Overview of MPLNET Version 3 Cloud Detection

JASPER R. LEWIS*

Joint Center for Earth Systems Technology,

University of Maryland Baltimore County, Baltimore, Maryland

JAMES R. CAMPBELL

Naval Research Laboratory, Monterey, California

ELLSWORTH J. WELTON

NASA Goddard Space Flight Center, Greenbelt, Maryland

SEBASTIAN A. STEWART

Science System and Applications, Inc., Lanham, Maryland

PHILLIP C. HAFTINGS

Science System and Applications, Inc., Lanham, Maryland

* *Corresponding author address*: NASA GSFC, Code 612, Greenbelt, MD 20771.
E-mail: jasper.r.lewis@nasa.gov

ABSTRACT

The National Aeronautics and Space Administration Micropulse Lidar Network Version 3 cloud detection algorithm is described and its differences relative to the previous version highlighted. Clouds are identified from normalized Level 1 signal profiles using two complementary methods. The first considers signal derivatives vertically for resolving low-level clouds. The second, which resolves high-level clouds like cirrus, is based on signal uncertainties given the relatively low signal-to-noise ratio exhibited in the upper troposphere by eye-safe network instruments, especially during daytime. Furthermore, a multi-temporal averaging scheme is used to improve cloud detection under conditions of weak signal-to-noise. Diurnal and seasonal cycles of cloud occurrence frequency based on one year of measurements at the Goddard Space Flight Center (Greenbelt, MD) site are compared for the new and previous versions. The largest differences, and perceived improvement, in detection occurs for high clouds (above 5-km, mean sea level) which increase in occurrence by nearly 6%. There is also an increase in the detection of multi-layered cloud profiles from 9% to 20%. Macrophysical properties and estimates of cloud optical depth are presented for a transparent cirrus dataset. However, the limit to which molecular signal can be reliably retrieved above cirrus clouds occurs between cloud optical depths of 0.5 and 0.8.

1. Introduction

Clouds play a critical role in the Earth's climate system because they are inextricably linked to the hydrological cycle and radiation budget (Liou 1986; Ramanthan et al. 1989). Information about cloud height, thickness, occurrence, and amount are critical inputs for a host of numerical applications involving climate research. Therefore, it is important to have highly accurate and quantitative data records of cloud properties that span several years and geographic regions. Verification of even the most basic modeling processes demands compulsory observations of global cloud occurrence, if there is to be any confidence in their fidelity.

Various methods of determining cloud climatologies exist, each with their own advantages and limitations. Visual observations from the surface (Warren et al. 1985; Hahn et al. 1996; Hahn and Warren 1999) provide cloud fraction and morphological cloud types. However, these can be biased by the quality of technician training, underestimation of high clouds, sparse global coverage, and nighttime bias. Passive radiometric sensors aboard satellites, which are the core input of the International Satellite Cloud Climatology Project (ISCCP; Rossow and Schiffer 1991, 1999), offer a true global representation and have the best (unobstructed) potential view of high clouds. However, these can undersample low-level maritime clouds and underrepresent optically-thin cirrus clouds (Holz et al. 2008).

Active sensors, like lidar and radar (Platt et al. 1994; Moran et al. 1998; Wang and Sassen 2001), are the primary tools for observing and profiling cloud vertical structure to high accuracy. When flown aboard satellites, like Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO; Winker et al. 2007) and CloudSat (Stephens et al. 2002), active sensors also provide global coverage. Even still, the relatively narrow profiling curtain of current active sensors limits observation densities. In the case of CALIPSO and CloudSat, these

54 missions provide at most two profiles per 24-hour period over most regions, which limits studies
55 of the diurnal impact of clouds on the Earth system. Fundamentally, an array of remote sensing
56 methods is needed in order to investigate the complexity of clouds (Schiffer and Rossow 1983).

57 ## 2. Micropulse Lidar Network

58 The National Aeronautics and Space Administration (NASA) Micropulse Lidar Network
59 (MPLNET; Welton et al. 2001, http://mplnet.gsfc.nasa.gov) is a federated network of micropulse
60 lidar (MPL) systems deployed worldwide in support of basic science and the NASA Earth
61 Observing System program (Wielicki et al. 1995). A benefit of MPLNET is the use of a
62 standardized instrument employing a common data processing algorithm with thorough
63 uncertainty characterization, which allows for straightforward comparisons between sites. With
64 sites in polar, mid-latitude, and tropical regions and continuous day/night, high temporal
65 resolution datasets going back as far as 1999, MPLNET datasets represent a valuable archive for
66 improving our understanding of global cloud macrophysical properties on diurnal, season, and
67 decadal scales.

68 There have been two versions of MPLNET data processing algorithms to date. The first,
69 referred to as Version 1, was released in 2000. Beginning in 2006, the project transitioned to
70 Version 2 (hereafter V2) data products which are currently available. Version 3 (hereafter V3)
71 data processing algorithms are currently in development.

72 The V2 Level 1 MPLNET data products contain the system diagnostics (e.g. solar
73 background counts, instrument temperature and energy, etc.), normalized relative backscatter
74 (NRB), and NRB uncertainty. The NRB is reported at one-minute temporal resolution up to 30-
75 km, typically at 75-m vertical resolution (some sites operate at 30-m). The NRB is defined as:

$$\mathrm{NRB}(z) = C\beta(z)T^2(z) \quad (1)$$

where C is the instrument calibration constant, β is the backscatter coefficient from both molecules and particles, T^2 is the corresponding total atmospheric two-way transmittance, and z is the altitude. NRB and the calculation of its uncertainty are discussed by Campbell et al. (2002) and Welton and Campbell (2002). Level 1 data are available in real time with no quality assurance. Cloud base and top heights are identified in the V2 Level 1.5 (real time, no quality assurance) and Level 2 (not real time, quality assured) MPLNET data products. The V2 Level 1.5 data products use temperature and pressure profiles from US Standard Atmospheres (COESA 1976) to determine molecular calculations, while National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis (Kalnay et al. 1996) is used for V2 Level 2 data.

MPLNET V2 data products have been used to distinguish cloud presence in a number of scientific investigations to date. For example, Campbell and Sassen (2008) use data from the South Pole to document polar stratospheric cloud occurrence over multiple seasons. Shupe et al. (2011) consider MPLNET measurements at Ny-Ålesund, Norway as context for evaluating Artic cloud properties. Others have investigated cirrus contamination of Aerosol Robotic Network (AERONET) aerosol optical depth in Southeast Asia (Chew et al. 2011; Huang et al. 2011) and globally (Huang et al. 2012). Lolli et al. (2013) use collocated 355/527-nm MPLNET observations to estimate the drizzle droplet size from stratocumulus and stratus clouds.

A new V3 cloud detection algorithm has been developed to improve the quality of MPLNET cloud products. The new algorithm uses a combination of signal-processing techniques and a multi-resolution temporal averaging scheme to resolve cloud boundaries. Meteorological profiles provided by the Goddard Earth Observing System – Version 5 (GEOS-5) Atmospheric General Circulation Model (AGCM; Rienecker et al. 2008; Molod et al. 2012) are used for

molecular calculations. Specifically, the Forward Processing for Instrument Teams (FP-IT) GEOS-5 Version 5.9.1 data are utilized (http://gmao.gsfc.nasa.gov/products). The model data are available at 3-hour intervals over 72 pressure levels at 0.625° longitude and 0.5° latitude resolution. Modeled profiles for this study were subsampled from the GEOS-5 grid containing the Goddard Space Flight Center (GSFC) site location, and interpolated to the MPLNET range and time resolutions (75 m, 1 minute).

The goals of this paper are to describe the new algorithm and demonstrate performance. We outline changes relative to V2 cloud detection and describe how the new algorithm is applied to a variety of cloudy scenes. We apply one year of data collected at the GSFC MPLNET site (38.99° N, 76.84° W, 0.05 km above mean sea level; MSL) to compare V2 and V3 results and highlight the impact of our upgraded techniques through differences in macrophysical cloud properties observed from this location.

3. Cloud detection algorithm description

Examples of daytime and nighttime NRB profiles at GSFC are shown in Fig. 1. Both profiles show high-level clouds with base heights near 10 km and top heights near 13 km, MSL. The daytime NRB profile exhibits relatively lower signal-to-noise compared with the nighttime case due to higher solar background, which makes detection of elevated layers an increasingly difficult task.

a. Version 2 cloud detection

Layers are identified in the V2 cloud detection algorithm by a combination of two retrieval methods applied to the Level 1 data products. The first method requires that the first derivative of the lidar signal exceed a minimum threshold in order to detect a layer. The assumption of strong signal gradients makes this well suited for detecting liquid-phase clouds, which are

122 frequently at lower levels in the NRB profile and correspond with higher signal-to-noise. This is
123 hereafter referred to as the gradient-based cloud detection method (GCDM). The second method
124 is designed for use in cases of low signal-to-noise ratio (SNR) and relies on uncertainties in the
125 lidar signal. This method uses two tunable thresholds and one objective threshold to identify
126 cloud boundaries, and is hereafter referred to as the uncertainty-based cloud detection method
127 (UCDM).

128 Given the relatively low SNR exhibited by the MPL in the upper troposphere at base one-
129 minute resolution (primarily during daytime), no single procedure is used to detect all cloud
130 types at all times. Thus the merger of these two methods offers the possibility to retrieve the
131 entire cloud vertical structure to the limit of signal attenuation. We describe the basis for each
132 method, as follows.

133 1) GRADIENT-BASED CLOUD DETECTION

134 Autonomous methods of cloud detection using gradients in the lidar signal, such as the
135 differential zero-crossing method described by Pal et al. (1992), are well established. The first
136 step in the GCDM is to normalize the NRB using the attenuated molecular backscatter
137 coefficient,

$$\beta'_m(z) = \beta_m T_m^2(z), \tag{2}$$

138 which produces an attenuated scattering ratio, β_r', multiplied by the instrument calibration
139 constant as

$$C\beta'_r(z) = \frac{C\beta(z)T^2(z)}{\beta'_m(z)} = \frac{C[\beta_m(z)+\beta_p(z)]T_m^2(z)T_p^2(z)}{\beta_m(z)T_m^2(z)} = C\left[1+\frac{\beta_p(z)}{\beta_m(z)}\right]T_p^2(z). \tag{3}$$

140 Here the subscripts *m* and *p* denote contributions from molecules and particles, respectively.

141 The first derivative of $C\beta_r'$ is used to identify clouds in the GCDM. Due to increasing
142 uncertainty in the profile with height, the GCDM retrieval is only performed up to a "noise

altitude", defined as the altitude at which the uncertainty, δNRB, exceeds half of the NRB (or conversely, analogous to an SNR of 2). Cloud presence corresponds with an increase in total backscatter, which results in a large positive gradient in $C\beta_r'$ with height as seen in Fig. 2. The threshold used to identify the cloud base is defined as

$$a_{\max} = K \cdot \overline{C\beta_r'}\,, \tag{4}$$

where $\overline{C\beta_r'}$ is the mean value of $C\beta_r'$ up to the noise altitude and K is an empirical parameter (unitless), set to 10 for this study. The value of K is chosen carefully so that it is high enough to reject insignificant peaks in the first derivative of $C\beta_r'$ (i.e. aerosol stratification in the surface-attached layer or signal noise) while remaining sensitive enough to identify weakly-scattering clouds.

The cloud base is identified at the altitude bin immediately preceding that where the first derivative of $C\beta_r'$ exceeds $a_{\max}$. Identification of cloud top is more ambiguous and is performed using one of two processes. A negative gradient in $C\beta_r'$ occurs near the top of a cloud, corresponding with the decrease in total backscatter and the impact of signal attenuation through the cloud. The first method used to identify the cloud top relies on a threshold defined as

$$a_{\min} = \overline{C\beta_r'} - a_{\max}\,. \tag{5}$$

The algorithm begins by looking for altitude bins above the cloud base where the first derivative of $C\beta_r'$ falls below $a_{\min}$. Then the altitude bin where the first derivative initially returns above $a_{\min}$ is identified as the cloud top. However, if this condition is not met, the cloud top is chosen as the altitude bin where the value of $C\beta_r'$ falls below the value at the cloud base or the noise altitude, whichever occurs first. If the lidar signal becomes significantly attenuated within the

cloud, the designation as an apparent cloud top is more appropriate in accordance with standards agreed upon by the Experimental Cloud Lidar Pilot Study (ECLIPS; Platt et al. 1994).

Two cloud layers are apparent in Fig. 2. Both cloud bases are identified at the altitudes immediately below the a_{max} exceedances (dashed line, positive derivative). The first (lowest) cloud top can be found using the a_{min} threshold (dashed line, negative derivative). But the derivative never falls below a_{min} for the second cloud layer. Therefore, the alternative process is used to identify the apparent cloud top. Only true (not apparent) cloud tops are reported in V2 MPLNET cloud products.

2) UNCERTAINTY-BASED CLOUD DETECTION

An alternative to algorithms that utilize gradients in the lidar return to identify clouds are approaches that compare cloudy lidar returns to clear sky returns (Clothiaux et al. 2007). Similarly, the UCDM uses a theoretical molecular return and the signal uncertainty to detect elevated clouds, and is fully described by Campbell et al. (2008, hereafter C08). However, a few comments regarding its implementation are warranted. The first step in the UCDM is to approximate the value of the instrument calibration constant. Level 1 NRB is divided by the attenuated molecular backscatter coefficient, resulting in an attenuated scattering ratio multiplied by the instrument calibration constant, as given in Eq. (3). Next, a clear-sky search is performed to locate a normalization region where we can approximate that β_p approaches zero over a certain number of range bins, N. The nature of the UCDM only allows for cloud detection at altitudes above the normalization region. The calibration constant is approximated by averaging Eq. (3) over the N bins. C08 stress that this final normalization value, C_f^*, must be distinguished from C due to unknown transmission losses below the normalization region.

As an example, a representation of the UCDM is shown in Fig. 3. Beginning at 1-km above ground level (AGL) and working upward, the uncertainty in $C\beta_r'$ is evaluated at each altitude bin according to the criteria set forth by C08 until a "clear-air" slot is found. In this case, the "clear-air" region used to determine the normalization value begins at r_1 = 6.46 km and ends at r_N = 8.71 km. Averaging $C\beta_r'$ between r_1 and r_N gives a value of C_f^* = 92.33 MHz km^3 sr μJ^{-1}.

Once C_f^* has been calculated, a so-called pseudo-attenuated backscatter and its uncertainty are solved as

$$\mathrm{PAB}(z) = \frac{\mathrm{NRB}(z)}{C_f^*}, \tag{6}$$

and

$$\delta\mathrm{PAB}(z) = \mathrm{PAB}(z)\sqrt{\left[\frac{\delta\mathrm{NRB}(z)}{\mathrm{NRB}(z)}\right]^2 + \left(\frac{\delta C_f^*}{C_f^*}\right)^2}. \tag{7}$$

Eq. (7) is then modified by substituting the attenuated molecular backscatter for the PAB to develop an objective threshold,

$$\alpha(z) = \beta_m(z)T_m^2(z) + \beta_m(z)T_m^2(z)\sqrt{\left[\frac{\delta\mathrm{NRB}(z)}{\beta_m(z)T_m^2(z)C_f^*}\right]^2 + \left(\frac{\delta C_f^*}{C_f^*}\right)^2}, \tag{8}$$

which is used to differentiate what are first presumed cloud returns from molecular return. Range bins above r_N that meet the condition

$$\mathrm{PAB}(z) - \delta\mathrm{PAB}(z) > \alpha(z) \tag{9}$$

are then evaluated to determine whether they represent particulate layer base heights. Range bins that do not meet the condition from Eq. (9) are used to establish "clear-air" slots and are

disqualified from the particulate base height search. In other words, a minimum detectable scattering ratio for particulates is defined as

$$\beta_{r,\min}(z) = \frac{\alpha(z) + \delta PAB(z)}{\beta'_m(z)}. \quad (10)$$

Running averages of PAB and δPAB are used in conjunction with two additional tunable thresholds, ϕ and κ (both analogous to a SNR), to determine the actual layer base and top heights. The threshold ϕ sets the minimum average value of PAB/δPAB for bins that exceed Eq. (10) in order to identify the layer base. At this point, and as described above, the UCDM assumes that any such layers detected are hydrometeor clouds, thereby leaving the potential for false detection of elevated aerosol layers. In the absence of supplementary information, however, such as color ratio (Liu et al. 2005) or depolarization (Cho et al. 2008, Omar et al. 2009) and combined with the goal of resolving as much thin cirrus as possible in the low SNR portions of the NRB profile, this is unavoidable. Mitigation strategies are described further below.

The threshold κ sets the minimum average value of PAB/δPAB for bins that do not exceed Eq. (10) in order to identify clear air layers and consequently particulate layer tops. In Fig. 3, the red line indicates the threshold used to distinguish particulate from molecular returns. Bins that exceed this objective threshold are evaluated using the tunable threshold ϕ, while bins with values lower than this threshold are evaluated using the tunable threshold κ. The sensitivity of V2 cloud detection to the tunable thresholds is evaluated in C08 and the values chosen for ϕ and κ will depend on the site location and instrument performance parameters.

3) V2 CLOUD RETRIEVALS

An example of V2 cloud retrievals is shown in the top panel of Fig. 4. Results from the GCDM and UCDM are integrated based on the noise altitude (described in Section 3.a.1). Clouds occurring below this height are reported from the GCDM. All clouds above the noise altitude are identified using the UCDM. At night, the noise altitude reaches above typical cirrus cloud heights at GSFC. Therefore, the GCDM is almost exclusively responsible for cloud detection. As a result, weakly-scattering cirrus can go undetected, since GCDM thresholds are tuned primarily with boundary layer phenomena in mind (i.e. suppression of aerosol identification). This can be observed frequently between 0300 and 0600 UTC in Fig. 4, where cirrus presence is underreported and cloud base heights are overestimated.

In the daytime, the noise altitude shown in Fig. 4 falls between 8–9 km and the UCDM is responsible for all cloud detection above it. In several instances, cloud bases (red markers) are shown while the corresponding cloud tops (orange markers) appear to be missing. In these cases the lidar signal is assumed to be significantly attenuated, and therefore no cloud top is reported.

b. Version 3 cloud detection

The V3 algorithm is based on V2 with a few meaningful changes to the UCDM. Consequently, the changes in V3 represent an update to C08 and how the GCDM and UCDM are merged. A schematic of the V3 cloud detection algorithm at the one-minute base NRB temporal resolution is shown in Fig. 5. Low-altitude obstructions (e.g. fog or low stratus decks) reduce SNR and limit the accuracy of cloud retrievals. Therefore, each profile is screened for these "beam-blocked" conditions by a process described further below. If no such obstructions are found, the first step in the UCDM is to calculate the normalization value.

1) NORMALIZATION REGION

As mentioned in C08, it is most practical to find a normalization region to calculate C_f^* nearest to the instrument in order to increase the depth of the profile analyzed for clouds. In V2,

243 the clear-sky search is initiated at 1 km AGL. However, in V3, an altitude of 5-km MSL is
244 chosen to reduce the likelihood of normalizing within relatively homogenous aerosol layers in or
245 near the boundary layer. Furthermore, based on recent global cloud and aerosol studies (Holz et
246 al. 2008; Sassen et al. 2008; Campbell et al. 2015), 5 km represents a reasonable height to begin
247 the search for high clouds in tropical and mid-latitude locations. A lower altitude may be needed
248 for polar sites, however. In the event that it is not possible to perform the normalization step at
249 5-km, the UCDM attempts to normalize lower in the atmospheric profile iteratively, to as low as
250 1 km AGL, as done in the V2 algorithm.

251 The normalization region also serves as the boundary between the GCDM and UCDM
252 retrievals in the V3 cloud algorithm, allowing the better-suited method (GCDM for low clouds
253 and UCDM for high clouds) to operate during both day and night. The V3 cloud retrievals in the
254 bottom panel of Fig. 4 can be compared with the V2 retrieval in the same figure to see the
255 relative apparent improvement. We also note that there are conditions when either method
256 (GCDM or UCDM) may be used to retrieve high-level or low-level clouds. Therefore, the
257 retrieval method for each cloud layer is provided as an output parameter.

258 2) OBJECTIVE THRESHOLD

259 The objective UCDM threshold, α, defined in Eq. (8) is now adjusted for attenuation within
260 cloud layers in V3. In order to attenuate the α threshold, we assume an initial unity transmission
261 at the base of the first detected cloud layer. Then the two-way transmittance is calculated at each
262 altitude bin Z within the cloud layer as

$$T_C^2(Z) = T_C^2(Z-1)\exp\left[-2S_C\beta_C(Z)\Delta z\right]$$
$$T_C^2(Z) = T_C^2(Z-1)\exp\left[-2S_C\left(\frac{\beta_r'(Z)}{T_C^2(Z-1)}-1\right)\beta_m(Z)\Delta z\right], \quad (11)$$

where β_C is the cloud backscatter coefficient, S_C is the extinction-to-backscatter ratio (or lidar ratio) of the cloud, Δz is the vertical resolution of the instrument, and β_r' is the attenuated scattering ratio given by

$$\beta_r'(z) = \frac{\mathrm{NRB}(z)}{C_f^* \beta_m'(z)} \approx \left[1 + \frac{\beta_C(z)}{\beta_m(z)}\right] T_C^2(z) . \tag{12}$$

Overestimating the single unknown term in these equations, the extinction-to-backscatter ratio, will lead to excessive attenuation of the objective threshold and eventual mischaracterization of clear sky as cloud. Therefore, in this step, a very conservative choice for S_C equal to the molecular extinction-to-backscatter ratio ($8\pi/3$) is used.

Once the transmittance has been determined, Eq. (8) is modified to recalculate the α threshold as

$$\alpha(z) = \beta_m(z) T_m^2(z) T_C^2(z) + \beta_m(z) T_m^2(z) \sqrt{\left[\frac{\delta \mathrm{NRB}(z)}{\beta_m(z) T_m^2(z) C_f^*}\right]^2 + \left(\frac{\delta C_f^*}{C_f^*}\right)^2} . \tag{13}$$

By attenuating the α threshold, we are able to better resolve cloud tops for optically-thick clouds. Furthermore, multilayer clouds, where the upper layer was missed entirely by the V2 algorithm, are now more effectively detected in some cases.

3) MULTI-TEMPORAL RESOLUTION

Cloud boundaries are only reported at one-minute temporal resolution in the V2 algorithm. However, instances of high solar background reduce UCDM performance. So, as described by C08, multi-temporal resolution settings are used in V3. In addition to the one-minute base temporal resolution, the UCDM is performed for intermediate (five-minute) and long (twenty-minute) temporal averages using a sliding window centered on a corresponding one-minute profile.

A flow chart describing the process is shown in Fig. 6. Retrievals at the base temporal resolution are used to screen profiles at longer averages, thus limiting the effects of "beam-blocked" profile contamination. Within a window of N profiles, k profiles are removed from the average if an attenuating structure is detected below 5 km at the base temporal resolution. If $k > N/2$, then the entire average profile is rejected.

A combined cloud scene is created using cloud boundaries retrieved from the three temporal resolutions. First, the combined scene uses the cloud boundaries from the base temporal resolution. Next, the intermediate and then long temporal resolutions are used to fill in the missing gaps as needed. Cloud boundaries are always reported at the highest temporal resolution possible to reduce the effects of cloud smearing caused by averaging. This is broadly consistent with the methodology used for NASA Cloud Aerosol Lidar with Orthogonal Polarization (CALIOP) Level 2 products and their gridding of multiple spatial resolutions from the selective iterated boundary locator (SIBYL) to their vertical feature mask (Vaughan et al. 2005). A noted difference is that SIBYL also uses an intensity-clearing process to remove features detected at finer resolutions from the coarser spatial averages. No such intensity clearing is performed with the V3 algorithm.

4) FALSE POSITIVES

As mentioned previously, use of the UCDM presumes first only cloud presence. Additional constraints are thus used to reduce the number of instances when noise excursions, elevated aerosol layers or poor normalizations produce false cloud retrievals. The first constraint establishes a minimum layer thickness of 150 m (i.e., two bins at 75-m resolution) in order to qualify a potential layer as a cloud. Therefore, we require all bins within a distance greater or equal to the minimum layer thickness to exceed the minimum detectable scattering ratio in Eq. (10) before a cloud base is established. Similarly, a minimum clear air distance of 150 m is used

306 to avoid falsely segmenting a single cloud into multiple layers. All bins within the minimum
307 clear distance must fall below the minimum detectable scattering ratio in order to establish a
308 cloud top. The second constraint requires that the standard deviation of the attenuated scattering
309 ratio (β_r') within the detected layer exceed an empirically-determined threshold, $\sigma_{\min}$, which
310 varies as a function of cloud top temperature. Cloud layers generally cause large variances in β_r',
311 either through attenuation effects in otherwise colloidally-stable liquid water clouds or since ice
312 crystals fall within cirrus cloud layers, creating complex structures. By contrast, aerosols in the
313 free troposphere settle in stratified stable layers absent of convection and are expected to be
314 homogeneous within each layer. Thus, clouds layers can be distinguished by their relatively
315 large standard deviations of β_r'.

316 In order to determine $\sigma_{\min}$, a dataset was developed consisting of 144 days at GSFC in 2012
317 and 27061 retrieved UCDM layers at one-minute resolution (18308 thin cirrus cloud layers, 3233
318 non-cirrus layers, 5520 aerosol layers) when the particulate type could be reasonably identified
319 from visual inspection (Fig. 7). Thin cirrus clouds are distinguished using a cloud top
320 temperature threshold of -37 °C (Sassen and Campbell 2001, Campbell et al. 2015) and a
321 maximum cloud optical depth (COD) of 0.3 (Sassen and Cho 1992). The COD calculation uses
322 a process described by Chew et al. (2011) and is discussed fully in Section 3.c.1. Non-cirrus
323 clouds are those with cloud top temperature warmer than -37 °C.

324 Several choices for $\sigma_{\min}$ were evaluated using error matrices (Congalton and Meade 1986)
325 and the corresponding values of accuracy and Matthews correlation coefficient (Matthews 1975;
326 MCC) which are defined as

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{14}$$

and

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (15)$$

where *TP* is the number of instances when clouds were correctly identified, *TN* is the number of instances when aerosols were correctly identified, *FP* is the number of instances when aerosols were misidentified as clouds, and *FN* is the number of instances when clouds were misidentified as aerosols. The choice for $\sigma_{\min}$ that resulted in the highest values of accuracy (0.92) and MCC (0.74) is given by

$$\sigma_{\min} = \begin{cases} 2, & \text{for } T > -37\,^{\circ}\text{C} \\ 10^{\frac{T+40}{10}}, & \text{for } -47\,^{\circ}\text{C} < T < -37\,^{\circ}\text{C} \\ 0.2, & \text{for T} < -47\,^{\circ}\text{C} \end{cases}. \quad (16)$$

The error matrix for the empirically-determined $\sigma_{\min}$ is provided in Table 1. However, we note that the dataset contains two atypical long-range smoke transport events (with corresponding top height temperatures between -50 °C and -60 °C), which comprised 35% of the aerosol category. If these two events are removed, the accuracy and MCC improve to 0.98 and 0.92, respectively. Similar results were observed at the five- and twenty-minute resolutions, but not shown for brevity.

The final constraint used to distinguish cloud from aerosol layers is that the estimated COD exceed a threshold, $\tau_{\min}$. Through empirical testing, we estimate COD and set $\tau_{\min} = 0.005$ based on analysis of these subsets relative to the perception of how noise impacts these sub-samples combined with a similar analysis by Thorsen et al. (2011).

We briefly note here that lidars with polarization capabilities have recently been incorporated into the MPLNET project. However, because the overwhelming majority of

existing data (which spans from 2000 – present) was collected without polarization, the algorithm presented here does not rely on such data. It remains as a future goal to demonstrate how polarization can be used to improve aerosol-cloud discrimination, once a sufficient amount of data is collected from the new polarized sites.

c. Version 3 algorithm output

A listing of the V3 cloud detection algorithm output parameters are provided in Table 2. The output parameters from all temporal averages are gridded to one-minute temporal resolution, as previously described in the combined cloud scene. The number of cloud layers detected, day flag, and attenuation altitude are given as a single value each minute, characterizing the atmospheric column. All other cloud products and data flags correspond with individual cloud layers, and are provided each minute with dimensions equal to the number of cloud layers detected. Meteorological values at the cloud boundaries are obtained from the interpolated GEOS-5 profiles described in Section 2.

1) CLOUD PHASE AND CIRRUS CLOUD OPTICAL DEPTH

In the absence of visual cloud observations, as is the case for autonomous lidar measurements made by MPLNET, Sassen and Campbell (2001) recommend using a minimum cloud top temperature of -37 °C to identify cirrus. In the V3 cloud algorithm, we use this thermal threshold to distinguish ice clouds (i.e. cirrus) from all other cloud phases. Because depolarization capabilities are not standard for all MPLNET instruments, no attempt is made to distinguish liquid from mixed phase clouds.

Campbell et al. (2015) evaluate the -37 °C cloud top temperature threshold globally versus the Level 2 CALIOP algorithms that identify ice-phase cloud layers and found that over 99% of clouds satisfying this thermal threshold were classified as ice. Furthermore, 81% of all ice clouds had cloud top temperatures less than -37 °C. They conclude, consistent with the findings

370 of Sassen and Campbell (2001), that this thermal threshold is stable for specifically
371 distinguishing cirrus cloud presence in lidar studies that lack depolarization, though there is some
372 ambiguity in cases of "warm" cirrus that likely coincide with cloud top temperature greater than
373 -37 °C.

374 An estimated COD is calculated for clouds distinguished as cirrus using the procedure
375 described by Chew et al. (2011). Two-way cloud transmission is calculated using Eq. (11).
376 However, now the value of S_C is selected based on the cloud top temperature. Reported values
377 of S_C are on the order of 16–18 sr for liquid water clouds (Pinnick et al. 1983, Yorks et al. 2011)
378 and 10–40 sr for cirrus (Sassen and Comstock 2001; Chen et al. 2002; Yorks et al. 2011; Garnier
379 et al. 2015). A value of S_C = 18 sr is chosen for layers with clouds top temperatures warmer than
380 -37 °C and S_C = 20 sr at colder temperatures where cirrus clouds are expected. We note that due
381 to uncertainty in the lidar ratio for cirrus clouds, these estimates may represent the lower limit of
382 COD.

383 Next, Eq. (12) is used to solve for the cloud backscatter coefficient and the estimated COD
384 is given by

$$\tau = S_C \int_{base}^{top} \beta_C(z)\,dz = S_C \int_{base}^{top} \beta_m(z)\left(\frac{\beta_r'(z)}{T_C^2(z)} - 1\right)dz\,. \tag{17}$$

385 The fidelity of the COD estimate is limited by the choice of S_C and accuracy of cloud boundaries
386 retrieved. The relative error in the lidar-derived optical depth is smallest for low optical depths
387 and proportional to $\Delta S_C/S_C$ as τ approaches zero (Winker et al. 2009). Lidar signals are unable to
388 penetrate through optically-thick clouds, which causes uncertainty in the value at the apparent
389 cloud top. In these cases, the estimated COD will be biased low. Similarly, attenuation from the
390 bottom-most cloud layer leads to uncertainty in corresponding retrievals of higher clouds for

391 lidar profiles containing multiple cloud layers. As a result, the estimated COD will be most
392 reliable for single-layer, optically-thin clouds.

393 2) RETRIEVAL INDEX

394 While cloud boundaries are only reported at a single temporal resolution, a retrieval index is
395 included to indicate whether the cloud was also detected at one or more of the other temporal
396 averages. Cloud layers at different temporal resolutions are considered the same if (i) they share
397 a common base or top height within a vertical depth of 250 m or (ii) one cloud layer is
398 completely enveloped within the other.

399 An example of a combined cloud scene, with corresponding retrieval indices, is shown in
400 Fig. 8. The value of the retrieval index is equal to the sum of the temporal resolutions used to
401 identify the cloud layer. For example, if a cloud is detected at all three temporal resolutions, the
402 value of the retrieval index is 1+5+20 = 26. The advantage of the multi-temporal averaging
403 scheme can be seen during the day between 1400–1500 UTC in Fig. 8. The elevated cloud layer
404 (~15 km) is mostly undetected at the one-minute resolution, but can be resolved using the longer
405 averages. The cloud layer at ~2 km produces "beam-blocked" conditions that prevent use of
406 higher-temporal averages for much of the cirrus cloud layer above it.

407 3) ATTENUATION ALTITUDE

408 Because the lidar signal can become completely attenuated within optically-thick clouds, it
409 is important to determine when a true cloud top is being reported as opposed to an apparent
410 cloud top. Nadir-pointing lidar instruments have an advantage of using the ground return to
411 determine if the lidar signal has been extinguished. However, with zenith-pointing lidar, that
412 determination is more tenuous. Winker and Vaughan (1994) defined a transmittance index to
413 determine when the lidar signal was fully attenuated based on the percentage of samples above
414 the cloud top that exceeded the background. Other techniques used for zenith-pointing lidar

415 have included the use of a minimum threshold lidar signal along with its slope (Wang and Sassen
416 2001) and comparisons with molecular profiles (Lo et al. 2006).

417 In V3, cloud tops (both true and apparent) are reported for all cloud layers along with the
418 altitude at which the lidar signal is determined to be fully attenuated. This attenuation altitude is
419 found by starting at the range bin of the highest reported cloud altitude and incrementally
420 moving upwards in the profile until, within a depth of 2 km, (i) the percent difference between
421 the mean pseudo-attenuated backscatter and modeled attenuated molecular backscatter falls
422 below some threshold, T_1, and (ii) either the backscatter signal falls below a minimum value or
423 the percentage of range bins where the backscatter signal is less than zero exceeds a threshold T_2.

424 This application pertains specifically to profiles that contain clouds or other obstructions,
425 since the attenuation thresholds can also be satisfied by other conditions that lead to low SNR
426 (e.g. high solar background). "Beam-block" conditions from low-altitude obstructions are found
427 with the same search criteria, though the search is limited to the first 2 km above the surface.

428 ## 4. Results

429 In order to demonstrate the effects of the changes implemented in the V3 algorithm, we
430 compare V2 and V3 cloud retrievals for one year at the GSFC MPLNET site. Table 3 and Fig. 9
431 show data sampling statistics for 2012, including the total number of profiles and percentage of
432 time when 1-minute NRB measurements were available monthly. Observable profiles are given
433 as the number and percentage of available profiles that are not "beam-blocked" below 2 km,
434 MSL. Profile attenuation was determined using the V2 method because it is the most restrictive
435 and ensures an even comparison between the two cloud detection algorithms. The diurnal
436 distribution of data recorded and successful V3 normalizations to calculate C_f^* are also shown.
437 There is very little differentiation between the distributions of V2 and V3 normalizations, so only

438 the number of successful V3 normalizations are shown for simplicity. Though MPLNET
439 measurements are continuous, there is a decrease in the number of successful normalizations
440 near solar noon in relatively low SNR conditions.

441 Because V3 uses a merged cloud scene and V2 is only processed at one-minute resolution,
442 V3 retrievals are evaluated using the base one-minute resolution (hereafter V3b) and the merged
443 cloud scene (hereafter V3m). Comparisons are limited to cloud base statistics because cloud
444 tops are not recorded for all V2 retrievals. Finally, we describe the macrophysical and optical
445 characteristics of cirrus clouds observed during this study, again adhering to the methodology
446 described in Campbell et al. (2015), using V3 retrievals.

447 *a. Vertical dependence*

448 Figure 10 shows the cloud base distributions retrieved from the V2 and V3 algorithms,
449 respectively, at GSFC during 2012. A bimodal distribution similar to that observed by Winker
450 and Vaughan (1994), with peaks at ~1-2 km and ~9-10 km is apparent. The total numbers of
451 cloud observations are 269505, 304363, and 332810 for the V2, V3b, and V3m retrievals,
452 respectively. Compared with V2, the number of cloud observations increases by 12.9% and
453 23.5% for V3b and V3m, respectively. The largest increase in the number of clouds observed
454 occurs at altitudes above 5 km.

455 Because the difference in the number of clouds retrieved shows a clear vertical dependence,
456 we examine them specifically for three sub-samples, by defining low clouds as those with base
457 heights less than 2 km, high clouds as those with base heights greater than 5 km, and middle
458 clouds as those with base heights between 2-5 km (WMO 1975). The number of lidar profiles
459 for each classification, along with occurrence frequency, is shown in Table 4. Cloud occurrence
460 frequency is defined as the number of lidar profiles containing a particular cloud classification

461 divided by the total number of observable profiles. Regardless of the retrieval method (V2, V3b,
462 and V3m), occurrence frequency is nearly identical for low clouds, which reflects the relative
463 consistency in GCDM application between V2 and V3 at one-minute resolution. High clouds
464 show the largest increase in occurrence frequency. For example, comparing the V2 and V3m
465 algorithms, the occurrence frequency of high clouds increases by 5.9% attributable to (i) the
466 increased identification of elevated, multi-layer cloud decks using an attenuated UCDM
467 threshold, (ii) increased use of the UCDM to identify high clouds at day and night, and (iii)
468 multi-temporal application of UCDM to increase SNR. To (i), V2 retrievals resulted in 91% of
469 cloudy lidar profiles containing single-layer clouds. The percentage of single-layer clouds
470 decreases to 83% and 80% for V3b and V3m, respectively.

471 *b. Seasonal dependence*

472 Figure 11 shows the annual cycle for low, middle, high, and total cloud classifications
473 during 2012. The low cloud occurrence frequency is nearly identical for all three retrieval
474 methods. Middle clouds retrieved using V3b and V3m exhibit a slight separation from V2. The
475 largest differences are again seen with high-cloud retrievals. While the annual cycles for high
476 clouds show similar patterns for all three retrievals, there is an increase in occurrence frequency
477 of ~3% and 6% for V3b and V3m, respectively. The increase in high-cloud occurrence
478 frequency when compared to V2 ranges from 1% to 4% using V3b and 4% to 10% for V3m.
479 The largest differences for high-cloud occurrence frequency between V2 and V3 occurs during
480 summer months, which is coincident with the period when the sun is at its highest elevation and
481 thus solar background is highest.

482 *c. Diurnal dependence*

483 Differences in the diurnal cycle show similar characteristics as the annual cycles for low and
484 middle clouds. As seen in Fig. 12, V2, V3b, and V3m are nearly identical for low clouds. While
485 V3b and V3m show slight differences from V2, they are indistinguishable from each other.
486 High-cloud diurnal cycles follow the same trends for all three retrievals. However, the cloud
487 occurrence frequency is higher for V3b and highest for V3m. No clear diurnal trend is apparent
488 at GSFC because some changes (e.g. using the UCDM at all times) affect both day and night
489 retrievals. At tropical sites, where the solar background is higher and longer temporal averaging
490 is necessary, there may be a more obvious diurnal trend.

491 *d. Macrophysical and optical cirrus properties*

492 Based on the greater detection of high clouds demonstrated above, we characterize cirrus
493 clouds over the GSFC site as detected by the V3 algorithm. As stated earlier, cirrus presence is
494 determined using a cloud top temperature threshold of -37 °C. Additionally, we limit the
495 analysis to cases when (i) only cirrus clouds (no underlying liquid water or mixed phase clouds)
496 were detected in the profile, (ii) the estimated COD was less than 3, based on the upper-limit for
497 cirrus clouds suggested by Sassen and Cho (1992), and (iii) the attenuation altitude was at least 2
498 km above the cloud top. The final constraint limits the analysis to "transparent cirrus" cases for
499 which the algorithm is more likely to identify the true cloud top.

500 The resulting dataset includes 57930 cirrus clouds. The majority of cloud detections (82%)
501 occur at the base one-minute temporal resolution. The largest occurrence rate of the coarse
502 temporal averages occurs at or near noon and during the summer months when the solar
503 background is highest.

504 Table 5 summarizes the seasonal and annual mean characteristics of the transparent cirrus
505 dataset. The monthly variation in the macrophysical properties is shown in Fig. 13. Cirrus

clouds over GSFC tend to be higher and thinner (geometrically and optically) in the spring and summer and lower and thicker in the fall and winter seasons. Cirrus also occur more frequently in the spring and summer months. The transparent cirrus dataset is composed almost entirely (~95%) of sub-visual (COD < 0.03) and thin (COD < 0.3) cirrus clouds. Uncertainties in the value of the extinction-to-backscatter ratio and cloud top height could lead to an exaggeration of this finding. However, it should be noted that when the extinction-to-backscatter ratio is increased from 20 sr to 30 sr (not shown), 86% of cirrus clouds still have a COD less than 0.3. These results are qualitatively consistent with the findings of Dupont et al. (2010), who reported 50-75% of non-opaque cirrus clouds had an optical thickness less than 0.3 based on ground-based lidar and CALIPSO observations at four mid-latitude sites.

Frequency distributions of the optical and macrophysical properties are presented in Fig. 14. The cloud optical depth peaks in the sub-visual range and has a positive skew. The transparent cirrus dataset suggests that the limit at which we are able to resolve molecular signal above cloud, and thus reliably determine the cloud top, occurs near a COD of 0.5. However, if an extinction-to-backscatter ratio of 30 sr is used, this limit occurs near a COD of 0.8.

A comparison of daytime and nighttime cloud retrievals is provided in Table 6. There are only slight differences in the occurrence frequency between day and night cases. However, the geometric and optical depths are considerably lower in the daytime. The thinning of daytime cirrus may be attributable to difficulty in correctly identifying cloud boundaries due to solar background effects (Thorsen et al. 2013). However, convective cloud remnants are also likelier to occur during daytime hours. Decoupling the two, aside from seasonal influence, is outside the scope of this analysis. In the same manner, the daytime retrievals are more likely to be

528 considered as totally attenuated due to the higher solar background reducing the possibility to
529 resolve molecular signal at cirrus heights.

530 ## 5. Summary and Discussion

531 A new Version 3 (V3) cloud detection algorithm has been developed within the NASA
532 Micropulse Lidar Network (MPLNET) that uses a combination of retrieval methods and a multi-
533 temporal averaging scheme. Most V3 changes represent updates to the Version 2 (V2)
534 uncertainty-based threshold algorithm introduced by Campbell et al. (2008). The threshold used
535 to identify cloud presence now accounts for attenuation losses within cloud layers, which allows
536 for better estimation of cloud tops and boundaries of overlying cloud layers in profiles where
537 multiple cloud layers are detected. A more synergistic merging of the gradient-based cloud
538 detection method (GCDM) and uncertainty-based cloud detection method (UCDM) improves
539 nighttime clouds detection of tenuous high clouds. The incorporation of coarser temporal
540 resolutions at intermediate (5-minute) and long (20-minute) averages improves detection in
541 situations with low SNR (e.g. high solar background). One year of data at the NASA Goddard
542 Space Flight Center (GSFC) in Greenbelt, MD is used to show the effect of these updates on
543 cloud retrievals.

544 The largest impact of the changes to the cloud detection algorithm is evident with high
545 clouds (those with cloud base > 5 km), while the diurnal and annual cycles of low and middle
546 clouds exhibit only slight changes from V2 to V3. The high-cloud occurrence frequency
547 increases by nearly 6% at GSFC when using the V3 merged cloud scene compared with the V2
548 retrieval. Furthermore, the ability to detect multi-layered cloud scenes is improved with the V3
549 algorithm. The results show that 91% of clouds in 2012 at the NASA GSFC project site were
550 recorded as single-layer clouds according to the V2 retrieval compared with 80% for V3.

551 A brief investigation of macrophysical properties for transparent cirrus clouds shows that
552 the mean base and top heights at GSFC occur at 10.17 ± 1.63 km and 11.07 ± 1.43 km,
553 respectively. The highest and thinnest (both geometrically and optically) cirrus are found during
554 the spring and summer months, which was coincident with the highest cirrus occurrence
555 frequency. There is no significant difference in occurrence frequency between daytime and
556 nighttime retrievals. However, cirrus clouds are thinner (both geometrically and optically) in
557 daytime than nighttime, which may be attributed to a combination of increased uncertainty due
558 to the solar background effects and higher occurrence of convective cloud remnants during the
559 day. Notably, the limit to which we are able to resolve molecular signal above cirrus clouds
560 occurs between cloud optical depths of 0.5 and 0.8, allowing for uncertainty in the extinction-to-
561 backscatter ratio.

562 The value of the MPLNET cloud datasets is in its continuous (both day and night) and long-
563 term measurements at polar, mid-latitude, and tropical sites using a standard instrument and data
564 processing algorithm. Incorporating the V3 cloud retrievals from MPLNET as part of a multi-
565 instrument investigation will enhance our current knowledge of clouds, in particular cirrus. As it
566 stands, the cloud products provide a unique validation dataset for the modeling community and
567 satellite measurements. With some MPLNET sites now well into their second decade of
568 continuous cloud and aerosol observations, the project has become an integral component of
569 ground-based evaluation of atmospheric processes and verification of NASA satellite missions.
570 This paper thus represents our continuing effort to optimize the fidelity of project datasets for the
571 benefit of the community and in sustaining general scientific inquiry.

572 *Acknowledgments.*

573 The authors acknowledge Larry Belcher for processing the V2 lidar data, and the MPLNET
574 PIs and staff for their efforts in establishing and maintaining the GSFC site. The GEOS-5
575 meteorological data were provided by the NASA Global Modeling and Assimilation Office
576 (GMAO) at GSFC. The NASA Micro-Pulse Lidar Network is funded by the NASA Earth
577 Observing System and Radiation Sciences Program. Author JRC acknowledges the support of
578 NASA Interagency Agreement NNG13HH10I on behalf of MPLNET.

REFERENCES

Campbell, J. R., D. L. Hlavka, E. J. Welton, C. J. Flynn, D. D. Turner, J. D. Spinhirne, V. S. Scott, and I. H. Hwang, 2002: Full-time, eye-safe cloud and aerosol lidar observation at Atmosphere Radiation Measurement program sites: Instrument and data processing. *J. Atmos. Oceanic Technol.*, **19**, 431-442.

Campbell, J. R. and K. Sassen, 2008: Polar stratospheric clouds at the South Pole from 5 years of continuous lidar data: macrophysical, optical and thermodynamic properties, *J. Geophys. Res.*, **113**, D20204, doi:10.1029/2007JD009680.

Campbell, J. R., K. Sassen, and E. J. Welton, 2008: Elevated cloud and aerosol layer retrievals from micropulse lidar signal profiles. *J. Atmos. Oceanic Technol.*, **25**, 685-700, doi:10.1175/2007JTECHA1034.1.

Campbell, J. R., M. A. Vaughan, M. Oo, R. E. Holz, J. R. Lewis, E. J. Welton, 2015: Distinguishing cirrus cloud presence in autonomous lidar measurements. *Atmos. Meas. Tech.*, **8**, 435-449, doi:10.5194/amt-8-435-2015.

Chen, W.-N., C.-W. Chiang, and J.-B. Nee, 2002: Lidar ratio and depolarization ratio for cirrus clouds. *Appl. Opt.*, **41**, 6470-6476, doi:10.1364/AO.41.006470.

Chew, B. N., J. R. Campbell, J. S. Reid, D. M. Giles, E. J. Welton, S. V. Salinas, and S. C. Liew, 2011: Tropical cirrus cloud contamination in sun photometer data. *Atmos. Environ.*, **45**, 6724-6731, doi:10.1016/j.atmosenv.2011.08.017.

Cho, H.-M., P. Yang, G. W. Kattawar, N. L. Nasiri, Y. Hu, P. Minnis, C. Trepte, D. Winker, 2008: Depolarization ratio and attenuated backscatter for nine cloud types: analyses based on collocated CALIPSO lidar and MODIS measurements. *Opt. Express*, **16**, 3931-3948, doi:10.1364/OE.16.003931.

602 Clothiaux, E. E., G. G. Mace, and T. P. Ackerman, 2007: An automated algorithm for detection
603 of hydrometeor return in micropulse lidar data. *J. Atmos. Oceanic Technol.*, **15**, 1035-1042.
604 Committee on Extension to the Standard Atmosphere (COESA), 1976: U.S. Standard
605 Atmosphere, 1976. U.S. Government Printing Office, Washington, D.C.
606 Congalton, R. G. and R. A Mead, 1986: Techniques used in assessing the accuracy of remotely
607 sensed data from error matrices. *IEEE Trans. Geosci. Remote Sens.*, **GE-24**, 169–174.
608 Dupont, J.-C., M. Haeffelin, Y. Morille, V. Noël, P. Keckhut, D. Winker, J. Comstock, P.
609 Chervet, and A. Roblin, 2010: Macrophysical and optical properties of midlatitude cirrus
610 clouds from four ground-based lidars and collocated CALIOP observations. *J. Geophys.*
611 *Res.*, **115**, D00H24, doi:10.1029/2009D011943.
612 Garnier, A., Pelon, J., Vaughan, M. A., Winker, D. M., Trepte, C. R., and Dubuisson, P. , 2015:
613 Optical depths of semi-transparent cirrus clouds over oceans from CALIPSO infrared
614 radiometer and lidar measurements, and an evaluation of the lidar multiple scattering factor.
615 *Atmos. Meas. Tech. Discuss.*, **8**, 2143-2189, doi:10.5194/amtd-8-2143-2015.
616 Hahn, C. J. and S. G. Warren, 1999: Extended edited synoptic cloud reports from ships and land
617 stations over the globe, 1952-1996. NDP-026C. Carbon Dioxide Information Analysis
618 Center, Oak Ridge National Laboratory, Oak Ridge, TN.
619 Hahn, C. J., S. G. Warren, and J. London, 1996: Edited synoptic cloud reports from ships and
620 land stations over the globe, 1982-1991. NDP-026B, Carbon Dioxide Information Analysis
621 Center, Oak Ridge National Laboratory, Oak Ridge, TN.
622 Holz, R. E., S. A. Ackerman, F. W. Nagle, R. Frey, S. Dutcher, R. E. Kuehn, M. A. Vaughan,
623 and B. Baum, 2008: Global Moderate Resolution Imaging Spectroradiometer (MODIS)

624 cloud detection and height evaluation using CALIOP. *J. Geophys. Res.*, **113**, D00A19,
625 doi:10.1029/2008JD009837.
626 Huang, J., N. Hsu, S.-C. Tsay, M.-J. Jeong, B. N. Holben, T. A. Berkoff, and E. J. Welton,
627 2011: Susceptibility of aerosol optical thickness retrievals to thin cirrus contamination
628 during the BASE-ASIA campaign. *J. Geophys. Res.*, **116**, D08214,
629 doi:10.1029/2010JD014910.
630 Huang, J., et al., 2012: Evaluations of cirrus contamination and screening in ground aerosol
631 observations using collocated lidar systems. *J. Geophys. Res*., **117**, D15204,
632 doi:10.1029/2012JD017757.
633 Kalnay E., et al., 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor.*
634 *Soc.*, **77**, 437–471.
635 Liou, K. N., 1986: Influence of cirrus clouds on weather and climate processes: A global
636 perspective. *Mon. Weather Rev*., **114**, 1167-1199.
637 Liu, Zhaoyan, A. Omar, Y. Hu, M. A. Vaughan, D. M. Winker, 2005: CALIOP algorithm
638 theoretical basis document, Part 3: Scene classification algorithms. Tech Rep. PC-SCI-202.,
639 NASA Langley Research Center, Hampton, VA.
640 Lo, C., J. M. Comstock, and C. Flynn, 2006: An Atmospheric Radiation Measurement value-
641 added product to retrieve optically thin visible optical depth using micropulse lidar. Tech.
642 Rep. DOE/SC-ARM/TR-077, Dept. of Energy, Washington, D.C.
643 Lolli, S., E. J. Welton, and J. R. Campbell, 2013: Evaluating light rain drop size estimates from
644 multiwavelength Micropulse Lidar Network profiling. *J. Atmos. Oceanic Technol.*, **20**,
645 2798-2807, doi:10.1175/JTECH-D-13-00062.1.

646 Matthews, B. W., 1975: Comparison of the predicted and observed secondary structure of T4
647 phage lysozyme. Bichim. Biophys. Acta, **405**, 442-451.
648 Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012: The GEOS-
649 5 atmospheric general circulation model: Mean climate and development from MERRA to
650 Fortuna. Tech. Rep. Ser. on Global Model. and Data Assim., Vol. 28, NASA Goddard Space
651 Flight Center, Greenbelt, MD.
652 Moran, K. P., B. E. Martner, M. J. Post, R. A. Kropfli, D. C. Welsh, and K. B. Widener, 1998:
653 An unattended cloud-profiling radar for use in climate research. *Bull. Amer. Meteor. Soc.*,
654 **79**, 442-455.
655 Omar, A. H., et al., 2009: The CALIPSO automated aerosol classification and lidar ratio
656 selection algorithm. *J. Atmos. Oceanic Technol.*, 26, 1994-2014,
657 doi:10.1175/2009JTECHA1231.1
658 Pal, S. R., W. Steinbrecht, and A. I. Carswell, 1992: Automated method for lidar determination
659 of cloud base height and vertical extent. *Appl. Opt.*, **31**, 1488-1494,
660 doi:10.1364/AO.31.001488.
661 Pinnick, R. G., S. G. Jennings, P. Chylek, C. Ham, and W. T. Grandy Jr. 1983: Backscatter and
662 extinction in water clouds. *J. Geophys. Res.*, **88 (C11)**, 6787-6796,
663 doi:10.1029/JC088iC11p06787.
664 Platt, C. M. et al, 1994: The experimental cloud lidar pilot study (ECLIPS) for cloud radiation
665 research. *Bull. Amer. Meteor. Soc.*, **75**, 1635-1654.
666 Ramanthan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D.
667 Hartmann, 1989: Cloud radiative forcing and climate: Results from the earth radiation
668 budget experiment. *Science*, **243**, 57-63.

669 Rienecker, M. M., et al., 2008: The GEOS-5 data assimilation system – documentation of
670 versions 5.0.1, 5.1.0, and 5.2.0. Tech. Rep. Ser. on Global Model. and Data Assim. Vol. 27,
671 NASA Goddard Space Flight Center, Greenbelt, MD.
672 Rossow, W. B. and R. A. Schiffer, 1991: ISCCP cloud data products. *Bull. Amer. Meteor. Soc.*,
673 **72**, 2-20.
674 Rossow, W. B. and R. A. Schiffer, 1999: Advances in understanding clouds from ISCCP. Bull.
675 *Amer. Meteor. Soc.*, **80**, 2261-2287.
676 Sassen, K. and J. R. Campbell, 2001: A midlatitude cirrus cloud climatology from the Facility
677 for Atmospheric Remote Sensing. Part I: Macrophysical and synoptic properties. *J. Atmos.*
678 *Sci.*, **58**, 481-496.
679 Sassen, K. and J. M. Comstock, 2001: A midlatitude cirrus climatology from the Facility for
680 Atmospheric Remote Sensing. Part III: Radiative properties. *J. Atmos. Sci.*, **58**, 2113-2127.
681 Sassen, K. and B. S. Cho, 1992: Subvisual-thin cirrus lidar dataset for satellite verification and
682 climatological research. *J. Appl. Meteor.*, **31**, 1275-1285.
683 Sassen, K., Z. Wang, and D. Liu, 2008: Global distribution of cirrus clouds from
684 CloudSat/Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO)
685 measurements. *J. Geophys. Res.*, **113**, D00A12, doi:10.1029/2008JD009972.
686 Schiffer, R. A. and W. B. Rossow, 1983: The International Satellite Cloud Climatology Project
687 ISCCP: The first project of the World Climate Research Programme. *Bull. Amer. Meteor.*
688 *Soc.*, **64**, 779-784.
689 Shupe, M. D., V. P. Walden, E. Eloranta, T. Uttal, J. R. Campbell, S. M. Starkweather, and M.
690 Shiobara, 2011: Clouds at Arctic atmospheric observatories, Part I : occurrence and

691 macrophysical properties, *J. Appl. Meteorol. Clim.*, **50**, 626-644, DOI:
692 10.1175/2010JAMC2467.1.
693 Stephens, G. L., et al. 2002: The CloudSat mission and the A-Train. *Bull. Amer. Meteor. Soc.*,
694 **834**, 1771-1790.
695 Thorsen, T. J., Q. Fu, and J. Comstock, 2011: Comparison of the CALIPSO satellite and ground-
696 based observations of cirrus clouds at the ARM TWP sites. *J. Geophys. Res.*, **116**, D21203,
697 doi:10.1029/2011JD015970.
698 Thorsen, T. J., Q. Fu, J. M. Comstock, C. Sivaraman, M. A. Vaughan, D. M. Winker, and D.
699 Turner, 2013: Macrophysical properties of tropical cirrus clouds from the CALIPSO satellite
700 and from ground-based micropulse lidar and Raman lidars. *J. Geophys. Res.*, **118**, 9209-
701 9220, doi:10.1002/jgrd.50691.
702 Vaughan, M., D. Winker, and K. Powell, 2005: CALIOP algorithm theoretical basis document,
703 Part 2: Feature detection and layer properties algorithms. Tech Rep. PC-SCI-202.01, NASA
704 Langley Research Center, Hampton, VA.
705 Wang, Z. and K. Sassen, 2001: Cloud type and macrophysical property retrieval using multiple
706 remote sensors. *J. Appl. Meteor.*, **50**, 1665-1682.
707 Warren, S. G., C. J. Hahn, and J. London, 1985: Simultaneous occurrence of different cloud
708 types. *J. Climate Appl. Meteor.*, **13**, 658-667.
709 Welton, E. J. and J. R. Campbell, 2002: Micropulse lidar signals: Uncertainty analysis. *J. Atmos.*
710 *Oceanic Technol.,* **19**, 2089-2094.
711 Welton, E. J., J. R. Campbell, J. D. Spinhirne, and V. S. Scott, 2001: Global monitoring of
712 clouds and aerosols using a network of micro-pulse lidar systems. *Lidar remote Sensing for*

713 *Industry and Environmental Monitoring*, U. N. Singh, T. Itabe, and N. Sugimoto, Eds., Proc.
714 SPIE, 4153, 151-158.
715 Wielicki, B. A., R. D. Cess, M. D. King, D. A. Randall, and E. F. Harrison, 1995: Mission to
716 planet Earth: Role of clouds and radiation in climate. *Bull. Amer. Meteor. Soc.*, **76**, 2125-
717 2153.
718 Winker, D. M. and M. A. Vaughan, 1994: Vertical distribution of clouds over Hampton, Virginia
719 observed by liar under the ECLIPS and FIRE ETO programs. *Atmos. Res.*, **34**, 117-133.
720 Winker, D. M., W. H. Hunt, and M. J. McGill, 2007: Initial performance assessment of CALIOP.
721 *Geophys. Res. Lett.*, **34**, L19803, doi:10.1029/2007GL030135.
722 Winker, D. M., M. A. Vaughan, A. Omar, Y. Hu, and K. A. Powell, 2009: Overview of the
723 CALIPSO mission and CALIOP data processing algorithms. *J. Atmos. Oceanic Technol.*,
724 **26**, 2310-2323, doi:10.1175/2009JTECHA1281.1.
725 World Meteorological Organization (WMO), 1975: International Cloud Atlas, vol. I, Manual on
726 the Observation of Clouds and Other Meteors. WMO No. 407, Geneva.
727 Yorks, J. E., D. L. Hlavka, W. D. Hart, and M. J. McGill, 2011: Statistics of cloud optical
728 properties from airborne lidar measurements. *J. Atmos. Oceanic Technol.*, **28**, 869-883,
729 doi:10.1175/2011JTECHA1507.1.

730 **List of Tables**

731

732 TABLE 1. Error matrix for cloud-aerosol discrimination using σ_{min}

	Predicted cloud	Predicted aerosol
True cloud	21242	299
True aerosol	1851	3669

733

734 Table 2. MPLNET V3 cloud detection algorithm output

Parameter	Description
Base and top altitudes	Cloud boundaries in km above MSL
N_{layers}	Number of cloud layers detected
$N_{profiles}$	Number of 1-min profiles used in average profile
Method of cloud detection	GCDM or UCDM
Cloud phase	Ice or liquid/mixed phase based on -37 °C threshold
Estimated cirrus COD	Calculated using lidar ratios of 20 and 30 sr, ranging from 0.005 to 3
Base and top temperatures	From FP-IT GEOS-5 Version 5.9.1
Base and top pressures	From FP-IT GEOS-5 Version 5.9.1
Base and top wind speeds	From FP-IT GEOS-5 Version 5.9.1
Base and top wind directions	From FP-IT GEOS-5 Version 5.9.1
Day/night flag	0 = Night, 1=Day (from ephemeris)
Retrieval index	Indicates temporal average used to identify the cloud layer
Attenuation altitude	Altitude where lidar beam is deemed to be significantly attenuated

735

TABLE 3. Summary of data collected at GSFC in 2012 (percentages shown in parentheses)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Profiles recorded	44615 (100)	41690 (100)	44611 (100)	43142 (100)	32095 (72)	41479 (96)	40817 (91)	44485 (100)	43078 (100)	29081 (65)	42946 (99)	44345 (99)	536999 (93)
Observable profiles	38022 (85)	38226 (92)	37303 (84)	39971 (93)	25870 (81)	38138 (92)	35339 (87)	39318 (88)	38104 (88)	23412 (81)	41265 (96)	35175 (79)	468165 (87)
V3 normali-zations	34798 (78)	34882 (84)	35248 (79)	36090 (84)	23644 (74)	35801 (86)	34303 (84)	39408 (89)	37158 (86)	20649 (71)	36224 (84)	31041 (70)	399246 (74)

736 TABLE 4. Number of lidar profiles and occurrence frequency at GSFC in 2012

V2	V3b	V3m
	Low clouds	
77943	78511	78513
0.181	0.183	0.183
	Middle clouds	
69612	75353	75462
0.162	0.175	0.175
	High clouds	
109366	120750	134514
0.254	0.281	0.313
	Total clouds	
246132	255079	267777
0.572	0.593	0.623

737

738 TABLE 5. Transparent cirrus cloud properties

	Spring (MAM)	Summer (JJA)	Fall (SON)	Winter (DJF)	Annual
Cloud base					
Height (km)	10.46 ± 1.44	11.17 ± 1.48	9.77 ± 1.33	9.13 ± 1.49	10.17 ± 1.63
Temperature (ºC)	-50.3 ± 9.7	-47.9 ± 9.2	-43.9 ± 8.9	-45.7 ± 10.2	-47.4 ± 9.9
Pressure (mb)	262.6 ± 57.6	243.9 ± 52.0	291.0 ± 53.7	312.4 ± 66.8	275.9 ± 63.7
Wind speed (m s^{-1})	23.3 ± 12.5	20.3 ± 11.0	26.6 ± 10.7	38.6 ± 14.5	27.1 ± 14.2
Wind direction (º)	277.2 ± 40.2	236.8 ± 110.3	259.4 ± 28.3	279.1 ± 20.5	264.6 ± 63.3
Cloud top					
Height (km)	11.39 ± 1.35	11.93 ± 1.34	10.74 ± 1.16	10.09 ± 1.33	11.07 ±1.48
Temperature (ºC)	-57.0 ± 8.1	-53.5 ± 8.0	-51.3 ± 7.3	-52.9 ± 8.4	-54.1 ± 8.3
Pressure (mb)	225.9 ± 47.0	215.4 ± 41.6	249.9 ± 41.2	267.6 ± 52.4	238.4 ± 50.5
Wind speed (m s^{-1})	24.5 ± 12.9	21.8 ± 11.8	29.6 ± 11.4	42.5 ± 15.6	29.3 ± 15.4
Wind direction (º)	277.2 ± 37.4	238.3 ± 108.0	259.5 ± 29.6	279.3 ± 20.6	265.0 ± 61.7
Cloud depth (km)	0.93 ± 0.61	0.76 ± 0.51	0.97 ± 0.70	0.96 ± 0.69	0.90 ± 0.63
Estimated COD	0.07 ± 0.10	0.06 ± 0.10	0.09 ± 0.14	0.09 ± 0.12	0.08 ± 0.11
Cirrus Type (%)					
Sub-visual	47	51	45	41	46
Thin	50	46	49	53	49
Opaque	3	3	6	6	5
Occurrence (%)	17	12	10	13	13
Cloud layers	18664	13889	10769	14608	57930

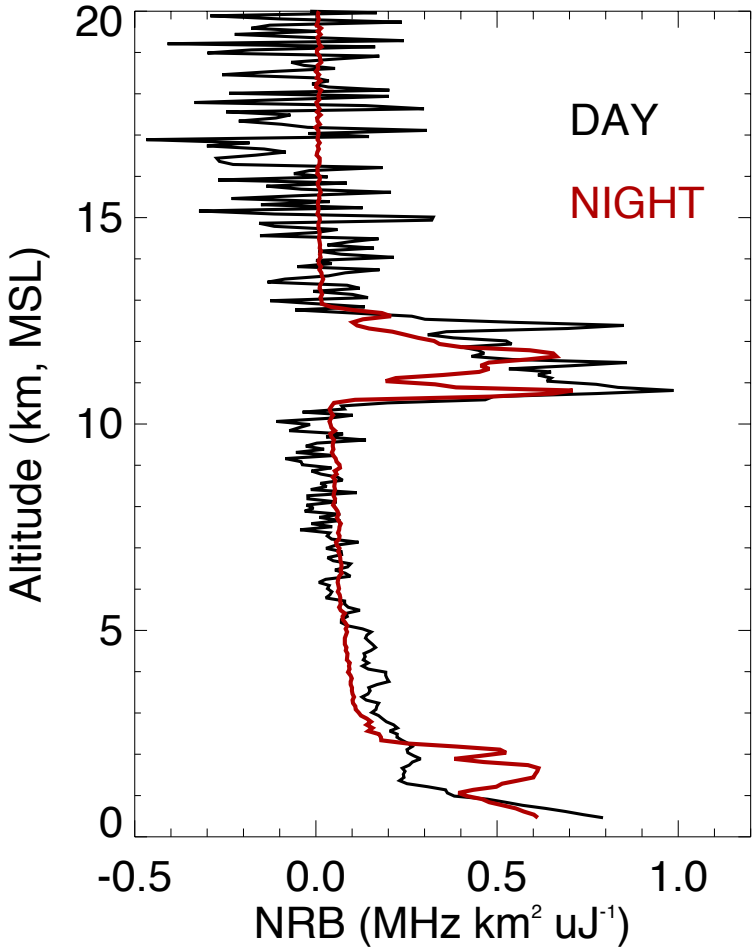
739 TABLE 6. Daytime and nighttime transparent cirrus properties

	Daytime	Nighttime
Cloud base		
Height (km)	10.15 ± 1.69	10.19 ± 1.58
Temperature (ºC)	-47.3 ± 10.0	-47.5 ± 9.8
Pressure (mb)	277.2 ± 65.8	274.9 ± 61.9
Wind speed (m s^{-1})	27.1 ± 14.0	27.0 ± 14.5
Wind direction (º)	258.6 ± 70.4	269.6 ± 56.2
Cloud top		
Height (km)	10.94 ± 1.51	11.19 ± 1.44
Temperature (ºC)	-53.2 ± 8.4	-54.8 ± 8.1
Pressure (mb)	243.5 ± 51.4	234.0 ± 49.3
Wind speed (m s^{-1})	29.1 ± 15.1	29.5 ± 15.6
Wind direction (º)	259.8 ± 68.3	269.4 ± 55.2
Cloud depth (km)	0.79 ± 0.56	1.00 ± 0.67
Estimated COD	0.07 ± 0.11	0.09 ± 0.12
Cirrus Type (%)		
Sub-visual	51	42
Thin	46	52
Opaque	3	6
Occurrence (%)	12	14
Cloud layers	26360	31570

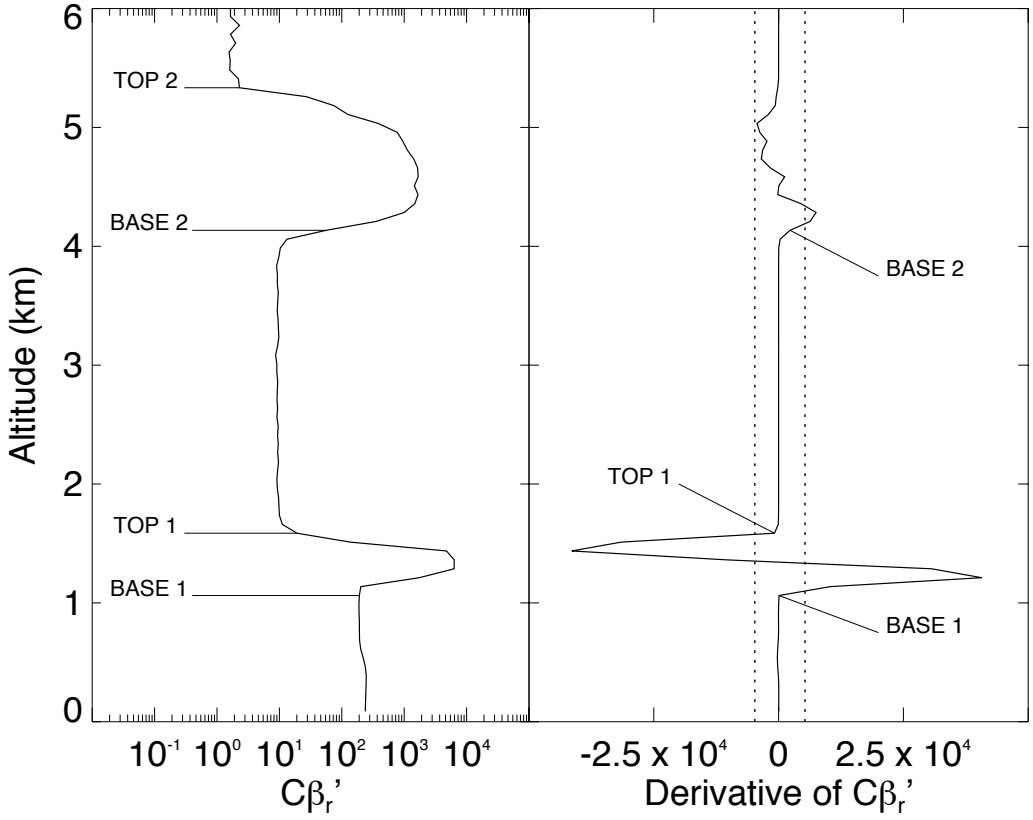
List of Figures

is 0.5 km for base and top altitudes and 0.25 km for cloud depth. (d) Frequency distribution (black) and cumulative frequency distribution (red). The vertical dashed lines indicate the thresholds for sub-visual (0.03) and thin (0.3) cirrus clouds. Horizontal axis bin size equals 0.001.

741



742
743 FIG. 1. Examples of the NRB at GSFC on 7 July 2012 at day (1800 UTC, black) and night (0300
744 UTC, red).



745
746 FIG. 2. Representation of the GCDM (GSFC; 0233 UTC 7 Nov 2012). The product of the
747 calibration constant and attenuated scattering ratio is shown in the left panel and its derivative is
748 shown on the right. The vertical dashed lines represent the a_{min} and a_{max} thresholds.

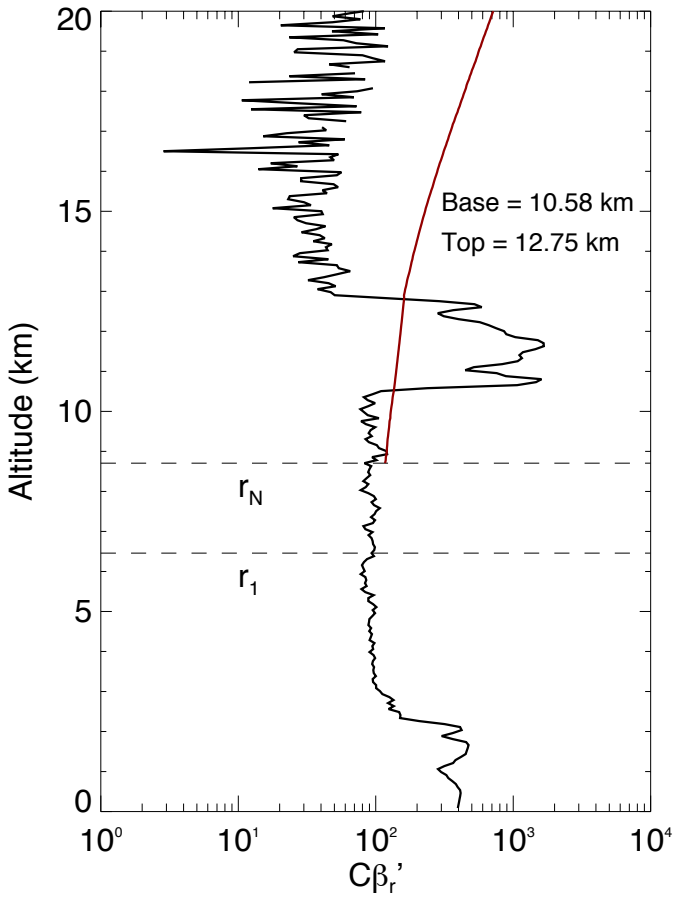


FIG. 3. Representation of the UCDM (GSFC; 0300 UTC 7 July 2012). The bottom (r_1) and top (r_N) of the normalization region are indicated by the horizontal dashed lines. The red curve represents the product of C_f^* and the minimum detectable scattering ratio used to determine the cloud boundaries.

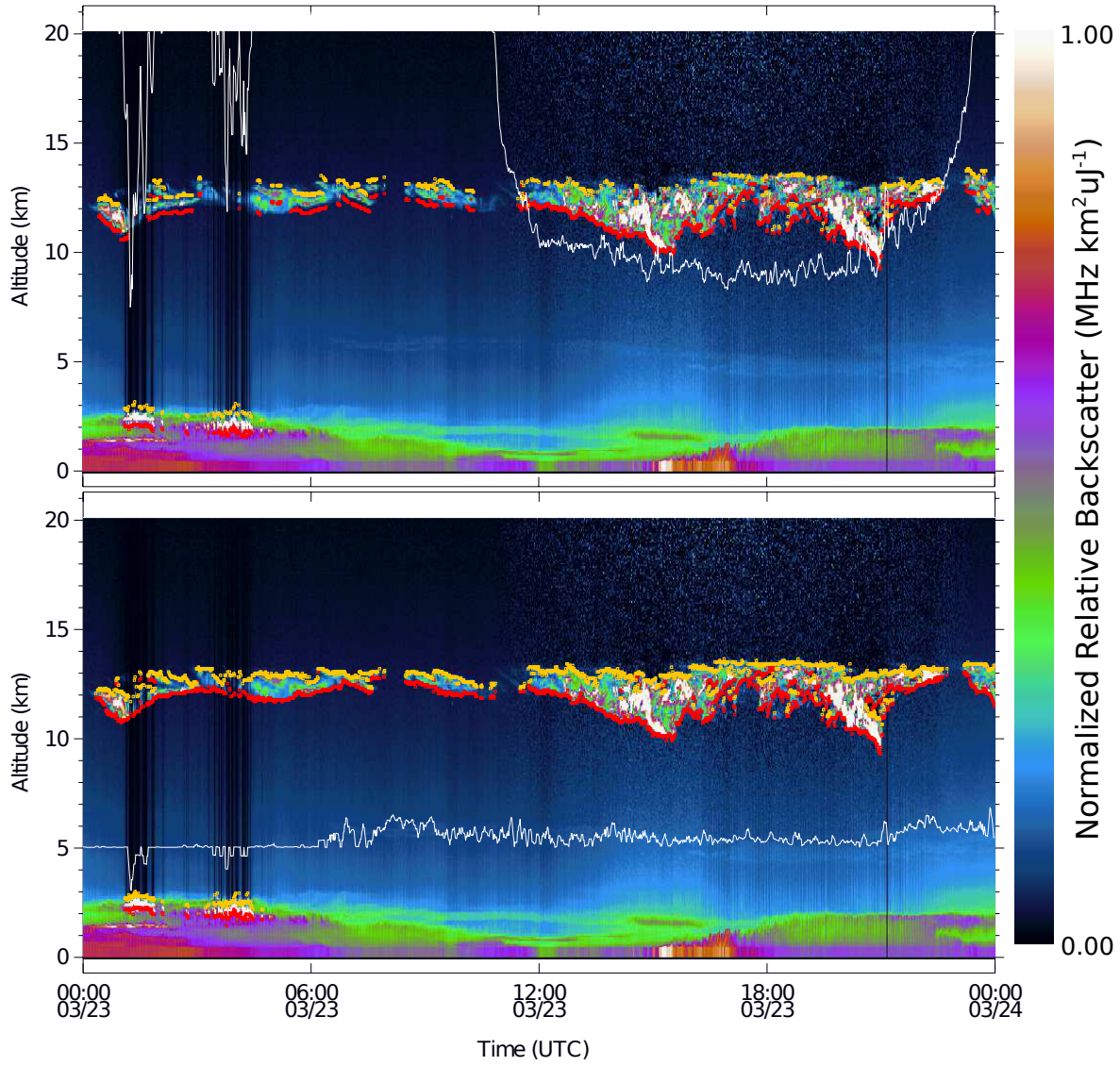
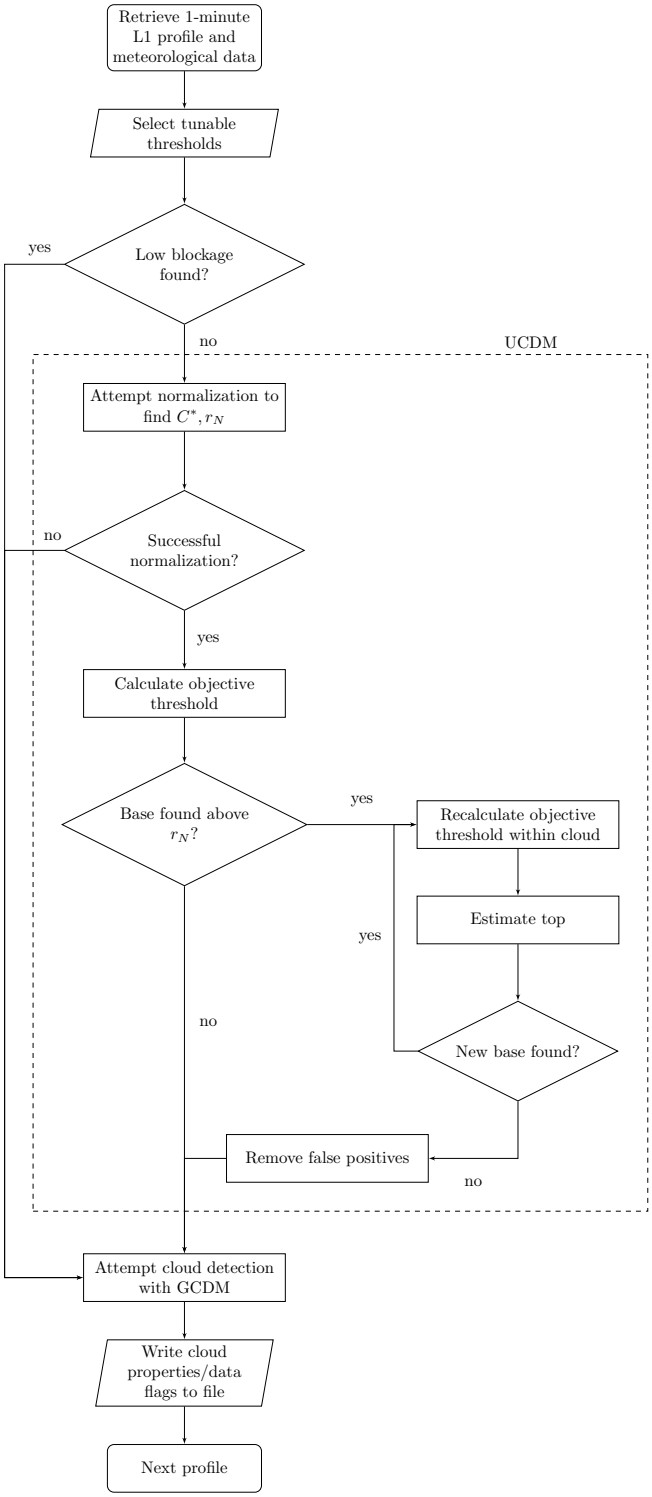
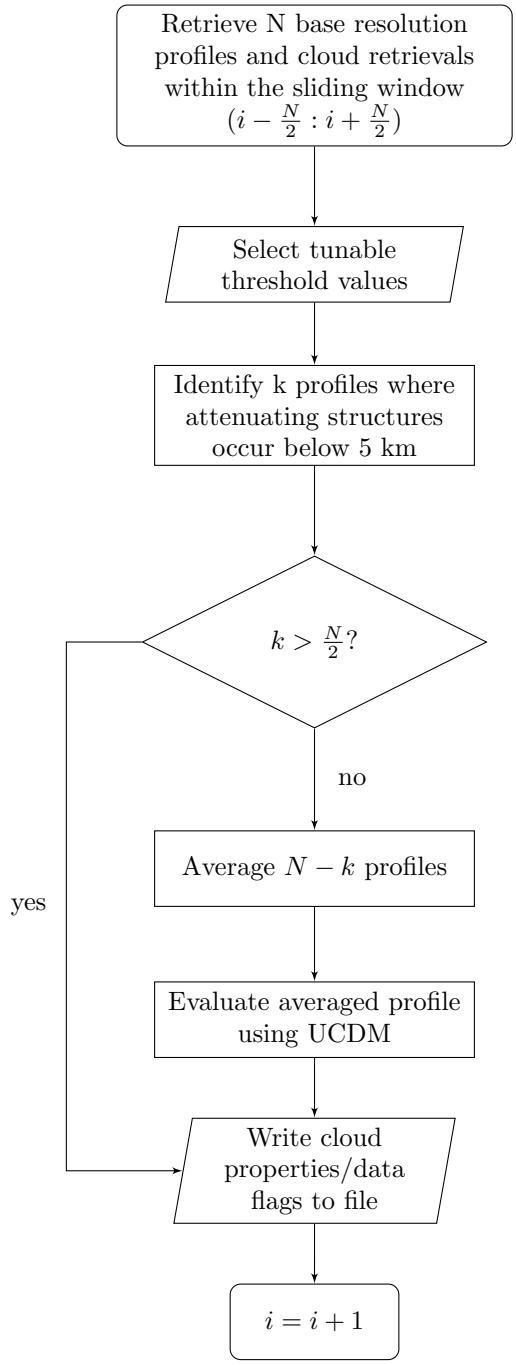


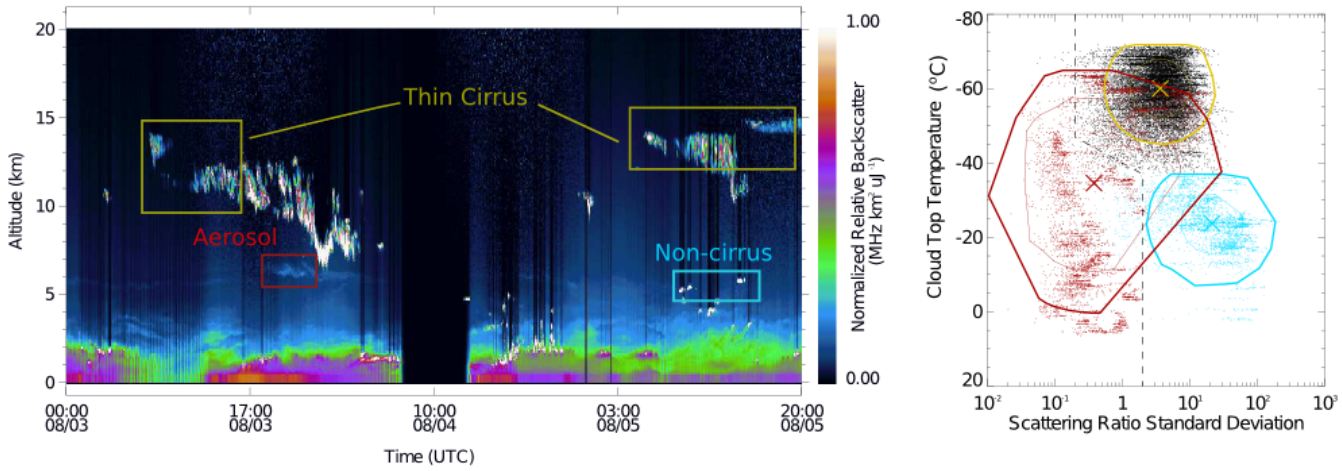
FIG. 4. (Top) Version 2 cloud retrievals at GSFC site on 23 March 2012 (one-minute resolution). Cloud bases and tops are indicated by red and orange markers, respectively. The noise altitude (i.e. δNRB/NRB > 0.5) is represented by the solid white line (four-minute smoothing applied for clarity). (Bottom) Same as top figure but using the V3 algorithm at one-minute resolution. The bottom of the normalization region is represented by the solid white line (four-minute smoothing applied for clarity).



761
762 FIG. 5. Schematic of the V3 cloud detection algorithm at the one-minute base resolution.

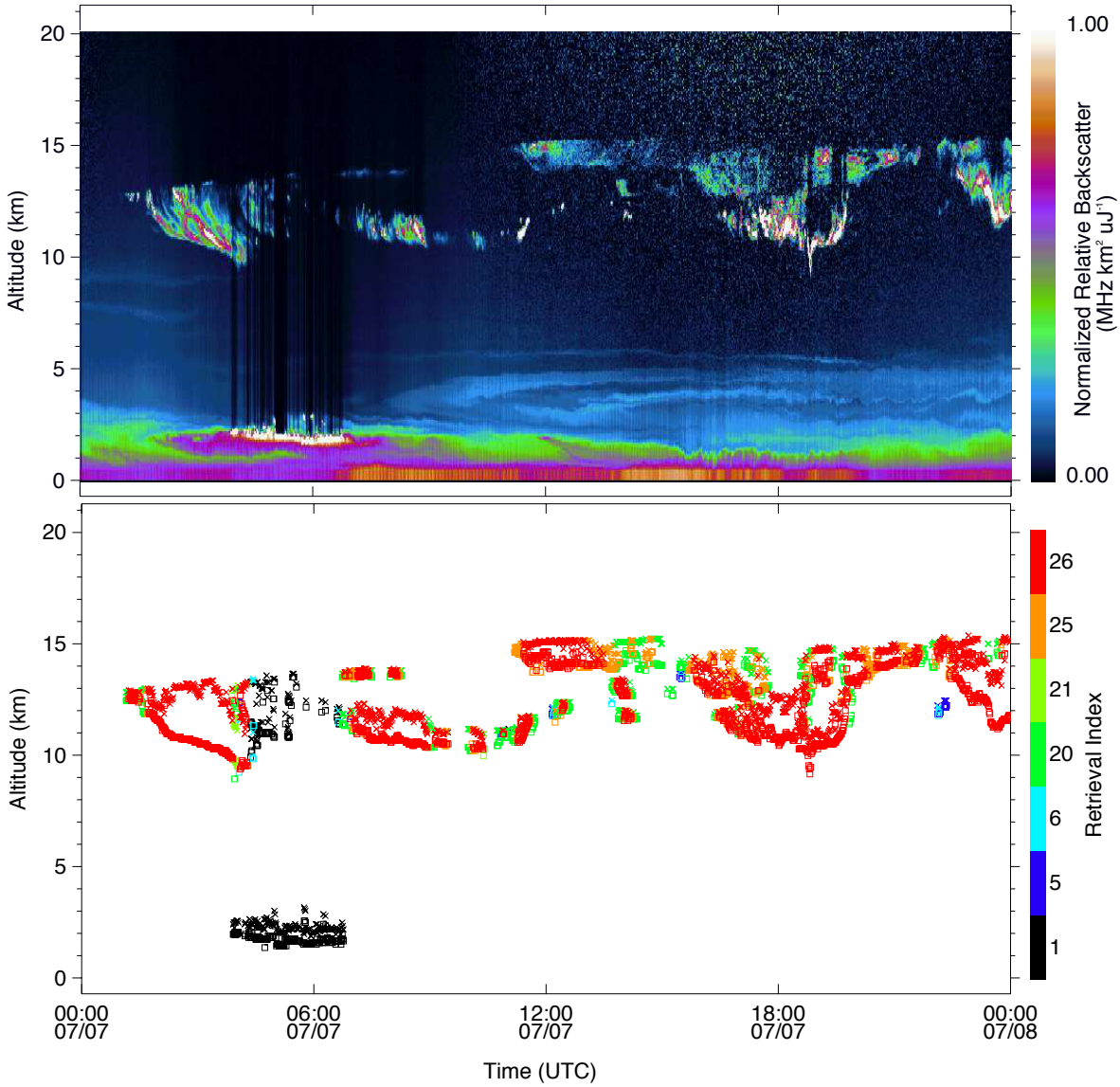


763
764 FIG. 6. Schematic of the V3 cloud detection algorithm for higher temporal averages.



765

766 FIG. 7. (Left) Representative groupings of cirrus clouds, non-cirrus clouds, and aerosols from
767 GSFC (3-5 Aug 2012). (Right) Scatterplot of full dataset used to determine σ_{min} threshold. The
768 Xs represent the median values of each group, the thin boundaries encloses 50% of the data
769 nearest the median and thick boundaries enclose 90% of the data. The dashed line indicates the
770 empirically-determined σ_{min}.



771
772 FIG. 8. (Top) NRB from lidar measurements at the GSFC MPLNET site on 7 July 2012.
773 (Bottom) Retrieval indices showing the temporal resolutions used to detect the combined cloud
774 scene. Square symbols are used to indicate the cloud base and the Xs are used for the cloud top.

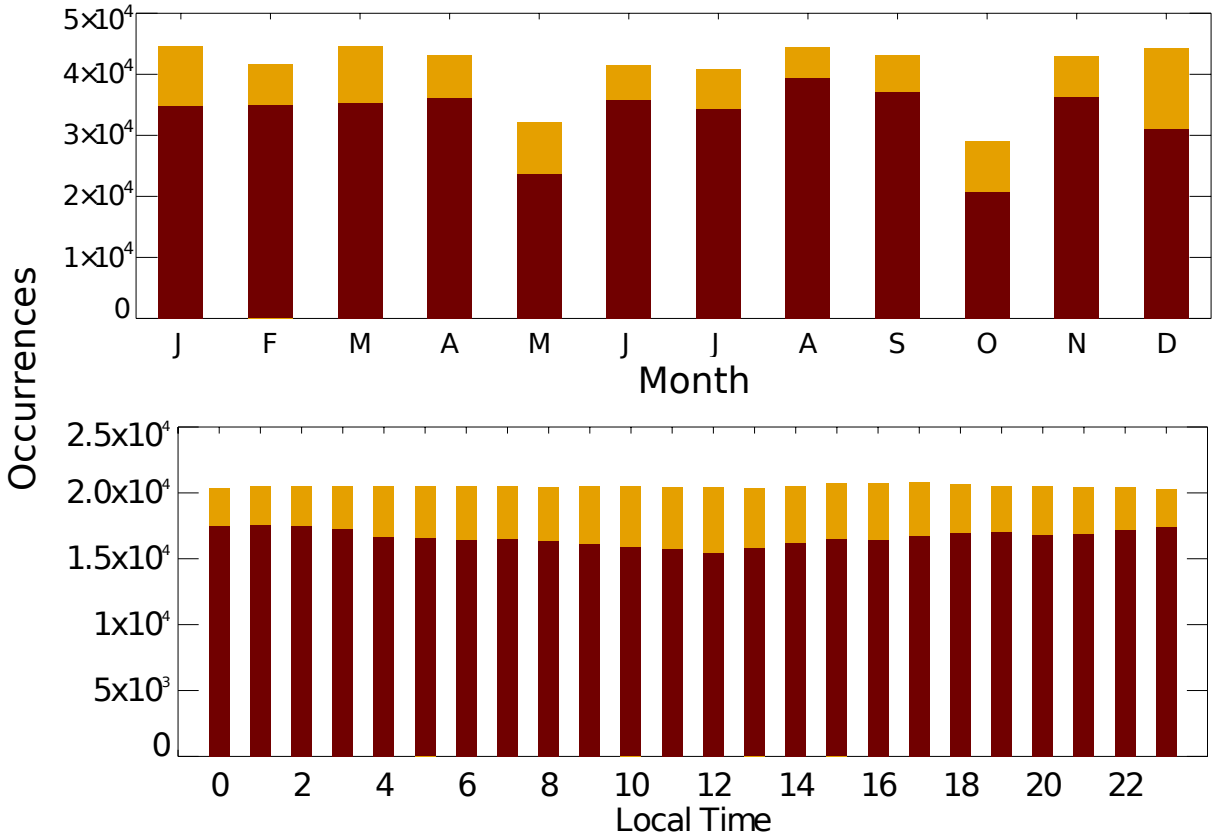
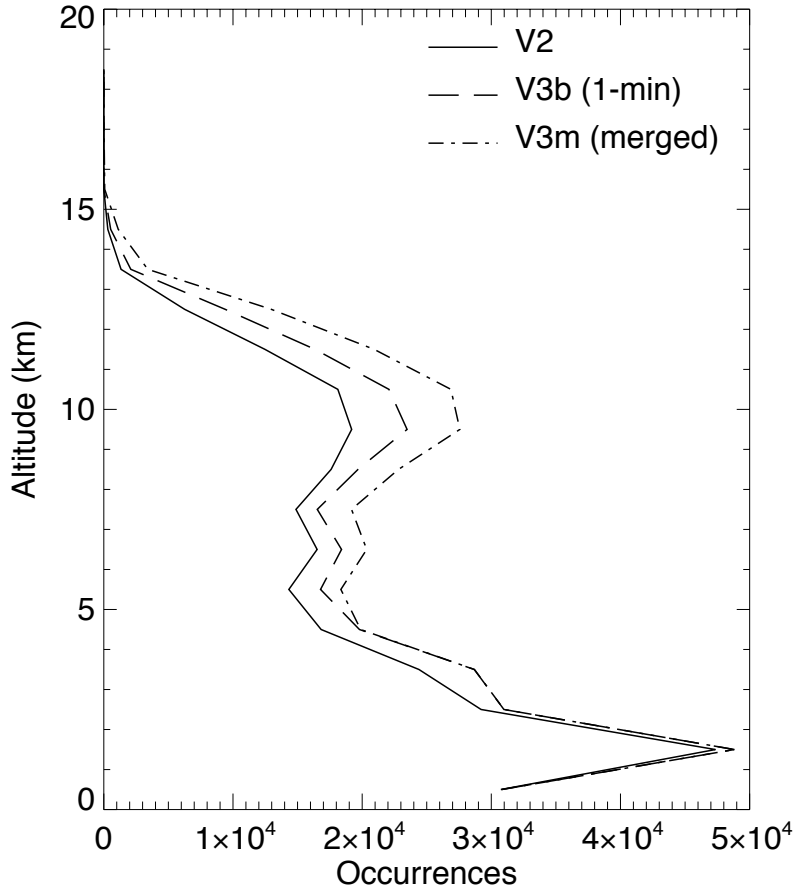


FIG. 9. (Top) Number of minutes when data was recorded by month. (Bottom) The number of minutes data was recorded each hour. The red colors are the successful C_f^* normalizations and yellow are the total data.



779
780 FIG. 10. Cloud base height distributions for GSFC during 2012 for V2 (solid line), V3b (dashed
781 line) and V3m (dash-dotted line). Vertical axis bin size equals 1 km.

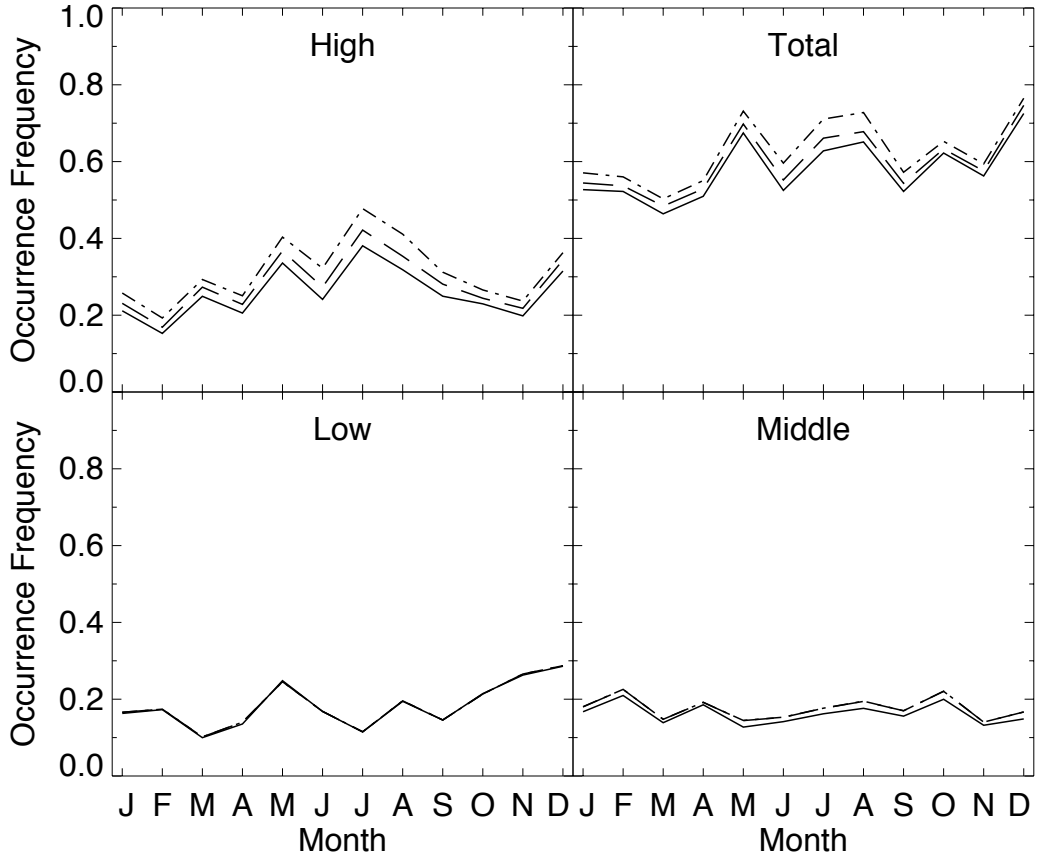
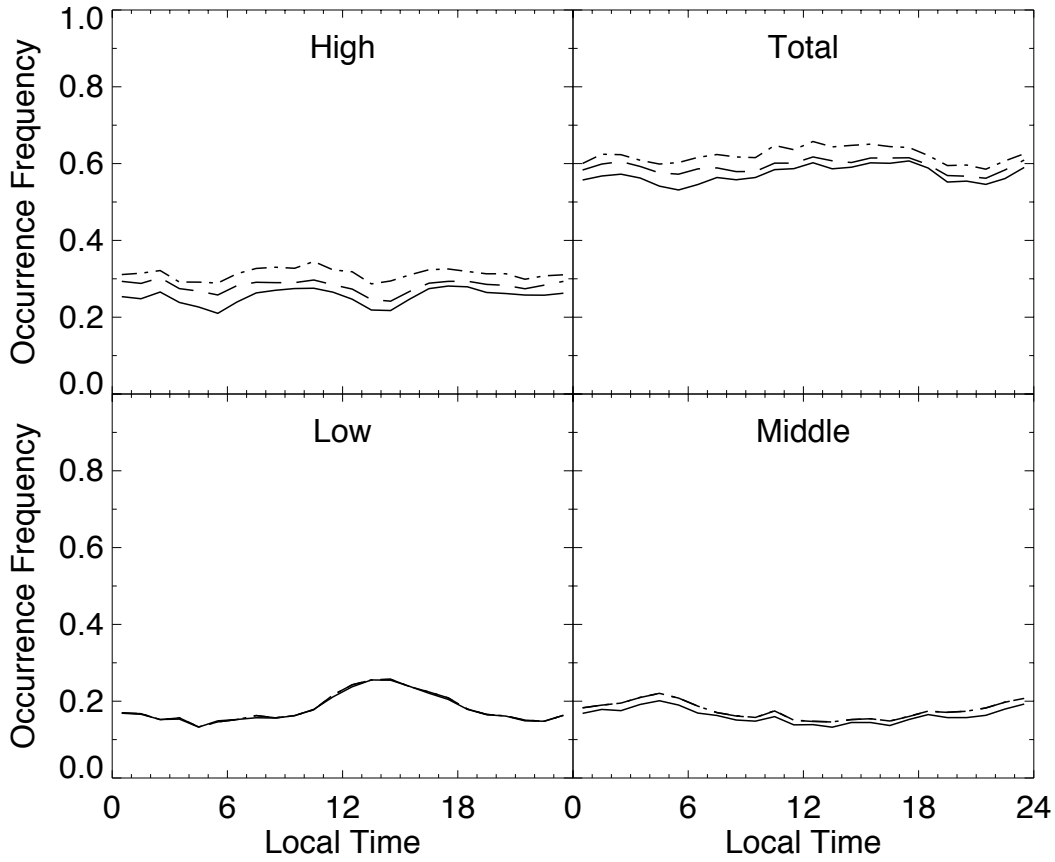
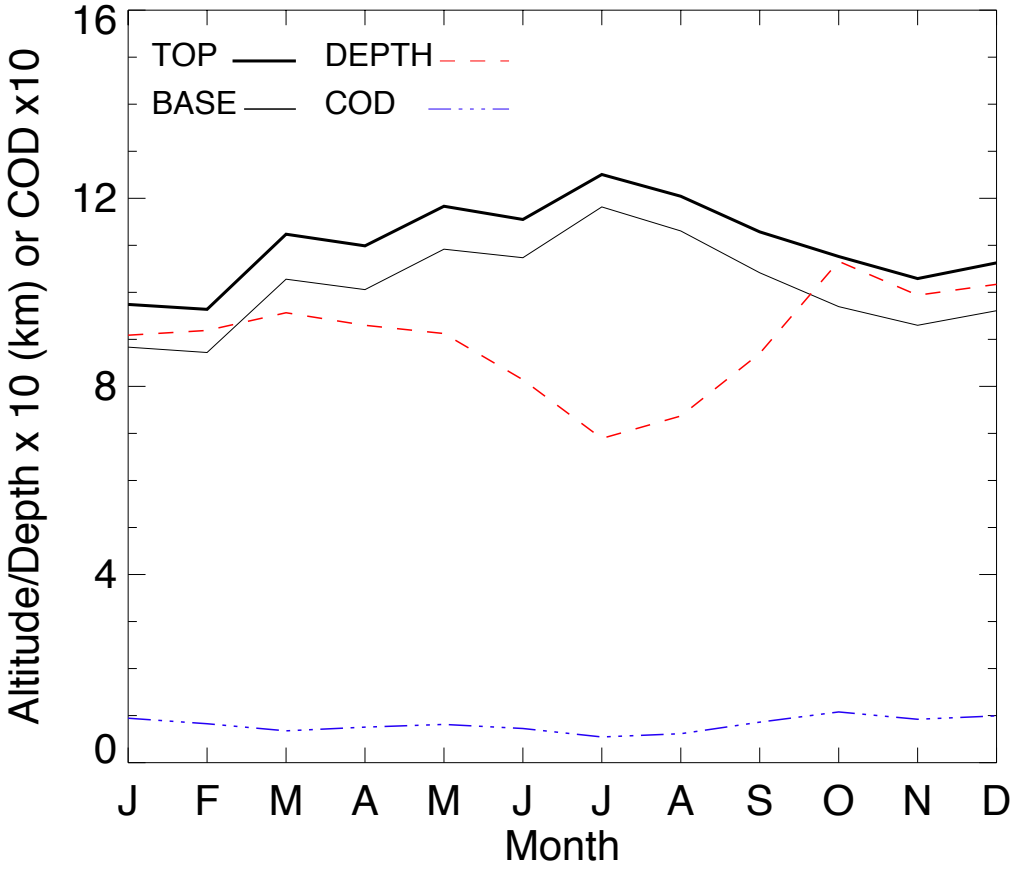


FIG. 11. Annual cycle of the occurrence frequency for low, middle, high and total clouds at GSFC during 2012. The solid line represents V2, the dashed line V3b, and the dash-dotted line V3m.



786
787 FIG. 12. Diurnal cycle of the occurrence frequency for low, middle, high and total clouds at
788 GSFC during 2012. The solid line represents V2, the dashed line V3b, and the dash-dotted line
789 V3m.



790
791 FIG. 13. Monthly averaged cloud top (thick solid line), cloud base (thin solid line), cloud depth
792 (dashed red line) and cloud optical depth (dashed-dotted blue line) for the transparent cirrus
793 dataset.
794

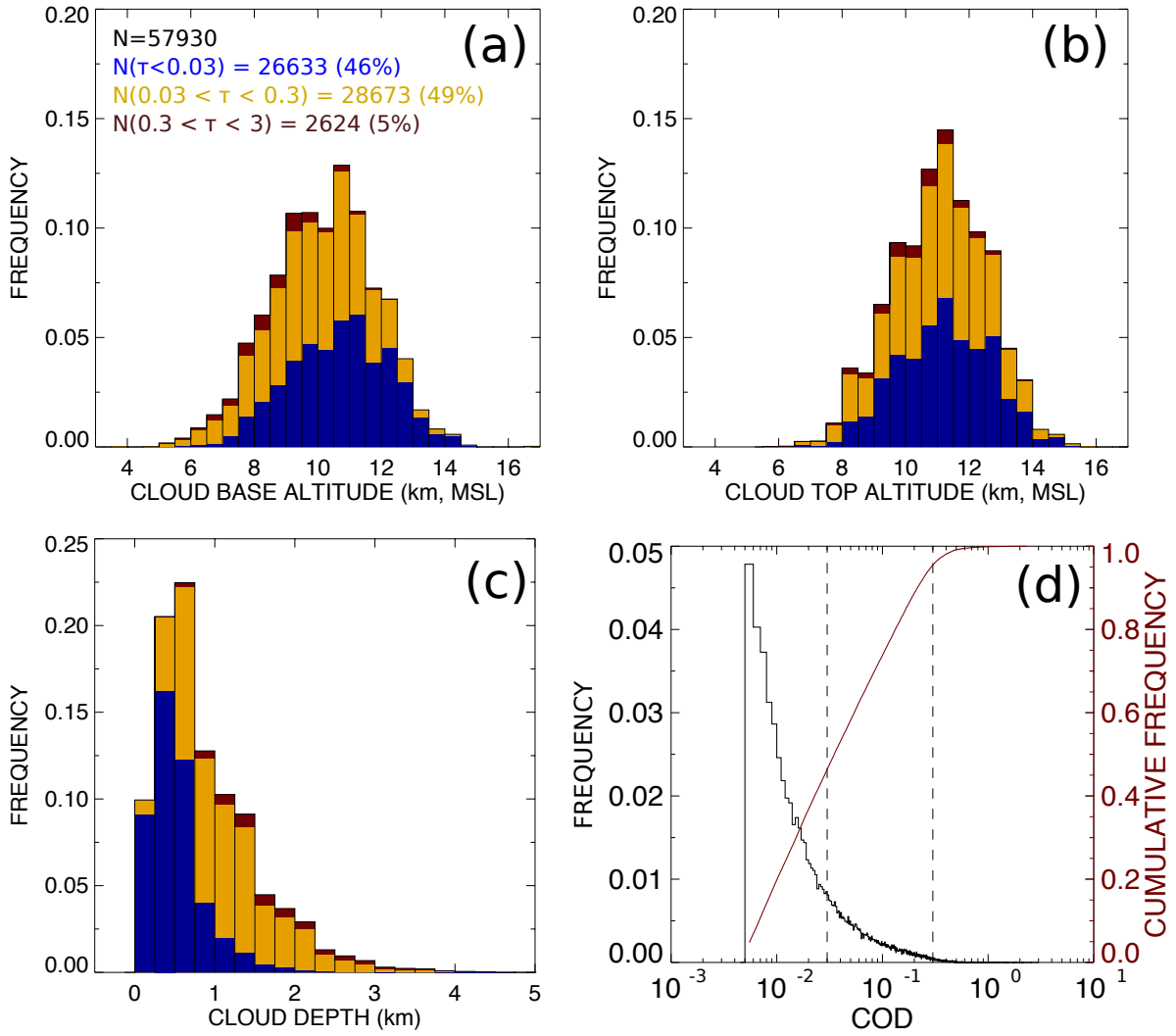


FIG. 14. (a) Cloud base altitude, (b) cloud top altitude and (c) cloud depth for the transparent cirrus dataset. The colors indicate the cirrus type based on the estimated COD (blue for sub-visual, yellow for thin, and red for opaque cirrus). Horizontal axis bin size is 0.5 km for base and top altitudes and 0.25 km for cloud depth. (d) Frequency distribution (black) and cumulative frequency distribution (red). The vertical dashed lines indicate the thresholds for sub-visual (0.03) and thin (0.3) cirrus clouds. Horizontal axis bin size equals 0.001.