Methods for Estimating 2D Cloud Size Distributions from 1D Observations

DAVID M. ROMPS

Department of Earth and Planetary Science, University of California, Berkeley, and Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

ANDREW M. VOGELMANN

Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, New York

(Manuscript received 5 April 2017, in final form 6 June 2017)

ABSTRACT

The two-dimensional (2D) size distribution of clouds in the horizontal plane plays a central role in the calculation of cloud cover, cloud radiative forcing, convective entrainment rates, and the likelihood of precipitation. Here, a simple method is proposed for calculating the area-weighted mean cloud size and for approximating the 2D size distribution from the 1D cloud-chord lengths measured by aircraft and vertically pointing lidar and radar. This simple method (which is exact for square clouds) compares favorably against the inverse Abel transform (which is exact for circular clouds) in the context of theoretical size distributions. Both methods also perform well when used to predict the size distribution of real clouds from a Landsat scene. When applied to a large number of Landsat scenes, the simple method is able to accurately estimate the mean cloud size. As a demonstration, the methods are applied to aircraft measurements of shallow cumuli during the Routine ARM Aerial Facility (AAF) Clouds with Low Optical Water Depths (CLOWD) Optical Radiative Observations (RACORO) campaign, which then allow for an estimate of the true area-weighted mean cloud size.

1. Introduction

For a given patch of sky, the distribution of horizontal cloud sizes plays an important role in setting the total cloud cover (e.g., Koren et al. 2008), the cloud radiative forcing (e.g., Marshak and Davis 2005), convective entrainment rates (e.g., Stirling and Stratton 2012; Neggers 2015), and the likelihood of precipitation (e.g., Jiang et al. 2010). Despite the importance of the cloud size distribution, it is not often measured directly. Instead, during field campaigns and at meteorological stations, cloud sizes are usually inferred indirectly from linear sampling by aircraft, radar, lidar, or radiometer. Unfortunately, the distribution of cloud-chord lengths measured in this way is not the same as the distribution of cloud sizes. This mismatch is primarily caused by two effects: 1) an off-center sampling of a cloud will tend to yield a chord that is smaller than the true diameter, biasing the distribution to smaller sizes, and 2) large clouds are more likely to be sampled than small clouds, biasing the distribution to larger sizes. The question addressed here is how to map from the observed distribution of cloud-chord lengths to the actual distribution of cloud sizes.

In this discussion, "linear sampling" refers to one of the following: either that the clouds at a particular height are sampled along a line, or that some vertically integrated cloud indicator is sampled along a line. For example: an aircraft records the presence of cloudy air along a line at its flight altitude; at each height, a vertically pointing lidar or radar records the presence of cloudy air along a line parallel to the wind direction; and a vertically pointing microwave radiometer records a vertically integrated indicator of cloudy air along a line parallel to the mean wind direction within the relevant cloud layer. Other examples include nadir-pointing spaceborne radar and lidar, which sample clouds at a given height along a line parallel to the satellite's orbit. As depicted in Fig. 1, these measurements give us a set of lengths L_i of sample cloud chords, with i ranging from one to the total number of cloud chords that were sampled. If we bin these L_i and normalize appropriately, we get the distribution of detected cloud chords $P(L)$ (m^{-2}), where $P(L)dL$ is

Corresponding author: David M. Romps, romps@berkeley.edu

DOI: 10.1175/JAS-D-17-0105.1

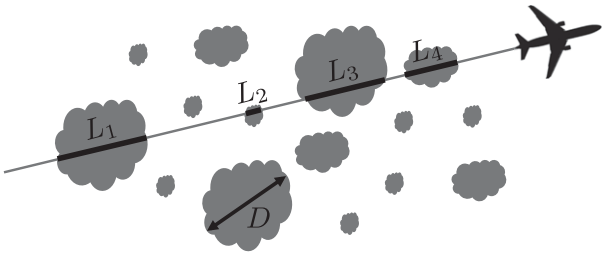

FIG. 1. An airplane, on a linear flight path, will sample cloud-chord lengths L_i, with i ranging from 1 to the total number of clouds sampled. Binning those L_i and normalizing appropriately gives the distribution of chord lengths $P(L)$, where $P(L)dL$ is the probability per distance traveled of detecting a chord between L and $L + dL$. The question addressed in this paper is how to convert from $P(L)$ to the cloud size distribution $n(D)$, where $n(D)dD$ is the number concentration (i.e., number per horizontal area) of clouds with linear sizes between D and $D + dD$.

the probability per distance sampled of detecting a chord between L and $L + dL$.

As mentioned above, the cloud-chord distribution P is not the same as the cloud size distribution, which we will denote by n (m^{-3}). Defining D as some measure of the linear size of a cloud (e.g., the square root of its horizontal area), $n(D)dD$ is the number of clouds per horizontal area that have sizes between D and $D + dD$. The fact that P and n are different distributions is readily apparent from the fact that they have different units. The central question of this paper, therefore, is how to map from P to n.

Some previous authors have solved this problem by assuming that clouds are circular, which allows the relationship between P and n to be written down in terms of the Abel transform and its inverse (Vul'fson 1964; Yau and Rogers 1984). That method, however, is a bit cumbersome and of questionable relevance. It is cumbersome because it expresses n as an integro-differential function of P with a divergent integrand, which can be difficult to evaluate when working with sparse data. And its relevance is questionable because clouds are not circular; for real, noncircular clouds, the Abel transform gives only an approximation of n. Since this method is only approximate, it is worth asking whether there are other approximate methods that are simpler to implement but that perform nearly as well. The next section presents such a method.

For the reader in a rush to calculate cloud sizes, there are four key equations. The first is Eq. (5), which states the obvious by saying that the best estimate of the cloud fractional area is equal to the sum of the cloud-chord lengths divided by the total sampling distance. The second is Eq. (11), which gives an expression for the area-weighted mean cloud size. Finally, there are two methods for calculating the distribution of cloud sizes from the distribution of measured cloud chords. The first is the simple method given in Eq. (8), which is easy to use, is guaranteed to give nonnegative number concentrations, and does a decent job of approximating the true distribution. The second is the Abel-transform method given in Eq. (14), which is more cumbersome to use, is potentially prone to generating negative number concentrations, but is typically better at approximating the true distribution.

2. Estimation methods

For simplicity in this section, let us assume that clouds are convex in the x–y plane so that any linear transect of the cloud is a single line segment. Let us characterize clouds by their maximum width D transverse to the sampling direction. Let us denote by $n(D)dD$ the number density (m^{-2}) of clouds with transverse widths in $[D, D + dD]$. Note that $n(D)$ has units of per cubic meter. For a cloud of width D, let us say that we "sample" it the moment our detector passes through a chosen transverse line. For example, for a circular cloud, we could choose its transverse diameter, and, for a square cloud oriented with sides parallel and perpendicular to the sampling direction, we could choose its transverse midline. A cloud of transverse width D presents a cross section of D to the sampling instrument, so the probability of sampling a cloud of width in $[D, D + dD]$ in a distance dx is the number density $n(D)dD$ times the cross section D times the sampling distance dx—that is, $n(D)dD\,D\,dx$. Assuming that our instrument is sampling a cloud of width D, let us denote by $p(L|D)dL$ the probability that it samples a cloud-chord length in $[L, L + dL]$. The probability density of sampling a cloud length L is

$$P(L) = \int_0^\infty dD\, n(D) D\, p(L|D). \tag{1}$$

Here, $P(L)dL$ is the probability per sampling distance of detecting a cloud chord with a length between L and $L + dL$.

Now, consider the cloud fractional area σ_0, which is the fraction of the horizontal area (either at a given height or projected) that is occupied by clouds. The value of σ_0 can be obtained directly from either n or P. If we know the mean area of clouds with width D, which we will denote by $A(D)$, then we can define the cloud fractional-area distribution $\sigma(D)$ as

$$\sigma(D) = n(D)A(D). \tag{2}$$

Then, σ_0 is given by

$$\sigma_0 = \int_0^\infty dD\, \sigma(D). \tag{3}$$

Here, $\sigma(D)dD$ is the fraction of area covered by clouds with widths between D and $D + dD$. Alternatively, we can obtain σ_0 from $P(L)$ through the relation

$$\sigma_0 = \int_0^\infty dL\, P(L)L. \tag{4}$$

Given the exact $\sigma(D)$ and $P(L)$, both of these equations give the exact value of σ_0. Of course, the simplest way to estimate σ_0 is by direct use of the observed chord lengths L_i using

$$\sigma_0 = \frac{\sum_{i=1}^{N} L_i}{\text{Total sampling distance}}, \tag{5}$$

where N is the total number of cloud chords that were sampled. These relations are all fairly straightforward, but the real challenge is figuring out how to calculate $n(D)$ from $P(L)$. For circular clouds and for square clouds, this problem can be solved exactly.

a. Square clouds (simple method)

Let clouds be nonoverlapping, square, and with sides oriented parallel to the sampling direction. For these square clouds,

$$p(L|D) = \delta(L - D), \tag{6}$$

where δ is the Dirac delta function. Then, by Eq. (1),

$$P(L) = Ln(L). \tag{7}$$

This is trivial to invert to get the expression for n in terms of P,

$$n(D) = n_{\text{simple}}(D) \equiv \frac{P(D)}{D}. \tag{8}$$

For these square clouds, n_{simple} is exactly equal to n; there was no off-center sampling bias to begin with in $P(L)$, and the large-cloud sampling bias has been corrected by the division of P by D. Since $A(D) = D^2$, we can combine Eqs. (2) and (8) to get

$$\sigma_{\text{simple}}(D) = P(D)D, \tag{9}$$

which is an expression that has been used by Rodts et al. (2003) and Berg and Kassianov (2008). This simple method for approximating n and σ is easily applied to any chord-length distribution P, including those constructed from sparse measurements.

Another benefit of this simple method is that the area-weighted mean cloud size can be calculated directly from the set of observed chord lengths. The area-weighted mean cloud size can be written as the integral over D of $\sigma_{\text{simple}}(D)D$ divided by the integral over D of $\sigma_{\text{simple}}(D)$, which, by Eq. (9), is equal to the integral over L of $P(L)L^2$ divided by the integral over L of $P(L)L$. Therefore, the simple-method estimate of the area-weighted mean cloud size $\langle D\rangle_{\text{simple}}$ can be calculated directly from the observed chord lengths L_i as

$$\langle D\rangle_{\text{simple}} = \frac{\sum_{i=1}^{N} L_i^2}{\sum_{i=1}^{N} L_i}, \tag{10}$$

where N is the total number of sampled cloud chords. As we will see in sections 4 and 5, this estimate of the area-weighted mean cloud size is biased low by ~40% for real clouds (which are not sampling-aligned squares), so we can attempt to correct this bias by writing

$$\langle D\rangle = k\frac{\sum_{i=1}^{N} L_i^2}{\sum_{i=1}^{N} L_i}, \tag{11}$$

where $k \approx 1.7 \pm 0.3$ is a correction factor (see section 5 for an explanation of this factor).

b. Circular clouds (Abel method)

Let clouds be nonoverlapping and circular. For these circular clouds,

$$p(L|D) = \frac{L}{D\sqrt{D^2 - L^2}}\mathscr{H}(L)\mathscr{H}(D - L), \tag{12}$$

where $\mathscr{H}$ is the Heaviside unit step function. Then, by Eq. (1),

$$P(L) = L\mathscr{H}(L)\int_L^\infty dD\frac{n(D)}{\sqrt{D^2 - L^2}}. \tag{13}$$

This is the Abel transform. The inverse Abel transform then gives us the expression for n in terms of P:

$$n(D) = n_{\text{Abel}}(D) \equiv -\frac{2D}{\pi}\mathscr{H}(D)\int_D^\infty dL\frac{\frac{d}{dL}\left[\frac{P(L)}{L}\right]}{\sqrt{L^2 - D^2}}. \tag{14}$$

For these circular clouds, n_{Abel} is exactly equal to n; both the off-center sampling bias and the large-cloud

sampling bias have been corrected by the inverse Abel transform. Since $A(D) = \pi D^2/4$, we can combine Eqs. (2) and (14) to get

$$\sigma_{\text{Abel}}(D) = -\frac{D^3}{2}\mathscr{H}(D)\int_D^\infty dL\,\frac{\dfrac{d}{dL}\left[\dfrac{P(L)}{L}\right]}{\sqrt{L^2 - D^2}}. \quad (15)$$

Because of the derivative and the divergent integrand, this method can be difficult to apply to sparse observational data. It can even predict negative values for n and σ in cases where $P(L)/L$ increases with L. This can occur either because the sizes, shapes, and orientations of the clouds guarantee that particular failure of the Abel method, or because spurious fluctuations in $P(L)/L$ caused by insufficient sampling lead to negative values in the numerical calculation of n_{Abel}.

3. Evaluation against theoretical distributions

As a first test of these methods, we will evaluate them against analytical size distributions. Consider a field of clouds with a power-law size distribution of the form

$$n(D) = \left(\frac{n_0}{D_0}\right)\left(\frac{D}{D_0}\right)^b \mathscr{H}(D)\mathscr{H}(D_0 - D), \quad (16)$$

where n_0 (m^{-2}), D_0 (m), and b (unitless) are constants. In order for this distribution to give a finite cloud fractional area σ_0, b must be greater than -3, which we will assume henceforth.

a. Power-law squares

Consider clouds that are nonoverlapping, square, and oriented with edges exactly parallel or perpendicular to the sampling direction. This gives them the $p(L|D)$ given by Eq. (6). Let us also assume that their widths D are distributed according to the power-law relation in Eq. (16).

For square clouds, the area per cloud is $A(D) = D^2$, so Eq. (3) tells us that the cloud fractional area σ_0 and n_0 are related by

$$\sigma_0 = \frac{1}{b+3} n_0 D_0^2. \quad (17)$$

Note that n_0 is not the number density here. The number density of clouds is given by $\int_0^\infty dD\, n(D)$, which is infinite for $-3 < b < -1$ even though n_0 and σ_0 are finite for all $b > -3$. Using Eqs. (16), (7), and (17), we find n and P to be

$$n(D) = \frac{(b+3)\sigma_0}{D_0^3}\left(\frac{D}{D_0}\right)^b \mathscr{H}(D)\mathscr{H}(D_0 - D) \quad \text{and} \quad (18)$$

$$P(L) = \frac{(b+3)\sigma_0}{D_0^2}\left(\frac{L}{D_0}\right)^{b+1} \mathscr{H}(L)\mathscr{H}(D_0 - L). \quad (19)$$

A sample field of such clouds, with $D_0 = 1$ km, $\sigma_0 = 0.1$, and $b = -2$, is shown in Fig. 2a. Since the finite resolution of the image precludes the plotting of infinitely small clouds, clouds with sizes less than 50 m are omitted. Figure 2b plots $n(D)$ and $P(L)$ in gray and blue, respectively.

What if, in the real world, our aircraft or vertically pointing instrument measures $P(L)$ of the form given by Eq. (19)? How would we reconstruct $n(D)$? From the previous section, we know we have two approaches: we can pretend the clouds are square (i.e., use the simple method) or we can pretend the clouds are circular (i.e., use the Abel method).

The simple method for estimating n from P, as given by Eq. (8), was derived from square clouds, so it should be exact in this case of square clouds. Indeed, the red curve in Fig. 2b, which plots Eq. (8) applied to the $P(L)$ from Eq. (19), matches n (the gray curve) exactly. The Abel method, on the other hand, is exact only for circular clouds, so it should only represent n approximately in this case. Indeed, the green curve in Fig. 2b, which plots Eq. (14) applied to the $P(L)$ from Eq. (19), deviates slightly from n (the gray curve) at the largest sizes.

b. Power-law circles

Now, consider clouds that are nonoverlapping and circular. This gives them the $p(L|D)$ given by Eq. (12). Let us also assume that their diameters D are distributed according to the power-law relation in Eq. (16).

For circular clouds, the area per cloud is $A(D) = \pi D^2/4$, so Eq. (3) tells us that the cloud fractional area σ_0 and n_0 are related by

$$\sigma_0 = \frac{\pi}{4}\frac{1}{b+3} n_0 D_0^2. \quad (20)$$

Using Eqs. (16), (13), and (20), we find n and P to be

$$n(D) = \frac{4(b+3)\sigma_0}{\pi D_0^3}\left(\frac{D}{D_0}\right)^b \mathscr{H}(D)\mathscr{H}(D_0 - D) \quad \text{and} \quad (21)$$

$$P(L) = \frac{2(b+3)\sigma_0}{\pi D_0^2}\left(\frac{L}{D_0}\right)^{b+1}\left[\sqrt{\pi}\,\frac{\Gamma(-b/2)}{\Gamma(1/2 - b/2)} - B_{L^2/D_0^2}(-b/2, 1/2)\right]\mathscr{H}(L)\mathscr{H}(D_0 - L), \quad (22)$$

where Γ is the Euler gamma function and B is the incomplete beta function. A sample field of such clouds, with $D_0 = 1$ km, $\sigma_0 = 0.1$, and $b = -2$, is shown in Fig. 3a. Again, clouds with sizes less than 50 m are omitted.

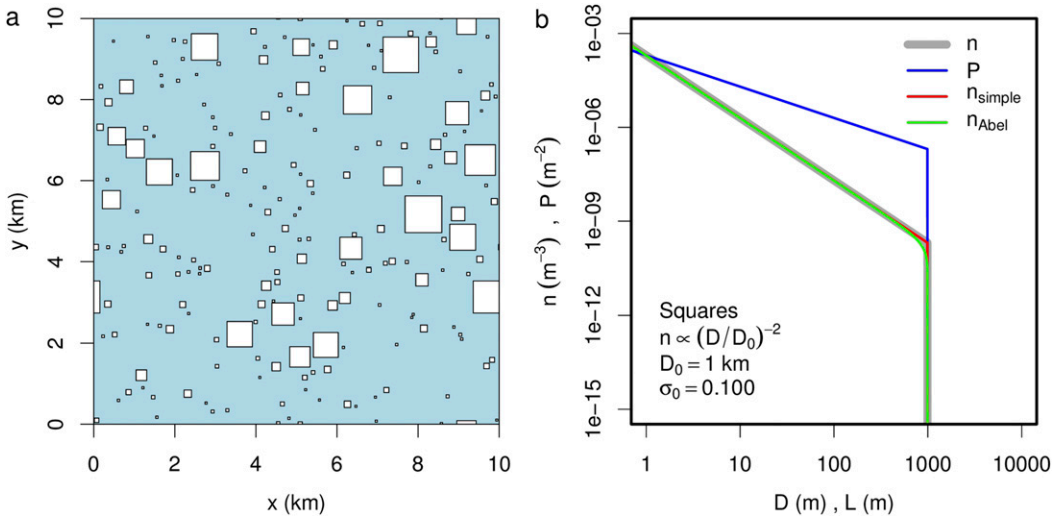

FIG. 2. (a) A sample field of square clouds generated from the n given by Eq. (18) for the case of $D_0 = 1$ km, $\sigma_0 = 0.1$, and $b = -2$. (b) For this case, plots of $n(D)$ from Eq. (18) (gray), $P(L)$ from Eq. (19) (blue), n_{simple} as calculated from Eq. (8) applied to the $P(L)$ from Eq. (19) (red), and n_{Abel} as calculated from Eq. (14) applied to the $P(L)$ from Eq. (19) (green).

Figure 3b plots $n(D)$ and $P(L)$ in gray and blue, respectively.

What if, in the real world, our aircraft or vertically pointing instrument measures a $P(L)$ of the form given by Eq. (22)? How would we reconstruct $n(D)$? From the previous section, we know we have two approaches: we can pretend the clouds are square (i.e., use the simple method) or we can pretend the clouds are circular (i.e., use the Abel method).

The Abel method for estimating n from P, as given by Eq. (14), was derived from circular clouds, so it should be exact in this case of circular clouds. Indeed, the green curve in Fig. 3b, which plots Eq. (14) applied to the $P(L)$ from Eq. (19), matches n (the gray curve) exactly. The simple method, on the other hand, is exact only for square clouds, so it should only represent n approximately in this case. Indeed, the red curve in Fig. 3b, which plots Eq. (8) applied to the $P(L)$ from

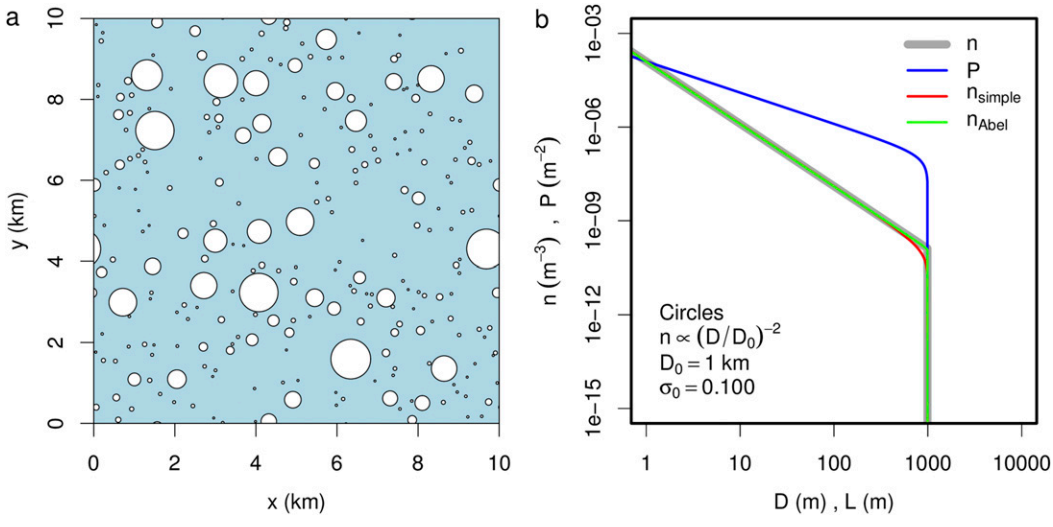

FIG. 3. (a) A sample field of circular clouds generated from the n given by Eq. (21) for the case of $D_0 = 1$ km, $\sigma_0 = 0.1$, and $b = -2$. (b) For this case, plots of $n(D)$ from Eq. (21) (gray), $P(L)$ from Eq. (22) (blue), n_{simple} calculated from Eq. (8) applied to the $P(L)$ from Eq. (22) (red), and n_{Abel} as calculated from Eq. (14) applied to the $P(L)$ from Eq. (22) (green).

Eq. (22), deviates slightly from n (the gray curve) at the largest sizes. Overall, though, it is remarkable how well both methods work for both square and circular clouds.

To quantify the error in the simple method in this case of circular clouds, we can compare the area-weighted mean size calculated with n_{simple} with the area-weighted mean size calculated with n. The area-weighted mean cloud size $\langle D \rangle$ is given by

$$\langle D \rangle = \frac{\int_0^\infty dD\, \sigma(D) D}{\int_0^\infty dD\, \sigma(D)}. \tag{23}$$

Similarly, we can define $\langle D \rangle_{\text{Abel}}$ and $\langle D \rangle_{\text{simple}}$ by equivalent equations using σ_{Abel} and σ_{simple} from Eqs. (15) and (9), respectively. In this case of circular clouds, σ_{Abel} is the same as the true σ, and they give an area-weighted mean size of

$$\langle D \rangle_{\text{Abel}} = \langle D \rangle = \frac{b+3}{b+4} D_0. \tag{24}$$

The simple method, on the other hand, uses Eq. (22) in Eq. (9) to get σ_{simple}, and then uses that expression in Eq. (23), which yields

$$\langle D \rangle_{\text{simple}} = \frac{8}{3\pi} \frac{b+3}{b+4} D_0. \tag{25}$$

Regardless of b, the ratio of these two estimates is $8/3\pi \approx 0.85$, a fact noted previously by Rodts et al. (2003). Therefore, the simple method gives an area-weighted mean size for circular clouds that is 15% smaller than that given by the Abel method. This is very close to the difference between these two methods when applied to Landsat imagery and aircraft data, as we will see in the next sections.

4. Evaluation against a Landsat scene

In reality, clouds are neither squares nor circles. To evaluate these inversion methods in a realistic cloud scene, we will use a Landsat image of maritime cumuli, which is shown (in negative, to save printer ink) in Fig. 4a. This image (Landsat scene identifier LC80100432013364LGN00) is about 200 km on a side with a resolution of 30 m. It was acquired at 1521 UTC 30 December 2013 over the Atlantic Ocean, about 400 km east of the Bahamas.

Figure 4 shows the image processing to which this scene is subjected. The original image (Fig. 4a) is rotated and cropped (Fig. 4b). Figure 4c zooms into one patch of the image. The scene is then filtered to set the ocean to white (Fig. 4d) and then to set all the clouds to black (Fig. 4e). Individual clouds are identified as connected patches of black; these individual clouds are colored in Fig. 4f. The D for each cloud is defined as the square root of its area. Although Figs. 4c–f show only a subset of the image, which is replicated in Fig. 5a, the entire scene is used in the calculation of P and n.

Since the scene is dominated by small clouds, the cloud sizes are sorted into nine bins that are logarithmically spaced. We then obtain $n(D)$ by normalizing the counts in each bin by the width of the bin and the total area of the Landsat scene. This is plotted as the gray curve in Fig. 5b. To calculate $P(L)$, a synthetic flight path is generated that starts in the upper-left corner of the scene, goes to the upper-right corner, turns around and moves down one row of pixels, flies back to the left edge, turns around and moves down one row, and repeats until it reaches the lower-right corner. This produces one continuous flight path that is roughly 1 000 000 km long. The observed chord lengths are then sorted into nine logarithmically spaced bins. We then obtain $P(L)$ by normalizing the counts in each bin by the width of the bin and the total length of the flight path. This is plotted as the blue curve in Fig. 5b.

How do the Abel method and the simple method perform in this case of real clouds? Since real clouds are neither circular nor square, neither method will be exact, but we might expect, from the outcome of the previous section, that both will do a good job of approximating n. The green curve in Fig. 5b plots n_{Abel}, which is obtained by applying Eq. (14) to P. This is an excellent approximation to the gray curve. The red curve in Fig. 5b plots n_{simple}, which is obtained by applying Eq. (8) to P. This is still a good approximation to the gray curve, although it has a larger bias toward small clouds.

Although plotting $n(D)$ on a log–log plot is standard practice, it is not the best way to convey an intuitive feeling for the cloud size distribution. An alternative is to plot the cloud fractional area distribution $\sigma(D)$ on linear axes. Figure 6a plots σ as measured directly from the 2D analysis, σ_{simple} as obtained from Eq. (9) applied to the 1D $P(L)$, and σ_{Abel} as obtained from Eq. (15) applied to the 1D $P(L)$, all on linear axes. Again, we see that the simple method is biased toward smaller clouds sizes, but the bias is small. We can quantify this bias by calculating the area-weighted mean cloud size $\langle D \rangle$, given by Eq. (23). The gray, red, and green circles on the abscissa of Fig. 6a denote the area-weighted mean cloud sizes $\langle D \rangle$, $\langle D \rangle_{\text{simple}}$,

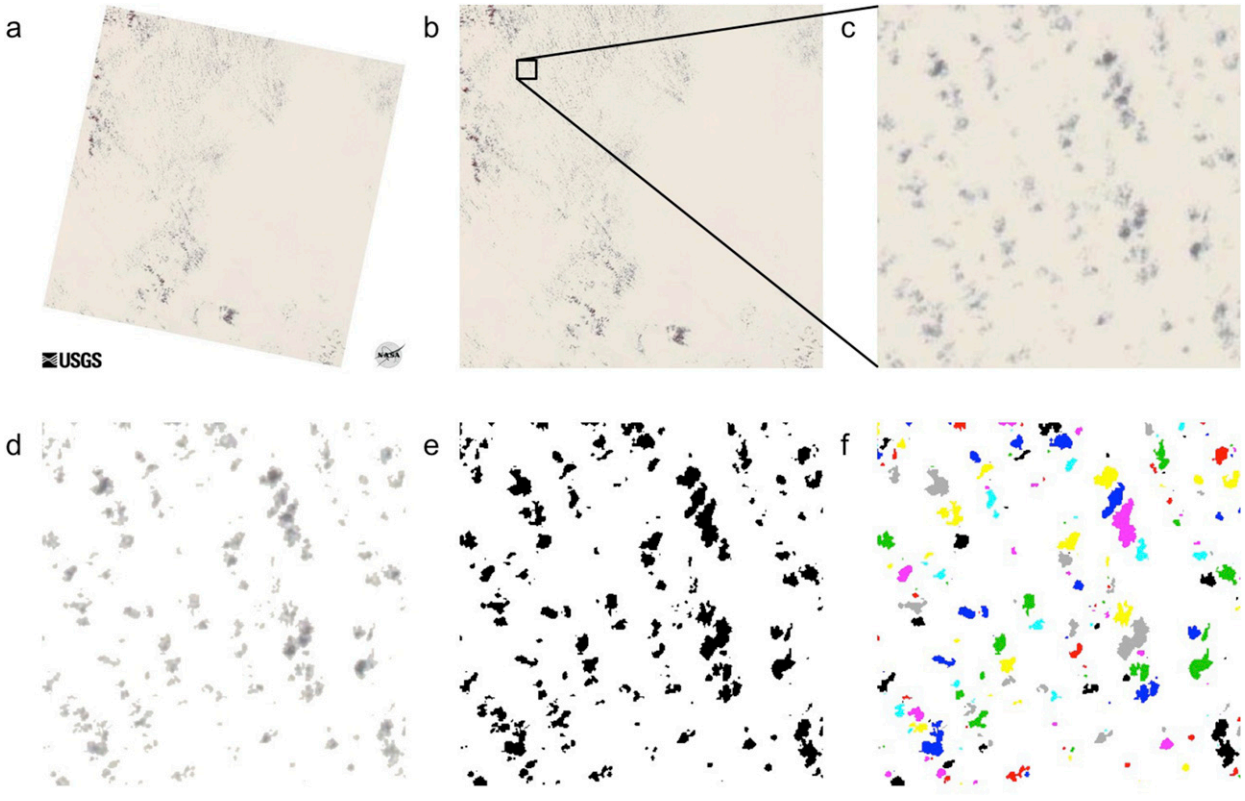

FIG. 4. (a) A 40 000-km^2 Landsat scene of maritime cumuli, in negative, that is then (b) rotated and cropped, (c) zoomed in on a 100-km^2 patch, (d) filtered to make the ocean white, (e) filtered to make the clouds black, and (f) color coded to illustrate individual clouds.

and $\langle D \rangle_{\text{Abel}}$, calculated from σ, σ_{simple}, and σ_{Abel}, respectively. The true area-weighted mean cloud size is 1450 m, and the Abel and simple methods underestimate this by 20% and 37%, respectively. Since $1/(1 - 0.37) = 1.6$, the true area-weighted mean cloud size is bigger than the estimate from the simple method by a factor of 1.6 in this case. Also, note that, in this case of real clouds, the simple method gives an

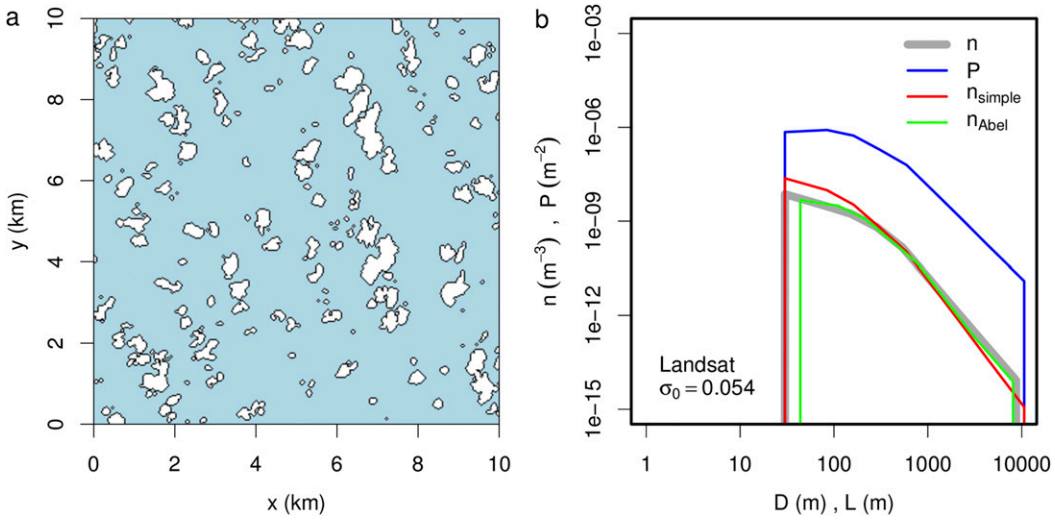

FIG. 5. (a) The same 10 km × 10 km sample of the Landsat scene shown in Figs. 4c–f. (b) Plots of $n(D)$ (gray) and $P(L)$ (blue) calculated directly from the Landsat image, as well as n_{simple} calculated from Eq. (8) applied to $P(L)$ (red) and n_{Abel} calculated from Eq. (14) applied to $P(L)$ (green).

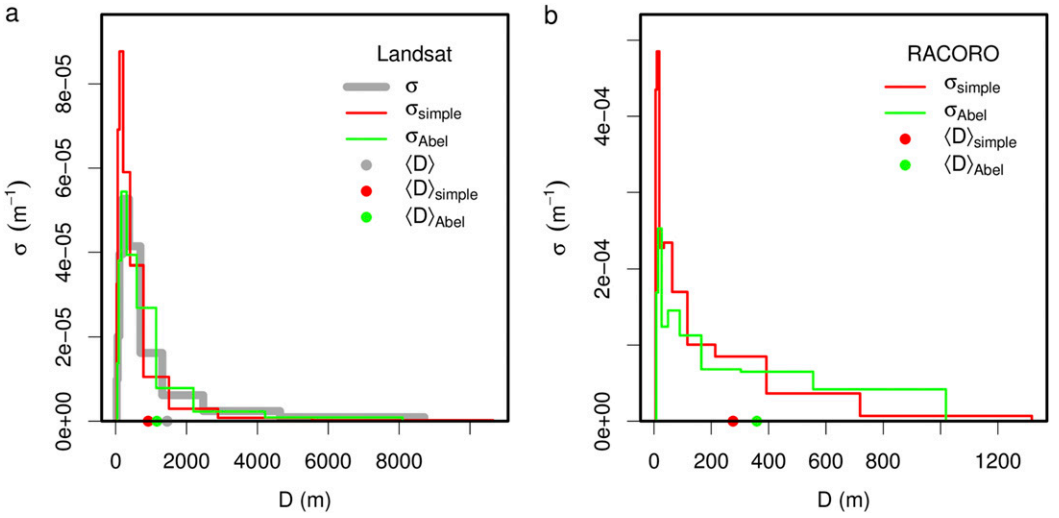

FIG. 6. (a) For the Landsat scene, plots of σ, σ_{simple}, and σ_{Abel} and the values of $\langle D\rangle$, $\langle D\rangle_{\text{simple}}$, and $\langle D\rangle_{\text{Abel}}$ marked on the abscissa. (b) As in (a), but for the RACORO data.

estimate for $\langle D\rangle$ that is 21% smaller than the estimate from the Abel method. This is very similar to the 15% difference found in section 3b for circular clouds.

5. Evaluation against 69 Landsat scenes

For the single Landsat scene analyzed in the previous section, we found that $\langle D\rangle/\langle D\rangle_{\text{simple}}$ equals 1.6. But how constant is this ratio? After all, we can easily imagine some cloud shapes that would lead to large deviations from this value. For clouds with an intensely nonconvex shape—for example, the shape of a comb—the sampled chord lengths would be biased low and the required correction factor would be much greater than 1.6. On the other hand, clouds that are elongated in the sampling direction with a large aspect ratio—for example, a very skinny rectangle or a highly eccentric ellipse—the sampled chord lengths would be biased high and the required correction factor would be much less than 1.6. (See the appendix for a mathematical treatment of rectangles and ellipses.) Fortunately, most convective clouds do not possess such exaggerated convexity or elongation. We might guess, therefore, that most naturally occurring convective clouds exhibit a similar $\langle D\rangle/\langle D\rangle_{\text{simple}}$ ratio.

To test this, we search for suitable *Landsat-8* scenes within Worldwide Reference System (WRS) rows 5–10 and WRS paths 38–43, which encompass the patch of the western North Atlantic shown in Fig. 7. Of the 229 "L8 OLI/TIRS C1 Level-1" scenes (L8: *Landsat-8*; OLI: Operational Land Imager; TIRS: Thermal Infrared Sensor; C1: Collection 1), 160 are deemed unsatisfactory owing to one or more of the following conditions: having elongated and wispy midtropospheric or upper-tropospheric clouds, being dominated by clouds comparable in size to the width of the scene, having an island in the scene that would interfere with the cloud-identification algorithm, or having solar reflection off the ocean surface that would interfere with the cloud-identification algorithm. The remaining 69 scenes are listed in Table 1. Each image is processed with

FIG. 7. The patch of the western North Atlantic encompassing WRS rows 5–10 and WRS paths 38–43. This is the region where the 69 *Landsat-8* scenes were collected.

TABLE 1. The 69 Landsat scenes used in section 5. The analysis used images from the combined OLI/TIRS product of *Landsat-8* acquired at the listed WRS path and row on the listed year, month, and day.

Path	Row	Year	Month	Day
005	038	2013	09	22
005	038	2015	07	10
005	038	2015	11	15
005	038	2016	02	03
005	039	2013	09	22
005	039	2013	12	11
005	039	2015	06	24
005	039	2015	08	27
005	039	2016	02	03
005	040	2016	02	03
005	041	2015	05	23
005	041	2015	07	26
005	041	2015	10	30
005	041	2015	12	17
005	041	2016	01	02
005	041	2016	02	03
006	039	2015	11	06
006	039	2016	01	09
006	039	2016	04	14
006	040	2015	07	01
006	040	2015	08	02
006	040	2015	10	05
006	040	2015	11	06
006	040	2016	04	14
006	041	2015	08	02
006	041	2015	11	06
006	041	2015	11	22
006	041	2016	04	14
007	038	2013	08	19
007	038	2013	12	09
007	038	2015	09	10
007	039	2013	08	19
007	039	2013	10	22
007	039	2015	07	08
007	039	2015	10	28
007	039	2015	11	13
007	039	2015	12	31
007	040	2015	07	08
007	040	2015	10	28
007	040	2015	11	13
007	040	2015	12	31
007	041	2015	10	12
007	041	2015	10	28
007	041	2015	11	13
007	041	2015	12	31
008	038	2015	05	12
008	038	2015	06	29
008	038	2015	07	31
008	039	2015	07	31
008	039	2016	02	08
008	039	2016	04	12
008	040	2015	07	31
008	040	2015	08	16
008	040	2015	09	17
008	040	2015	12	22
008	040	2016	02	08
008	040	2016	03	11
008	040	2016	04	12
008	041	2015	07	31
008	041	2015	08	16
008	041	2015	09	17
008	041	2015	11	04
008	041	2016	03	11
010	043	2013	07	23
010	043	2013	08	24
010	043	2013	09	09
010	043	2013	09	25
010	043	2013	11	12
010	043	2013	12	30

the methods described in the previous section, which generate a binary cloud–ocean image similar to that depicted in Fig. 4e. Figure 8 shows a low-resolution thumbnail of these binary images for each of the 69 Landsat scenes.

As in the previous section, we identify individual clouds as connected regions of black in the binary images. For each image, this allows us to calculate the 2D area of each cloud. From those areas, we calculate the true $\langle D \rangle$ for each of the 69 scenes. Sampling each binary image with linear transects as in the previous section, we also generate a set of 1D cloud-chord lengths. From those lengths, we calculate $\langle D \rangle_{\text{simple}}$ from Eq. (11) for each of the 69 scenes. Figure 9 plots the 69 actual $\langle D \rangle$ against the corresponding $k\langle D \rangle_{\text{simple}}$ using a single best-fit value of k.

The agreement in Fig. 9 is striking. Despite a 40-fold range in sizes (from 500 m to 20 km), $k\langle D \rangle_{\text{simple}}$ with a single value of k does an excellent job of estimating the true area-weighted mean cloud size. By calculating the mean and standard deviation of the 69 values of $\langle D \rangle / \langle D \rangle_{\text{simple}}$, we find that $k = 1.7 \pm 0.3$.

6. Application to RACORO

The real power of these methods lies in their ability to calculate the size distribution n in cases where n is not directly observable. Although there is no benchmark size distribution to compare against in these cases, we can perform a sanity test by checking that n_{simple} and n_{Abel} give similar results.

For demonstration, we use here cloud data collected by an aircraft during the Routine ARM Aerial Facility (AAF) Clouds with Low Optical Water Depths (CLOWD) Optical Radiative Observations

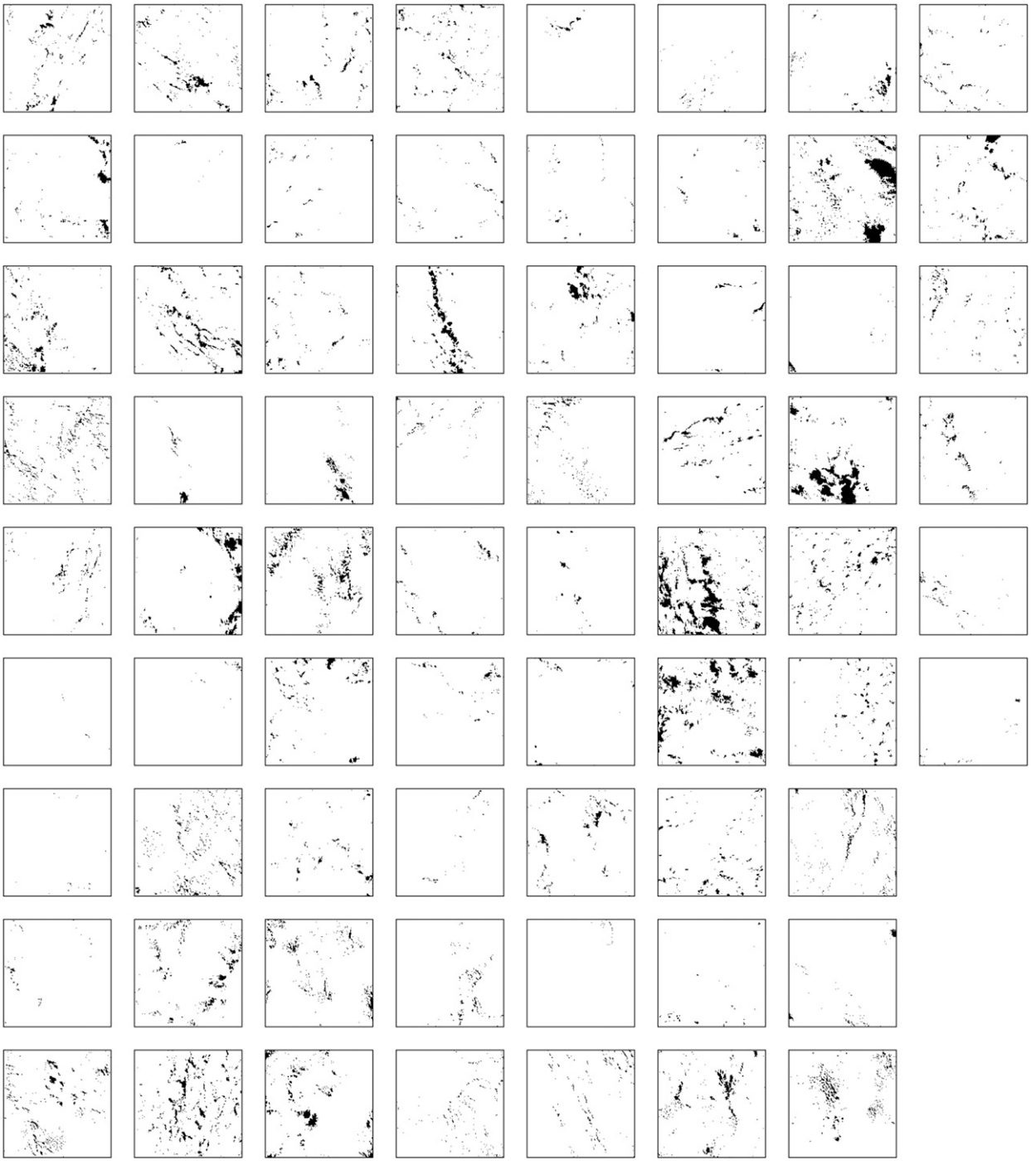

FIG. 8. Low-resolution (40 000 pixels) thumbnails of the 69 Landsat scenes processed into binary images, with clouds shown in black. Note that the analysis uses the full-resolution images, each of which has about 40 000 000 pixels.

(RACORO) campaign (Vogelmann et al. 2012), which took place over Oklahoma from 22 January to 30 June 2009. For our purposes, we focus on three consecutive days (22, 23, and 24 May) that had similar shallow cumuli. During those three days, the aircraft flew a total of 937 km of straight legs within the cumulus layers. The mixing ratio of condensate was measured at 10 Hz, which, at an airspeed of 60 m s^{-1}, gives a sampling

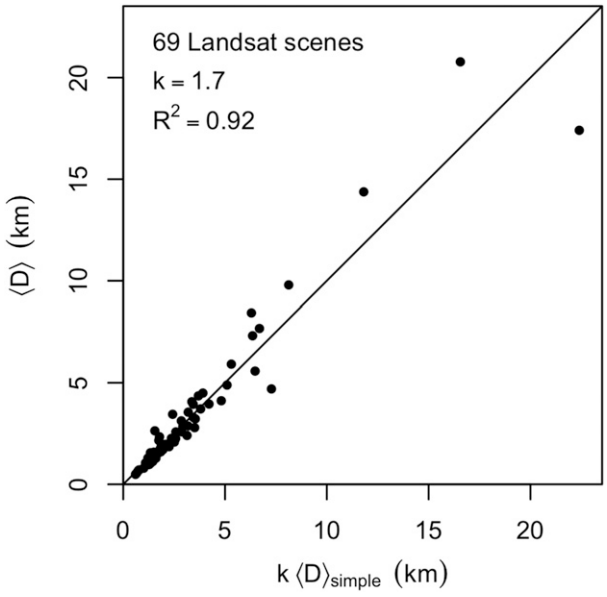

FIG. 9. Scatterplot of the actual area-weighted mean D for each of the 69 Landsat scenes vs the simple estimate from Eq. (11) using $k = 1.7$. The solid curve marks the one-to-one line.

FIG. 10. The clouds as seen on 24 May from the RACORO aircraft during a flight above the cloud layer.

distance of 6 m. The presence or absence of cloud was defined by a mixing ratio threshold of 10^{-5}. The estimate of the cloud fraction σ_0, as obtained from the fraction of distance the aircraft spent in cloud, is 0.078. Note that this is the mean cloud fraction for a single altitude within the cloud layer; the vertically projected cloud fraction, as obtained from a nadir-pointing satellite or a zenith-pointing camera, would be larger. Figure 10 shows a photograph of the cumulus layer taken from the aircraft on 24 May during a flight leg that was above the cloud layer (i.e., a flight leg not used in the calculation of P).

By binning the observed cloud-chord lengths and normalizing as before, we obtain the $P(L)$ shown as the blue curve in Fig. 11b. Using Eqs. (8) and (14) on this observed P generates n_{simple} and n_{Abel}, plotted in Fig. 11b in red and green, respectively. These two estimates for n agree fairly well, with the same tendency for the simple method to be biased toward smaller cloud sizes. To put this into a more intuitive form, Fig. 6b plots σ_{simple} and σ_{Abel} with the area-weighted mean sizes denoted by the circles on the abscissa. In this case, $\langle D\rangle_{\text{simple}} = 284$ m, which is 23% smaller than $\langle D\rangle_{\text{Abel}}$. Note that this is similar to the 15% and 21% differences found in sections 3b and 4, respectively. Using Eq. (11) with $k = 1.7 \pm 0.3$, we can estimate that the true area-weighted mean size of the clouds at the altitudes sampled by the RACORO aircraft was 480 ± 90 m. To illustrate the estimated distribution of cloud sizes sampled by these RACORO flights, Fig. 11a plots a sample field of circular clouds using n_{Abel} with, as before, a lower cutoff size for plotting purposes of 50 m.

7. Summary

Given a set of cloud-chord lengths L_i observed from an unbiased straight-line sampling strategy, the cloud cover is given by Eq. (5) and the area-weighted mean cloud size is given by Eq. (11). To calculate the distribution of cloud sizes $n(D)$, one must first bin the observed chord lengths and normalize the counts by the bin widths and total sampling distance to generate $P(L)$. One then has a choice to estimate $n(D)$ using either the simple method in Eq. (8) or the Abel method in Eq. (14); the former is exact in the case of sampling-aligned square clouds, and the latter is exact in the case of circular clouds.

As its name would suggest, the simple method is simple to use. It is also guaranteed to give nonnegative $n(D)$ and it will work on small sets of observed L_i. The Abel method is more complicated to use since it requires differentiating $P(L)$ and integrating a divergent function, and it can generate negative $n(D)$. It is, therefore, not suitable for use on sparse data. For rich datasets, however, the Abel method will tend to give more accurate results, undoubtedly because real cloud shapes are closer to being circles than sampling-aligned squares. Even if the more accurate Abel method is used, it is recommended that it be compared to the simple method to check for errors in the numerics; if everything is working properly, the area-weighted mean cloud

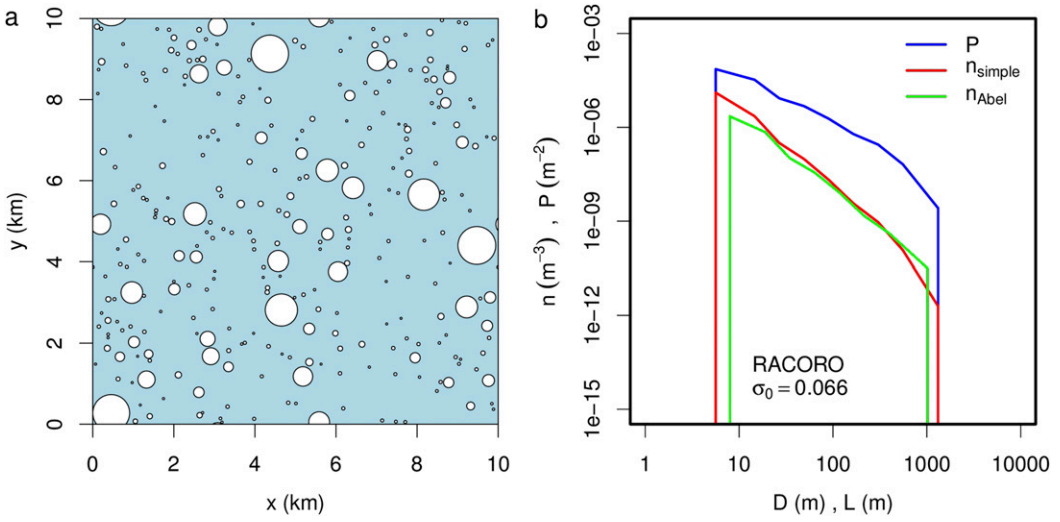

FIG. 11. (a) A sample field of circular clouds generated from the n_{Abel} calculated from Eq. (14) applied to the aircraft-measured $P(L)$. (b) Plots of $P(L)$ (blue) calculated directly from the RACORO aircraft data, as well as n_{simple} calculated from Eq. (8) applied to $P(L)$ (red), and n_{Abel} calculated from Eq. (14) applied to $P(L)$ (green).

sizes predicted by the two methods should differ by only about 20%.

Acknowledgments. This work was supported by the U.S. Department of Energy's Climate Model Development and Validation (CMDV), an Office of Science, Office of Biological and Environmental Research activity, under Contract DE-AC02-05CH11231. Thanks are due to three anonymous reviewers whose suggestions improved the paper.

APPENDIX

Mathematical Treatment of Rectangles and Ellipses

a. Simple method applied to rectangles

Consider a population of sampling-aligned rectangles, all of which have an aspect ratio r^2, whereby their width w (perpendicular to the sampling direction) is r^2 times their length l (parallel to the sampling direction); that is, $w = r^2 l$. For these clouds, the effective size, defined as the square root of the area, is $D = \sqrt{wl} = rl$. Note that $w = rD$ and $l = D/r$.

For these rectangular clouds, Eq. (6) generalizes to

$$p(L|D) = \delta(L - D/r) = r\delta(rL - D).$$

Equation (1) becomes

$$P(L) = \int_0^\infty dD\, n(D)\, rD\, p(L|D).$$

Here, the extra factor of r represents the fact that, compared to a square of the same D, a rectangle of aspect ratio r^2 is a factor of r wider and, therefore, r times more likely to be sampled. Combining these two equations, we get

$$P(L) = r^3 n(rL) L.$$

Therefore, by Eq. (8),

$$n_{\text{simple}}(D) = \frac{P(D)}{D} = r^3 n(rD).$$

Clearly, $n_{\text{simple}}(D)$ does not equal $n(D)$ for $r \neq 1$. The estimated area-weighted mean cloud size will be

$$\langle D \rangle_{\text{simple}} = \frac{\int_0^\infty dD\, n_{\text{simple}}(D) D^3}{\int_0^\infty dD\, n_{\text{simple}}(D) D^2} \tag{A1}$$

$$= \frac{\int_0^\infty dD\, n(rD) D^3}{\int_0^\infty dD\, n(rD) D^2} \tag{A2}$$

$$= \frac{1}{r} \frac{\int_0^\infty dD\, n(D) D^3}{\int_0^\infty dD\, n(D) D^2} \tag{A3}$$

$$= \frac{1}{r} \langle D \rangle, \tag{A4}$$

where the $1/r$ is generated in the second-to-last line by replacing the dummy integration variable D with D/r.

b. Abel method applied to ellipses

Consider a population of sampling-aligned ellipses, all of which have an aspect ratio r^2, whereby their width w (perpendicular to the sampling direction) is r^2 times their length l (parallel to the sampling direction); that is, $w = r^2 l$. Letting y measure distance parallel to the sampling direction, such an ellipse is described by

$$\frac{1}{r^2}x^2 + r^2 y^2 = \frac{D^2}{4}.$$

Here, the effective size D is defined as $D = 2\sqrt{A/\pi}$, where A is the area of the ellipse. Note that $w = rD$ and $l = D/r$.

For these elliptical clouds, Eq. (12) generalizes to

$$p(L|D) = \frac{r^2 L}{D\sqrt{D^2 - r^2 L^2}}\mathcal{H}(L)\mathcal{H}(D/r - L).$$

Equation (1) becomes

$$P(L) = \int_0^\infty dD\, n(D)\, rD\, p(L|D).$$

Here, the extra factor of r represents the fact that, compared to a circle of the same D, an ellipse of aspect ratio r^2 is a factor of r wider and, therefore, r times more likely to be sampled. Combining these two equations, we get

$$P(L) = r^3 L\mathcal{H}(L)\int_{rL}^\infty dD\,\frac{n(D)}{\sqrt{D^2 - r^2L^2}} \tag{A5}$$

$$= L\mathcal{H}(L)\int_L^\infty dD\,\frac{r^3 n(rD)}{\sqrt{D^2 - L^2}}. \tag{A6}$$

Comparing to Eq. (13), we see that the inverse Abel transform will generate

$$n_{\text{Abel}}(D) = r^3 n(rD).$$

Clearly, $n_{\text{Abel}}(D)$ does not equal $n(D)$ for $r \neq 1$. In fact, n_{Abel} is related to the actual n for ellipses in the same way that n_{simple} is related to the actual n for squares. Proceeding as before, we get

$$\langle D\rangle_{\text{Abel}} = \frac{1}{r}\langle D\rangle.$$

REFERENCES

Berg, L. K., and E. I. Kassianov, 2008: Temporal variability of fair-weather cumulus statistics at the ACRF SGP site. *J. Climate*, **21**, 3344–3358, doi:10.1175/2007JCLI2266.1.

Jiang, H., G. Feingold, and A. Sorooshian, 2010: Effect of aerosol on the susceptibility and efficiency of precipitation in warm trade cumulus clouds. *J. Atmos. Sci.*, **67**, 3525–3540, doi:10.1175/2010JAS3484.1.

Koren, I., L. Oreopoulos, G. Feingold, L. A. Remer, and O. Altaratz, 2008: How small is a small cloud? *Atmos. Chem. Phys.*, **8**, 3855–3864, doi:10.5194/acp-8-3855-2008.

Marshak, A., and A. B. Davis, Eds., 2005: *3D Radiative Transfer in Cloudy Atmospheres*. Springer, 686 pp., doi:10.1007/3-540-28519-9.

Neggers, R. A. J., 2015: Attributing the behavior of low-level clouds in large-scale models to subgrid-scale parameterizations. *J. Adv. Model. Earth Syst.*, **7**, 2029–2043, doi:10.1002/2015MS000503.

Rodts, S., P. Duynkerke, and H. Jonker, 2003: Size distributions and dynamical properties of shallow cumulus clouds from aircraft observations and satellite data. *J. Atmos. Sci.*, **60**, 1895–1912, doi:10.1175/1520-0469(2003)060<1895:SDADPO>2.0.CO;2.

Stirling, A. J., and R. A. Stratton, 2012: Entrainment processes in the diurnal cycle of deep convection over land. *Quart. J. Roy. Meteor. Soc.*, **138**, 1135–1149, doi:10.1002/qj.1868.

Vogelmann, A. M., and Coauthors, 2012: RACORO extended-term aircraft observations of boundary layer clouds. *Bull. Amer. Meteor. Soc.*, **93**, 861–878, doi:10.1175/BAMS-D-11-00189.1.

Vul'fson, N. L., 1964: *Convective Motions in a Free Atmosphere*. Israel Program for Scientific Translation, 188 pp.

Yau, M. K., and R. R. Rogers, 1984: An inversion problem on inferring the size distribution of precipitation areas from raingage measurements. *J. Atmos. Sci.*, **41**, 439–448, doi:10.1175/1520-0469(1984)041<0439:AIPOIT>2.0.CO;2.