance measurements during EUCREX'93, *Q. J. R. Meteorol. Soc.*, *124*, 1273–1291.
- Fu, Q. (1996), An accurate parameterization of the solar radiative properties of cirrus clouds for climate models, *J. Climate*, *9*, 2058–2082.
- Fu, Q., and K. N. Liou (1992), On the correlated k -distribution method for radiative transfer in nonhomogeneous atmospheres, *J. Atmos. Sci.*, *49*, 2139–2156.
- Heymsfield, A. J., D. Winker, and G.-J. van Zadelhoff (2005), Extinction-ice water content-effective radius algorithms for CALIPSO, *Geophys. Res. Lett.*, *32*, L10807, doi:10.1029/2005GL022742.
- Heymsfield, A. J., et al. (2008) Testing IWC retrieval methods using radar and ancillary measurements with in situ data, *J. Appl. Meteor. Climatol.*, *47*, 135–163.
- Hogan, R. J. (2006), Fast approximate calculation of multiply scattered lidar returns, *Appl. Opt.*, *45*, 5984–5992.
- Hogan, R. J., M. P. Mittermaier, and A. J. Illingworth (2006), The retrieval of ice water content from radar reflectivity factor and temperature and its use in evaluating a mesoscale model, *J. Appl. Meteor.*, *45*, 301–317.
- Hogan, R. J., L. Tian, P. R. A. Brown, C. D. Westbrook, A. J. Heymsfield, and J. D. Eastment (2012), Radar scattering from ice aggregates using the horizontally aligned oblate spheroid approximation, *J. Appl. Meteor. Climatol.*, *51*, 655–671.
- Illingworth, A. J., et al. (2007), Cloudnet, *Bull. Amer. Meteor. Soc.*, *88*, 883–898.
- Intrieri, J. M., G. L. Stephens, W. L. Eberhard, and T. Uttal (1993), A method for determining cirrus cloud particle sizes using lidar and radar backscatter technique, *J. Appl. Meteor.*, *32*, 1074–1082.
- Liu, H.-C., P. K. Wang, and R. E. Schlessinger (2003), A numerical study of cirrus clouds. Part II: Effects of ambient temperature, stability, radiation, ice microphysics, and microdynamics on cirrus evolution, *J. Atmos. Sci.*, *60*, 1097–1119.
- Liebe, H. (1985), An updated model for millimeter wave propagation in moist air, *Radio Sci.*, *20*, 1069–1089.
- Liu, C.-L., and A. J. Illingworth (2000), Toward more accurate retrievals of ice water content from radar measurements of clouds, *J. Appl. Meteor.*, *39*, 1130–1146.
- Loeb, N. G., S. Kato, K. Loukachine, and N. Manalo-Smith (2005), Angular distribution models for top-of-atmosphere radiative flux estimation from the Clouds and the Earth's Radiant Energy System instrument on the Terra satellite. Part I: Methodology, *J. Atmos. Oceanic Technol.*, *22*, 338–351.
- Long, C. N., and T. P. Ackerman (2000), Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects, *J. Geophys. Res.*, *105*, No. D12 (2000), pp. 15609–15626.
- Long, C. N., and Y. Shi (2006), The QCRad value added product: Surface radiation measurement quality control testing, including climatology configurable limits. DOE/SC-ARM/TR-074, 43 pp., Dep. of Energy, Washington, D. C. (Available at <http://www.arm.gov/publications/tech-reports/>).
- Long, C. N., T. P. Ackerman, K. L. Gaustad, and J. N. S. Cole (2006), Estimation of fractional sky cover from broadband shortwave radiometer measurements, *J. Geophys. Res.*, *111*, D11204, doi:10.1029/2005JD006475.
- Mace, G. G., T. P. Ackerman, P. Minnis, and D. F. Young (1998), Cirrus layer microphysical properties derived from surface-based millimeter radar and infrared interferometer data, *J. Geophys. Res.*, *103*, 23,207–23,216.
- Mace, G. G., A. J. Heymsfield, and M. R. Poellot (2002), On retrieving the microphysical properties of cirrus clouds using the moments of the millimeter-wavelength Doppler spectrum, *J. Geophys. Res.*, *107*(D24), 4815, doi:10.1029/2001JD001308.
- Mace, G. G., M. Deng, B. Soden, and E. Zipser (2006), Association of tropical cirrus in the 10–15-km layer with deep convective sources: An observational study combining millimeter radar data and satellite-derived trajectories, *J. Atmos. Sci.*, *63*, 480–503.
- Mather, J. H., S. A. McFarlane, M. A. Miller, and K. L. Johnson (2007), Cloud properties and associated radiative heating rates in the tropical western Pacific, *J. Geophys. Res.*, *112*, D05201, doi:10.1029/2006JD007555.
- Matrosov, S. Y., B. W. Orr, R. A. Kropfli, and J. B. Snider (1994), Retrieval of vertical profiles of cirrus cloud microphysical parameters from Doppler radar and infrared radiometer measurements, *J. Appl. Meteor.*, *33*, 617–626.
- Matrosov, S. Y. (1999), Retrievals of vertical profiles of ice cloud microphysics from radar and IR measurements using tuned regressions between reflectivity and cloud parameters, *J. Geophys. Res.*, *104*(D14), 16,741–16,753, doi:10.1029/1999JD900244.
- Matrosov, S. Y., A. V. Korolev, and A. J. Heymsfield (2002), Profiling cloud ice mass and particle characteristic size from Doppler radar measurements, *J. Atmos. Oceanic Technol.*, *19*, 1003–1018.
- May, P. T., C. N. Long, and A. Protat (2012), The diurnal cycle of the boundary layer, convection, clouds, and surface radiation in a coastal monsoon environment (Darwin, Australia), *J. Climate*, *25*, 5309–5326.
- Minnis, P., et al. (2008), Near-real time cloud retrievals from operational and research meteorological satellites, *Proc. SPIE Europe Remote Sens.* 2008, Cardiff, Wales, UK, 15–18 September, 7107–2, 8 pp.
- Mitchell, D. L. (1996), Use of mass- and area-dimensional power laws for determining precipitation particle terminal velocities, *J. Atmos. Sci.*, *53*, 1710–1723.
- Mitchell, D. L., and A. J. Heymsfield (2005), Refinements in the treatment of ice particle terminal velocities, highlighting aggregates, *J. Atmos. Sci.*, *62*, 1637–1644.
- Okamoto, H., S. Iwasaki, M. Yasui, H. Horie, H. Kuroiwa, and H. Kumagai (2003), An algorithm for retrieval of cloud microphysics using 95-GHz cloud radar and lidar, *J. Geophys. Res.*, *108*(D7), 4226, doi:10.1029/2001JD001225.
- Plana-Fattori, A., A. Protat, and J. Delanoë (2010), Observing ice clouds with a Doppler cloud radar. *C. R. Phys.*, *11*, 96–103.

- Platt, C. M. R. (1973), Lidar and radiometric observations of cirrus clouds, *J. Atmos. Sci.*, *30*, 1191–1204.
- Protat, A., and C. Williams (2011), The accuracy of radar estimates of terminal fall speed from vertically-pointing Doppler cloud radar measurements. *J. Appl. Meteor. Clim.*, *50*, 2120–2138, doi:10.1175/JAMC-D-10-05031.1.
- Protat, A., J. Delanoë, E. O'Connor, and T. L'Ecuyer (2010), The evaluation of CloudSat-derived ice microphysical products using ground-based cloud radar and lidar observations, *J. Atmos. Oceanic Tech.*, *27*, 793–810.
- Protat, A., J. Delanoë, A. Plana-Fattori, P. T. May, and E. O'Connor (2009), The statistical properties of tropical ice clouds generated by the West-African and Australian monsoons from ground-based radar-lidar observations, *Quart. J. Roy. Meteor. Soc.*, *136*(S1), 345–363, 2010, doi:10.1002/qj.490.
- Protat, A., A. Armstrong, M. Haeffelin, Y. Morille, J. Pelon, J. Delanoë, and D. Bouniol (2006), Impact of conditional sampling and instrumental limitations on the statistics of cloud properties derived from cloud radar and lidar at SIRTA, *Geophys. Res. Lett.*, *33*, L11805, doi:10.1029/2005GL025340.
- Protat, A., J. Delanoë, D. Bouniol, A. J. Heymsfield, A. Bansemmer, and P. R. A. Brown (2007), Evaluation of ice water content retrievals from cloud radar reflectivity and temperature using a large airborne in-situ microphysical database, *J. Appl. Meteorol.*, *46*(5), 557–572.
- Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, 238 pp., World Sci., Hackensack, N. J.
- Sakai, T., T. Nagai, M. Nakazato, and T. Matsumura (2003), Ice clouds and Asian dust studied with lidar measurements of particle extinction-to-backscatter ratio, particle depolarization, and water vapor mixing ratio over Tsukuba, *Appl. Opt.*, *42*, 7103–7116.
- Sassen, K., and B. S. Cho (1992), Subvisual-thin cirrus lidar dataset for satellite verification and climatological research, *J. Appl. Meteor.*, *31*, 1275–1285.
- Shupe, M. D. (2007), A ground-based multisensor cloud phase classifier, *Geophys. Res. Lett.*, *34*, L22809, doi:10.1029/2007GL031008.
- Stephens, G. L., S.-C., Tsay, P. W. Stackhouse, and P. J. Flatau (1990), The relevance of the microphysical and radiative properties of cirrus clouds to climate and climatic feedback, *J. Atmos. Sci.*, *47*, 1742–1754.
- Stith, J. L., J. E. Dye, A. Bansemmer, A. J. Heymsfield, C. A. Grainger, W. A. Petersen, and R. Cifelli (2002), Microphysical observations of tropical clouds, *J. Appl. Meteor.*, *41*, 97–117.
- Tinel, C., J. Testud, J. Pelon, R. J. Hogan, A. Protat, J. Delanoë, and D. Bouniol (2005), The retrieval of ice-cloud properties from cloud radar and lidar synergy, *J. Appl. Meteor.*, *44*, 860–875.
- Troyan, D. (2010), *Merged Sounding Value-Added Product*, DOE/SC-ARM/TR-087, 8 pp., Dept. of Energy, Washington, D. C. (Available at <http://www.arm.gov/publications/tech-reports>)
- Turner, D. D. (2005), Arctic mixed-phase cloud properties from AERI lidar observations: Algorithm and results from SHEBA, *J. Appl. Meteor.*, *44*, 427–444.
- Turner, D. D., et al. (2007), Thin liquid water clouds: Their importance and our challenge, *Bull. Amer. Meteor. Soc.*, *88*, 177–190.
- Vogelmann, A. M., and T. P. Ackerman (1995), Relating cirrus cloud properties to observed fluxes: A critical assessment, *J. Atmos. Sci.*, *52*, 4285–4301.
- Wang, Z., and K. Sassen (2001), Cloud type and macrophysical property retrieval using multiple remote sensors, *J. Appl. Meteor.*, *40*, 1665–1682.
- Wang, Z., and K. Sassen (2002), Cirrus cloud microphysical property retrieval using lidar and radar measurements. Part I: Algorithm description and comparison with in situ data, *J. Appl. Meteor.*, *41*, 218–229.
- Yang, P., K. Wyser, and D. Mitchell (2000), Parameterization of the scattering and absorption properties of individual ice crystals, *J. Geophys. Res.*, *105*, 4699–4718.