Oceanic Weather Cloud-Top Height Quality Assessment Report

FAA Aviation Weather Research Program
Quality Assessment Product Development Team

3 March 2005

Contributors:

Barbara Brown[1,2], Eric Gilleland[1] , Robert Hueftle[1], Lacey Holland[1], Mike Kay [2,5], Sean Madine [3,4], Jennifer Mahoney[2,3], Agnes Takacs[1] ,and Jamie Wolff[1]

[1] Research Applications Program, National Center for Atmospheric Research, Boulder CO 80307-3000

[2] Corresponding authors: bgb@ucar.edu and Jennifer.Mahoney@noaa.gov

[3] NOAA Research – Forecast Systems Laboratory, Boulder, Colorado, 80305

[4] Joint collaboration with Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University, Fort Collins, CO

[5] Joint collaboration with Cooperative for Research in the Environmental Sciences (CIRES), Boulder, CO

Contents

Section **Page**

Oceanic Weather Cloud-top height Product
Quality Assessment Report

Quality Assessment Product Development Team

3 March 2005

Executive Summary

The Cloud-Top Height (CTOP) product is being developed by the Oceanic Weather Product Development Team (OW PDT) of the Federal Aviation Administration Aviation Weather Research Program (FAA/AWRP), and is currently being considered for transition to an experimental product through the Aviation Weather Technology Transfer (AWTT) process. The product, which diagnoses clouds above 15,000 ft, is intended to diagnose clouds that are hazardous to aviation.

This report summarizes the performance of CTOP during two periods, 12 February-23 April and 15-August-15 September 2004, for the oceanic domains Pacific, North Pacific, and Gulf of Mexico as defined by the OW PDT. For testing purposes, the OW PDT generated CTOP diagnoses over the Continental U.S. (CONUS) for the period 15 August – 15 September 2004. These experimental diagnoses were also evaluated and results are presented in this report.

The overall verification strategy accounts for the relative lack of direct observations of cloud top in general, and in the oceanic regions specifically. The strategy first employs an intercomparison of CTOP with a number of other inference-based cloud-top height alogorithms applied in the data-rich coastal, island and CONUS regions. The data sets include rawinsondes that are both land and ship based, WSR-88D radar echo tops (ET), and the NESDIS cloud-top pressure product (NCTP). After demonstrating the level of agreement between CTOP and the various other measures of cloud-top height, the strategy extends the evaluation to data-poor oceanic regions with a grid-to-grid comparison between the satellite products. In an attempt to delineate different cloud regimes, the comparison stratifies results by different effective cloud amount (ECA) ranges as inferred by the NESDIS product.

Results of the comparison in the data-rich areas indicate the CTOP is generally consistent with the cloud-top heights determined from ground-based observations under 30,000 ft. The results of the satellite comparison in the data-poor areas demonstrate very good agreement, with respect to values established by other cloud-top height validation studies, between CTOP and NCTP for opaque and thick clouds, particularly at the upper levels. The statistics for the thin cloud comparison show significant disagreement, an expected result given the theoretical capabilities of the products.

1. Introduction

This report summarizes the results from an evaluation of diagnoses produced by a Cloud-Top Height (CTOP) diagnostic product. The CTOP was developed by the Federal Aviation Administration Aviation Weather Research Program's (FAA/AWRP) Oceanic Weather Product Development Team (OW PDT) and is based on a combination of GOES IR emittance values with temperature and pressure profiles from the Global Forecast System (GFS) numerical weather prediction model. The goal of the CTOP diagnosis is to provide a mechanism for detecting cloud tops that are dangerous to the flow of air traffic over oceanic domains. The CTOP product is being considered by the Aviation Weather Technology Transfer (AWTT) Board for transition from a test product to an experimental product. The evaluation presented considers both the overall ability of CTOP to capture all cloud-top heights and those that pose a danger to aviation.

In support of the AWTT process, the Quality Assessment Product Development Team (QA PDT) was tasked with evaluating the quality of the CTOP algorithm. A major challenge for this project has been the development of verification methodologies that use datasets that are independent of those used to create the algorithm; identification of independent observations is particularly difficult in data sparse environments, such as over the oceans, but it is a critical aspect of verification studies. Therefore, the QA PDT has developed creative approaches for evaluating the CTOP product by extracting as much information as possible from the limited amount of global data, such cloud-top heights derived from radar, rawinsonde observations, and satellites. It is important to note that none of these data sources provides a "true" observation of cloud-top height, but rather an inferred value. Hence, the CTOP evaluation in essence can only provide a comparison of the various approaches, to evaluate the consistency of the CTOP in the context of the other measures. The period of investigation occurred from 12 February-23 April 2004 and 15 August-15 September 2004.

The report is organized in five sections as follows. The evaluation approach is presented in Section 2. Section 3 briefly describes the techniques for determining cloud-top heights. The verification methods are described in Section 4. Results of the study are presented in Section 5. Finally, Section 6 includes the conclusions.

2. Approach

The assessment techniques used to evaluate the CTOP are quite different from those used in previous QA PDT evaluations. Since cloud-top heights are only directly observed by lidar observations, which were not available for this evaluation, cloud-top heights were inferred from several observation data sets that are relatively (although not completely) independent of the data sets used to create the CTOP product, including Echo Tops (ET) derived from radar, cloud tops derived from rawinsonde observations (RCTP), and cloud-top pressure derived from satellite observations (NCTP). As a first step in the evaluation, the cloud-top heights produced using these various observations were intercompared with the CTOP product for limited areas and point locations. This intercomparison provided the framework and appropriate background for the grid-to-grid comparison between the CTOP product and a satellite-derived cloud-top pressure product (NCTP) over oceanic regions, where rawinsonde and radar data are limited or unavailable. Takacs et al. (2004 a,b) summarized the global observational datasets available for use in verification studies. These datasets are described in greater detail in the following sections.

The intercomparison and the NCTP/CTOP grid comparison were conducted for the periods 12 February – 23 April (depicted in the left hemisphere view in Fig. 1) and 15 August – 15 September 2004 (depicted in the right hemisphere view in Fig. 1). The oceanic domains: Pacific, North Pacific (Hawaii), and the Gulf of Mexico (GOMEX), and the CONUS are shown in Fig. 1. Issuance schedules vary for each of the derived cloud-top height products used in this study, thus the time matching methods (described in Section 4.1) were adjusted to account for the time differences of each domain.

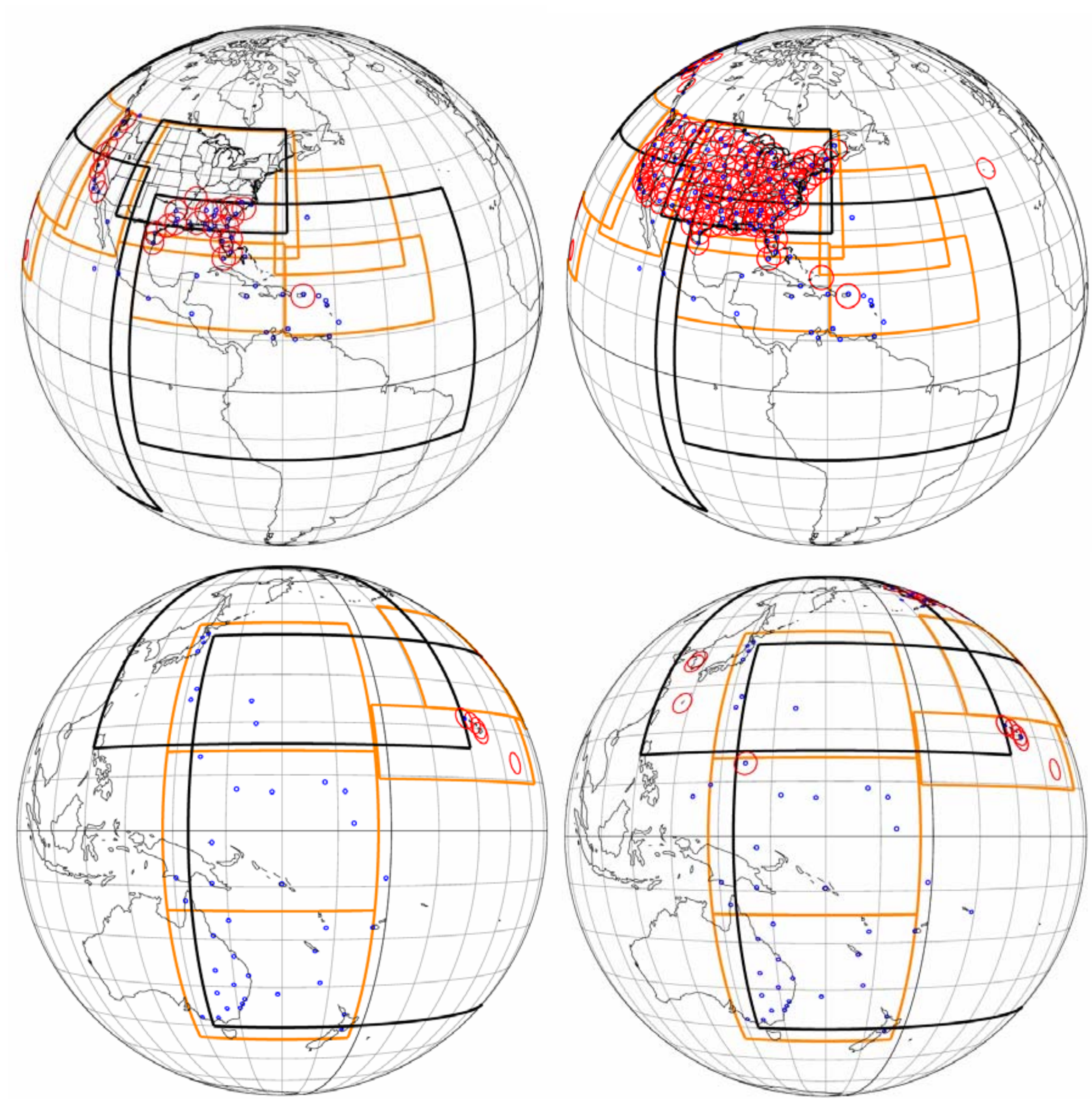


Figure 1. Overlap of CTOP (black), RCTP (blue), ET (red), and NCTP (orange) datasets over the globe. Left hemisphere 12 February-23 April 2004; right hemisphere 15 August-15 September 2004.

3. Techniques for Measuring Cloud-Top Height

Since direct measures of cloud-top height were unavailable for verification, cloud-top heights were derived from weather variables measured by radar, rawinsonde, and satellite observing instruments. The CTOP diagnostic and the derived cloud-top height techniques are described in this Section.

3.1 CTOP Diagnostic Product

The OW PDT utilizes the IR Window technique to create the CTOP product covering three oceanic domains, and for this evaluation only, the CONUS domain. This approach combines a brightness temperature, measured by the infrared window channel of the GOES Imager, with a temperature profile from the Global Forecast System (GFS) numerical weather prediction model to estimate the cloud height for a given pixel. An updated version of the procedure described to the OW PDT in a presentation created by Miller et al. (2002; http://www.rap.ucar.edu/projects/owpdt/documents/cloudtops2.html) follows:

- The geostationary IR data from GOES 9, 10,12 Imagers is ingested to create a "stitched" image over the domain of interest.
- The closest temporal match between the GOES Imager IR data and the GFS analysis over the same domain is determined.
- The intersection between each IR pixel in the domain of interest, and the GFS is determined by looping downward from the top of the atmosphere until intersection between the IR pixel and the GFS profile is achieved or the pressure level exceeds the 850 hPa cutoff.
- If an intersection is found, the GFS geopotential height value is interpolated to the pixel location and an estimate of cloud-top height is produced.
- An image representation of the cloud-top height is then produced.

The authors of the presentation also identified the following qualitative algorithmic cloud-top height detection strengths and weaknesses:

Strengths of the CTOP algorithm include the detection of

- Clouds over the oceans, because the IR technique performs best with a warm stable background
- Clouds that are optically thick
- Cloud regions that are characterized by a well-behaved lapse rate and well-defined tropopause

Weaknesses of the CTOP algorithm include the detection of

- Clouds over land, because of the highly variable temperature background
- Clouds that are optically thin
- Cloud regions that are characterized by a strong mid-level inversion

Due to the varying availability of the GOES Imager coverage over the globe, the issuance times and intervals for the CTOP product differs for each of the domains used in this evaluation. Through OW PDT processing, the product is updated every 20 min for the Pacific domain, every 30 min for the Gulf of Mexico domain, every 15 min for the CONUS domain, and roughly every 3 h for the North Pacific domain. The CTOP product has a nominal resolution of 4 km, the same as the GOES Imager IR window channel scan.

3.2 Rawinsonde-derived Cloud-Top Height (RCTP)

The instrument package on a rawinsonde measures the temperature, moisture, pressure, and winds from the Earth's surface to the stratosphere. Rawinsonde observations for this study were available at 141 locations over the CONUS, coastal areas, and islands twice a day, at 0000 and 1200 UTC. In a few cases, 0600 and 1800 UTC soundings were also available. Since the cloud-top height is not directly measured, a technique developed by Wang and Rossow (1995) and Landolt et al. (2004) was used to derive the cloud-top height from those variables that are directly measured by the rawinsonde instruments. To reduce these inaccuracies, Wang and Rossow (1995) suggested calculating RH with respect to both water and ice. The technique determines the cloud top by examining the relative humidity from the top of the atmosphere to the surface with respect to water and ice. The cloud top is set at the highest level when the relative humidity with respect to water (RH_w) or ice (RH_i) either (a) exceeds 87% or (b) exceeds 84% and the level above had RH_w or RH_i that was at least 3% lower than RH_w or RH_i of the level in question.

The rawinsonde vertical sounding can overestimate the cloud-top pressures, which results in an underestimate of cloud-top height. This overestimation of cloud-top pressure (underestimation of cloud-top height) in extreme cases can reach 170 hPa (especially in the Tropics), and it occurs more frequently in regions where cloud tops are colder than -40^{o}C, as in deep convective environments (regions that are important in the evaluation of the CTOP product). Landolt et al. (2004) compared satellite and sounding-derived cloud-top temperatures, and found that the Wang and Rossow technique for identifying cloud layers and cloud top temperatures works well for relatively thick clouds, but often fails when broken or thin cloud layers are present. Cloud thickness was a key factor for the accuracy of the method. Given the characteristics of clouds over ocean and land-surface areas described in Wang et al. (2000), the RCTP and CTOP should show closer agreement over oceanic areas.

When the Wang and Rossow technique was applied, the drift of the rawinsonde was estimated to provide more accurate measurement of the location of the cloud-top height. In particular, it usually takes one and a half to two hours for the ground receiving system to obtain all of the information from a rawinsonde system. Due to the drifting, the true location and time of the cloud-top height based on the rawinsonde observations can be far from the place and time of the launch of the balloon (Fig. 2). For this purpose, a 5.5 m/s ascension rate for the rawinsonde was used, as applied in other studies (e.g., Roy et al. 2004). The sum of the drift from the starting point at the rawinsonde station to a given level (e.g., at an estimated cloud-top height) provides a measure of the horizontal drift from the station's location to the point at the given level. By determining the drift, the Wang and Rossow (1995) and Landolt et al. (2004) methods were modified to allow for a more precise determination of the cloud-top height location.

In addition to land-based rawinsondes, Navy ship soundings were used. Like ground-based rawinsonde, the Wang and Rossow (1995) method was used on this dataset, after drift was taken into account. Figure 3 shows the global distribution of ship soundings over the Pacific and Gulf of Mexico.

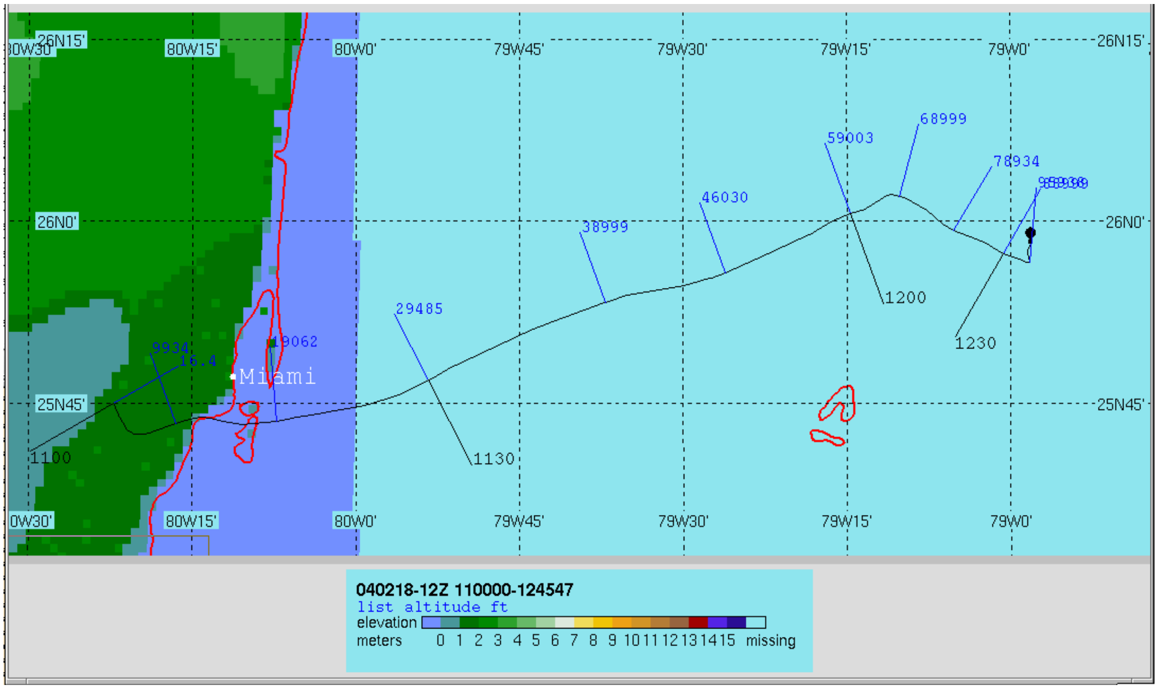


Figure 2. Example of the drift of a rawinsonde launched from Miami at 1100 UTC on 2/18/2004. Black and blue numbers connected to the drift path indicate the time (UTC) and altitude (ft), respectively.

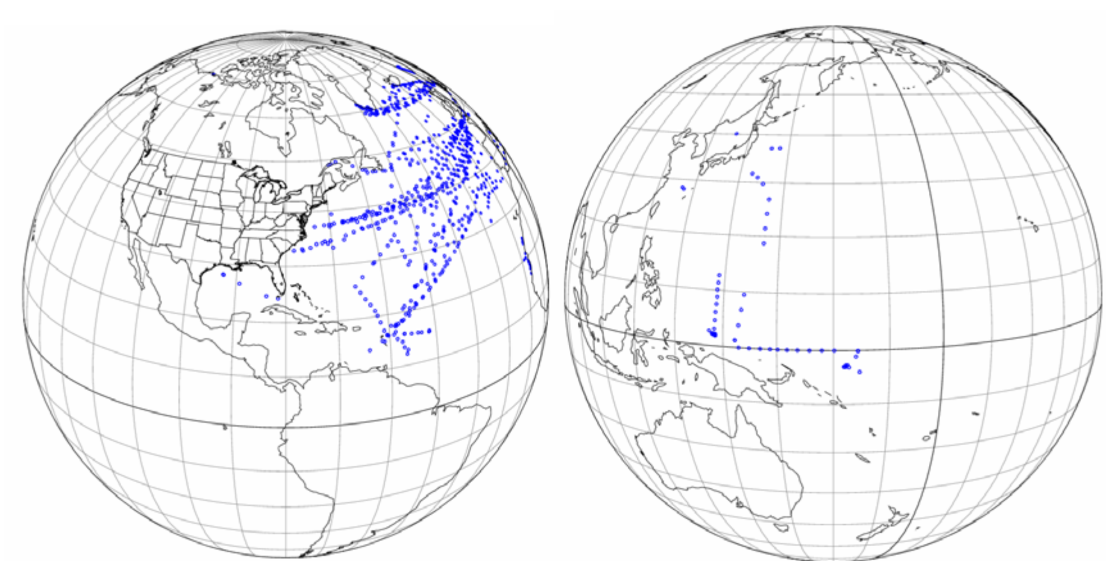


Figure 3. Location of Navy ship soundings over the globe.

3.3 Radar-derived Echo Top (ET)

By definition, the radar-derived echo top (ET) is the maximum vertical extent of precipitation-sized particles within a cloud and will always be less than or equal to the height of the cloud. The echo top height represents the top of the precipitation and is typically lower in height than the cloud top. The echo top measurement was used in this evaluation to provide a lower bound check to ensure that the CTOP heights are within the correct range. ET values are provided by the National Weather Service (NWS) WSR-88D radars and are measured in thousands of feet above mean sea level. The lowest detectable tops are those at 5,000 ft, while the highest detectable tops are at heights of 70,000 ft. The observations have a 4-km spatial and 6-min temporal resolution with a range of 230 km and a vertical resolution of 5,000 ft. Nearly 30 radars are located inside the verification domain in coastal areas and on islands.

The ET algorithm is constrained in several ways that reduce the vertical accuracy of the product. For instance, precision of echo top heights decreases with range due to beam broadening. At a range of 230 km, the beam (half power) is 4,000 m wide. In addition, the ET algorithm uses a threshold value of 18.5 dBZ to define echo tops, which does not correspond to what a pilot sees (i.e., pilots see clouds, whereas radars detect precipitation). Further reduction in height accuracy stems from the resolution of the mapping routine. That is, the highest of several possible echo top values above a 4x4-km box of 1-km resolution radar reflectivity is used in the ET product.

ET uncertainty is also associated with the scanning strategies, or volume coverage pattern (VCPs), that are used to detect and monitor convective storms. Since different VCPs have gaps between various elevation angles, there is uncertainty in echo top height in the neighborhood of the intersection of the center of the beam and the radar return (i.e., the radar does not "see" at all elevation angles). The new VCP 12 is the best for ET (personal communication with R.M. Steadham, NOAA Radar Operations Center). However, since each radar site is independently operated, the VCP is often different for each radar, thus possibly contributing to the height differences presented in this study. For more information on the details of the VCP used in this study, see Brown et al. (2000). The same study also indicated an optimum distance range from 45 to 120 km should be used to reduce the height uncertainty related to VCPs. At an elevation angle of 19.5 degrees, the beam would "undershoot" ET heights that are less than 45 km in range and greater than 50,000 ft in height. (Note that the WSR-88D radars never scan directly overhead so they are unable to detect the true top of the reflectivity echoes directly over the radar site.)

3.4 NESDIS Cloud-Top Pressure (NCTP) and Effective Cloud Amount (ECA)

This Section describes the characteristics of the NESDIS cloud products, which include the cloud top pressure (NCTP) and effective cloud amount (ECA) used for the grid-to-grid comparison with CTOP as well as the overall stratification of statistics. The validation studies summarized in the sub-sections provide acceptable cloud-top height comparison values that are used for interpreting the results of this evaluation.

The generation of the GOES Sounder-based derived cloud parameters, cloud top pressure and ECA is described by Schreiner et al. (2001). In this study, of the 77% of cloudy pixels examined, 55% were determined by the CO_2-slicing method and 45% by the IR window technique. The algorithm primarily relies on the CO_2-slicing technique, derived from radiative transfer principles, to determine cloud top pressure and ECA (Menzel et al. 1983; Wylie and Menzel 1989). In cases where the CO_2-slicing calculation fails due to the instrument noise (which typically occurs for very thin, high clouds or low, opaque clouds) the algorithm adopts the IR Window technique to determine the pixel cloud top pressure. A brightness temperature, measured by the GOES Sounder, provides the value for lookup in the GFS temperature profile. In these cases, the value for the effective cloud amount is set to 100%, a value never inferred by the CO_2-slicing technique.

Each pixel in the NCTP product has a nominal resolution of 10 km. When inferred by the CO_2-slicing approach, the assigned cloud-top height value, consists of the single pixel value while the assigned ECA consists of a 3x3 pixel averaged value. The maximum cloud top pressure value for the NCTP is either 150 hPa, which is roughly 45,000 ft in the standard atmosphere, or the tropopause, whichever height is lower in the atmosphere. The product covers the domains viewed by the Sounder instrument on GOES-9 (GMS replacement), GOES-10 (West), and GOES–12 (East), as is shown in Fig. 1.

The remote sensing community has generally accepted the CO_2-slicing algorithm as useful for determining cloud top pressure and effective cloud amount for clouds above 600hPa (Zhang and Menzel 2002). The technique, however, is known to have difficulty detecting the following types of clouds:

- Optically thin cirrus clouds (ECA < 10%)
- Multi-layered clouds (e.g., transmissive cloud above a lower opaque cloud)
- Low-level clouds (signal-to-noise problem)
- Clouds existing in an isothermal atmosphere (e.g. polar regions)

Figure 4 allows qualitative comparison of the CTOP and NCTP products with a GOES visible image taken at nearly the same time. The images, from 22 February 2005, highlight a typical situation over the Gulf of Mexico where a variety of cloud types are present. Of great interest to aviation, the area of deep convection over the east-central part of the Gulf is embedded in a variety of other clouds including a large canopy of cirrus of varying thickness. Overall, NCTP and CTOP show similar cloud-top heights in this region. However, notable differences between the products become apparent outside of the immediate areas of deep convection and in the surrounding cirrus clouds. The NCTP shows a much more continuous, albeit coarse, field of cirrus, evident in the visible image, while CTOP detects only a limited region of high cloud. Similar observations apply in the western Gulf where scattered deep convection surrounded by a large area of cirrus exists. These examples illustrate the known strengths and weaknesses of the IR Window algorithm used by CTOP, and the CO_2-slicing technique used primarily by NCTP.

3.4.1 Validation Studies of the NESDIS Cloud-Top Pressure (NCTP) Product

Since the NCTP product is the dataset that is used for the overall grid-to-grid comparison of the CTOP product, it is important to understand the validation studies that have been done to determine the accuracy of the NCTP product. The developers of the NCTP have studied the performance of the product as it has matured into operations. The studies do not stratify comparison results based on the technique used by the algorithm (i.e., CO_2-slicing or IR Window techniques) to compute the cloud top pressure pixel. In addition, the algorithm may perform differently in land and ocean domains, a perspective not examined in these reports. Summaries of the studies presented in this section should be interpreted with that understanding. The various validation measures are intended to approximate “acceptable” values for the comparison of CTOP and NCTP.

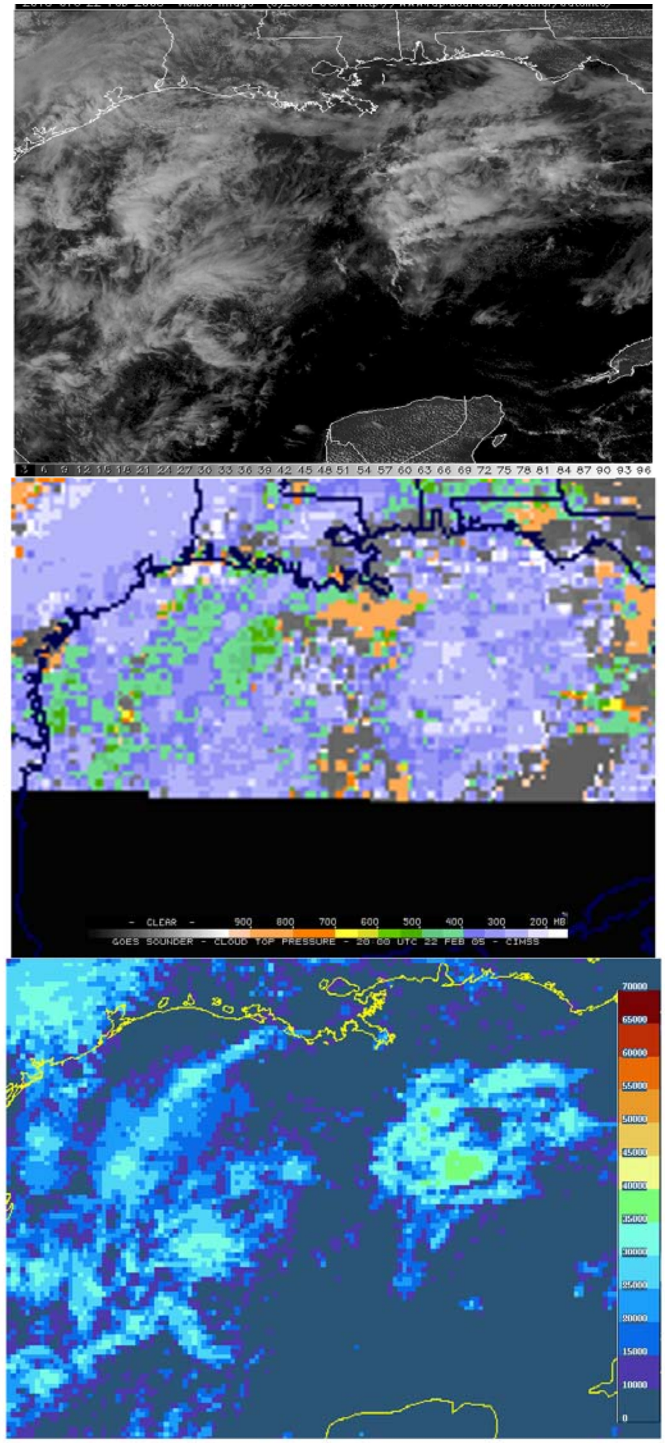


Figure 4. Qualitative comparison of three cloud product images in a small region of the Gulf of Mexico – GOES Visible (top), NCTP (middle), and OW PDT CTOP (bottom). Nominal resolutions are 1 km, 10 km, 4 km respectively

One validation study compared pilot reports (PIREPs) of cloud-top height to NCTP (Schreiner et al. 2001). The analysis, using a relatively small sample of observations, found NCTP to be on average ~3,100 ft lower than the PIREPs at flight levels between 15,000 and 35,000 ft. The comparison also determined a correlation coefficient of 0.83 for the same set of observations. Another study by the same group compared cloud-top height measurements from a ground-based combination Micropulse Lidar/Millimeter Cloud Radar (MPL/MMCR) (Hawkinson et al. 2001). After identifying and filtering incidents of multiple cloud layers, the analysis yielded a bias of ~3000 ft (low) and a correlation coefficient of 0.93. Concerns about these results include the observation filtering, which is a process not applied to the intercomparison in this report, and the inclusion of data below the flight levels of interest. The results are, however, consistent with the PIREP-based comparison.

3.4.2 Validation Studies of the CO_2-slicing Algorithm and Effective Cloud Amount

Frey et al. (1999) performed a validation study to assess the accuracy of the CO_2-slicing cloud height algorithm. Comparisons were made for cloud-top height values inferred by the CO_2-slicing technique utilizing data from the MODIS (Moderate-Resolution Imaging Spectroradiometer) Airborne Simulator (MAS) and those values provided by the Cloud Lidar System (CLS). The study, geared toward understanding detection of mid- and upper-level transmissive clouds, primarily evaluated values for clouds with effective cloud amounts between 30% and 60%. Assuming the difference distribution provided by the study to be approximately normal, a calculation estimates a mean absolute difference of ~2,000 ft. The NCTP cloud heights inferred by the CO_2-slicing technique utilizing data from the GOES Sounder– an older instrument than MODIS – would be expected to show a larger mean absolute difference.

An earlier study compared the CO_2-slicing inferred cloud heights utilizing data from the Visible-Infrared Spin Scan Radiometer Atmospheric Sounder (VAS), the predecessor to the GOES Sounder, to values determined by a variety of other techniques (Wylie and Menzel 1989). Studies of cloud characteristics often use rawinsonde observations as well as other ground-based observations (Menzel et al 1983). Comparison with rawinsonde inferred pressure, mostly consisting of reports of opaque or nearly opaque clouds from 800 hPa to 300 hPa, showed an average deviation of 40 hPa. Fig. 5 shows a representation of the width, in feet, of a 50 hPa layer as a function of pressure. A 50 hPa layer centered on 850 hPa is much smaller than the corresponding interval at 200 hPa. Thus, care must be used when interpreting errors expressed in hPa from those using height as the unit of measure given the rapid increase in layer depth as pressure decreases. In a comparison that included significantly more cirrus observations, the VAS CO_2-slicing derived pressures were on average 70 hPa greater than (i.e, lower in the atmosphere) those measured by lidar. Finally, comparison with stereo parallax measurements derived from GOES-East and GOES-West visible images, when restricted to single-layer cirrus clouds, demonstrated a bias of 25 hPa. This part of the study found that, in general, the VAS CO_2-slicing technique underestimated cloud heights above 400 hPa and overestimated them below that level.

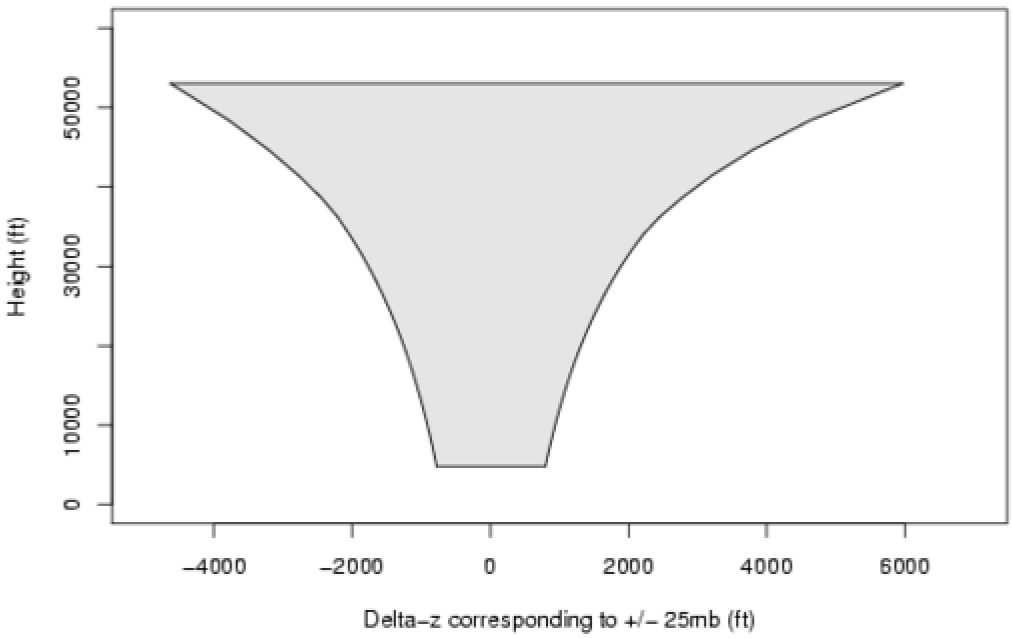


Figure 5. A representation of the width, in feet, of a 50 hPa layer as a function of pressure

The Wylie and Menzel (1989) study also compared VAS derived estimates of ECA to a statistical cloud cover climatology compiled from 11 years of land-based cloud observations (Warren et al. 1986). Differences between the datasets, revealed by the analysis, were traced to pixels with small partial coverage or very thin clouds. After filtering these cases, the authors concluded there was good agreement between the ECA measurements and the cloud cover climatology. The study also noted that up to 5% of the opaque (ECA between 95% and 100%) cloud reports were possibly cirrus of ~10% emissivity (very low ECA) that had been misclassified. Another investigation, in which an error analysis of ECA measurements was performed in the context of a different instrument, a similar conclusion was reached regarding the general ECA validation (Wielicki and Coakley 1981).

Sections 3.4.1 and 3.4.2 indicate that the CO_2-slicing technique, as used by NCTP, is a good indicator for inferring cloud-top heights for clouds above 600 hPa with ECA values from about 10% to 100%. The ECA measurement, even with known difficulties, provides a good stratification mechanism for variations in the cloud regime (i.e., opaque, thick, or thin cloud types).

Table 1. Summary of observation-derived cloud top measures.

Observation Type	Spatial and Temporal Resolution	Measured Quantity	Strengths	Weaknesses
Rawinsonde-derived cloud top (RCTP)	Available at points; every 12 hours; worldwide	Cloud top pressure (hPa)	Cloud-top height derived from direct measurement of meteorological state variables; measures clouds at all altitudes, but not for all temperatures;	Coarse temporal resolution, typically available every 12 hours; spatially limited; uncertainties in areas of deep convection
Radar-derived cloud top (ET)	Available at limited locations; 4-km/6-min resolution;	Echo top (ft)	Provides a lower bound check for cloud-top height accuracy	Always underestimates cloud tops; optimum range is 45-120-km from radar location due to beam broadening effects
Satellite –derived cloud top (NCTP) **(CO2-slicing technique)**	Available over oceanic domains (see Fig. 1); 10-14 km effective spatial resolution ; available hourly	Cloud top pressure (hPa); Effective Cloud Amount (%)	Detects clouds with a wide range of emissivity above 600 hPa with errors of ~50 hPa	Course resolution both spatially and temporally
OW CTOP diagnostic **(IR-window technique)**	Available over oceanic domains (see Fig. 1); 4-km/20-min. resolution	Cloud-top height (ft)	Detects cloud tops that are optically thick over oceans; high resolution both spatially and temporally.	Difficulty with detecting clouds that are less than optically thick.

4. Methods

The methods used to match the observation-derived cloud-top heights to the CTOP are presented in the following subsection. In addition, the statistical measures used to evaluate the diagnostic are summarized in Section 4.2.

4.1 Matching Methods

4.1.1 Intercomparison Matching Mechanics

For the intercomparison, the CTOP and cloud-top height values from the rawinsonde-, radar-, and the NCTP-derived cloud-top heights were matched at rawinsonde locations. The initial matching process involved selecting the value of each of the inferred cloud-top measures (CTOP, NCTP, and ET) at the gridpoint located spatially closest to the RCTP location. Secondly, the cloud-top height values (again for CTOP, NCTP,and ET) lying within a radius of 6, 12, and 24 km surrounding the RCTP location were considered. In particular, the median and peak values of CTOP, NCTP, and ET within each radius were determined, where the “peak” is the maximum derived cloud-top height value located within the radius. In addition, the “best” matched value within each radius (i.e., the derived cloud-top height closest in value to the RCTP) was also measured. With increasing radii, a larger number of derived cloud-top height values from each instrument were considered in each of these calculations. Table 2 shows the numbers of cloud-top height values included in the 6-, 12-, and 24-km radii for each observation instrument.

Table 2. The average number of derived cloud top values obtained from each observation instrument contained in a radius of each size.

Radius	ET	RCTP	CTOP	NCTP
6 km	7	1	7	1
12 km	28	1	28	2
24 km	112	1	112	9

With increasing radii within each range, some of the statistics have certain expected behaviors. For example, the “best” value should improve (or at least not get worse) with increasing radii, and the median would be expected to be relatively insensitive to the size of the radius.

4.1.2 Grid-to-Grid Comparison Mechanics (NCTP vs. CTOP)

The mechanics of the grid-to-grid comparison between the NCTP and CTOP were designed to account for differences in both spatial resolution and measurement time between associated values in CTOP and NCTP, issues that have plagued at least some other cloud-top height validation studies (Wylie and Menzel 1989). All CTOP pixels are marked with the valid time of the product while the scanline time, included in the NCTP product for each pixel, marks the valid time stamp for those pixels. A standard atmosphere calculation is used to convert the NCTP pressure values to height in units of feet. The effective resolution in mid-latitudes of a NCTP pixel is 10-14 km, which varies as a function of latitude due the field of view of the sounding instrument. This resolution roughly corresponds to a 3x3 set of CTOP pixels; the spatial window for a “one-to-one” comparison.

In an attempt to account for cloud movement in the time window by factoring in a mean zonal flow of 30 km/hr (Hansen and Sutera 1987), a 9x9 CTOP spatial window is used to provide a “time corrected” comparison. Two measures are estimated using the pixels within the spatial window; the median and the best match. The median value, a robust measure of the distribution, may not provide a good comparison in regions where the cloud field is discontinuous within the spatial window. For example, if one-third of the pixels contain the intended cloud height while two-thirds contain a cloud at a different height, then the median choice will be penalized by the verification measures. The “best” choice accounts for uncertainties in the cloud field and offers a comparison measure that is too liberal in many circumstances. Together these approaches provide bounds for the grid-to-grid comparisons.

The overall procedure is as follows:

1. For a given CTOP valid time and domain, select all NCTP pixels within the time window and the appropriate CTOP domain.

2. For each of the NCTP pixels, select the 3x3 and 9x9 CTOP spatial windows centered on the NCTP pixel.

3. Create four NCTP/CTOP comparison pairs by connecting the NCTP value with the CTOP 3x3 median, 3x3 best match, 9x9 median, and 9x9 best match.

In the analysis some of the pairs with the following properties are excluded:

- One or both of the values is set to clear (not cloud present)
- The CTOP value is below 15,000 ft (the algorithm minimum)
- The NCTP value is set to 150 hPa (the algorithm maximum), which is about 45,000 ft.

4.2 Statistical Measures

This evaluation of CTOP provides statistics based on a measures-oriented approach for evaluation of forecasts/diagnoses of continuous variables and a distributions-oriented approach (Murphy, 1993). In general, given the nature of the data available for these comparisons, it is appropriate to look at descriptive, rather than inferential, statistics. The types of approaches and statistics used are described further in this section.

For the intercomparison, robust measures such as the median and interquartile range[1] (IQR), lower quartile (i.e., 0.25th quantile) and upper quartile (i.e., 0.75th quantile) are considered in most cases. In addition, the mean and standard deviation are computed for some comparisons. These statistics are generally applied to the differences in values between two measures of cloud-top height (e.g., RTCP and CTOP), for subsets of the data (i.e., based on the stratifications presented in Section 4.3). The number of pairs of values also is an important parameter that can indicate the validity of the comparisons (e.g., a mean value based on two cases is not informative). The summary statistics (e.g., mean, median, IQR) provide information about the characteristics of the distributions of the differences. Actual (rather than

[1] The interquartile range is the range of values between the 0.25th and 0.75th quantile values, Half (50%) of the values of a variable are located between these two quantiles. The IQR is a robust measure of variability; larger IQR values indicate greater variability in the distribution.

absolute) differences are used for this analysis to provide indications of skewness and bias. Note that because the median is the midpoint of a distribution, a distribution is symmetric if the mean and median are equal.

A number of different types of exploratory plots are utilized for the intercomparisons, including histograms, scatter plots, height diagrams, and maps. The histograms display individual distributions of values, whereas scatterplots provide a way to examine the joint distribution of two cloud-top variables. Specifically, values of one product are plotted against the corresponding values for another product. If the two products provide equivalent values of cloud-top height, all of the points would fall on the 1:1 line. A scatterplot provides a detailed look at the relationship between two products, and can provide hints about ranges of values where the products are similar or different. Biases can also be evaluated using these plots; for example, if one product measures consistently higher or lower than the other, the distribution of differences is biased away from zero. Height diagrams show the summary statistics as a function of altitude (one of the stratification variables described in Section 4.3); these diagrams illustrate how the distributions of values change with height. Spatial distributions of some of the overall summary measures (e.g., median, IQR) provide information about regional variations in the comparisons.

For the grid-to-grid comparisons, the measures-oriented statistics provide values for comparison with validation studies associated with evaluations of the NESDIS cloud-top product and the CO_2-slicing approach, as summarized in sections 3.4.1 and 3.4.2. Results, stratified as described in section 4.3, include bias scores as well as mean absolute differences (MAD), where

Bias = avg CTOP value – avg NCTP value

MAD = avg absolute difference between the CTOP and the NCTP pixel height values

4.3 Stratifications

The statistics for the intercomparison and grid-to-grid comparisons are stratified using the following criteria:

Cloud Height as determined by CTOP

The height bins, each covering 5,000 ft, extend from the ground to 70,000 ft. This stratification allows comparison of the products at all levels as well as levels (middle and upper) for which the algorithm development suggests there should be good agreement and where the product is intented to be used.

Effective Cloud Amount as determined by NCTP

As explained in Section 3.4, the ECA provides a measure of cloud opacity. Statistics, for the CO_2-slicing derived values only (i.e., ECA < 100%), are stratified into the following ECA categories.

Intercomparison

- All clouds ECA between 0 and 100% (excluding the exact values of 0 and 100%)
- Thick ECA between 90% and 99% inclusive

Grid to Grid

- Opaque ECA between 95% and 99% inclusive
- Thick ECA between 51% and 94% inclusive
- Thin ECA between 1% and 50% inclusive

These values were chosen based on general ranges that have been suggested in the remote sensing literature. The ECA range chosen for the intercomparison of thick clouds is larger than for the grid-to-grid comparison and was adjusted to allow for a meaningful statistical sample size. The opaque range defined for the grid-to-grid comparison provides a rough indicator of convective clouds, particularly for levels in the atmosphere above 600 hPa. Remote sensing experts have studied this loose correlation using imagery case studies and ground-based lidar measurements (Personal communication with A. Schreiner). In addition, some numerical weather prediction models, including the operational Rapid Update Cycle (RUC), accept a threshold near 95% as an indicator of the presence of convection. In the case of the RUC convective parameterization, initialization with regions of ECA values great than or equal to 96% is designed to improve convective activity in the early stages of the model forecast (Personal communication with J. Brown, S. Weygandt, and R. Aune). Along with the comparison of the grids for all ECA values, the use of statistics in the opaque and thick ranges allows comparison in a regime where theory predicts agreement.

OW PDT Regions

Statistics are presented for both the intercomparison and the the grid-to-grid comparison in each of the domains defined by the CTOP product (GOMEX, N. Pacific, and Pacific; the CONUS domain was used for this comparison study only and will not provided as an experimental product to end users). As noted in Section 2, the GOES Sounder domains provide only partial coverage of the OW PDT regions.

Spatial Window

As described in Sections 4.1.1 and 4.1.2, statistics for the intercomparison are for the 6-km spatial radius and CTOP choice of median, and the grid-to-grid comparison are presented for the two spatial windows, 3x3 and 9x9, as well as the CTOP choice of median or best match. Additional results from the intercomparison are shown in Appendix A.

5. Results

The results of the cloud-top height intercomparison and the grid-to-grid comparison of NCTP and CTOP are summarized in this Section.

5.1 Intercomparison of Derived Cloud-Top heights at Rawinsonde Locations

The results from the intercomparison of cloud-top heights produced by the CTOP, RCTP, ET, and NCTP at rawinsonde locations are presented. The overall results are initially summarized. Secondly, the results are stratified by region, spatial variation, and by ECA. These comparisons focus on the median height and utilize the 6-km radius. All comparisons involving NCTP for this Section exclude ECAs of exactly 100%. Figures for other comparisons are shown in Appendix A. For a more detailed look at the intercomparison results, a case study involving Hurricane Frances is shown in Appendix B.

5.1.1 Overall Results from the Intercomparison

Since the CTOP product is aimed at diagnosing cloud-top heights at upper flight levels, the figures presented in this Section show results for CTOP over 15,000 ft. Because of the height limitations of NCTP, clouds above 45,000 ft are excluded. These overall results include all regions, time periods, and all cloud types.

Histograms of the differences between the cloud-top heights measured by the CTOP and the RCTP are shown in Figs. 6-8. Histograms of differences between values matched for two products are useful for discerning many important properties of the various quantities and demonstrate the mean and spread of the differences. The histograms are pictorial representations of the frequency distributions of the data. The abscissa represents the values of the differences between the products being compared and the abscissa are "binned" into 5,000 ft classes. The ordinate axis represents the frequency of values for each bin. Thus, the area of each rectangle is proportional to the frequency (or number of values) for a given interval (or bin).

As shown in Fig. 6, for most cases the RCTP reports a lower height than the CTOP for clouds between 15,000 and 45,000 ft. Because the RCTP generally tends to underestimate cloud-top heights, the results show that the CTOP values are nearly consistent with the RCTP values with errors that are typically ± 5,000 ft. However, in a number of cases, the CTOP is quite a bit higher than the RCTP, as shown by the blue bars. In fact, the shape of the histogram indicates that the distribution is somewhat skewed toward positive values and higher CTOP heights.

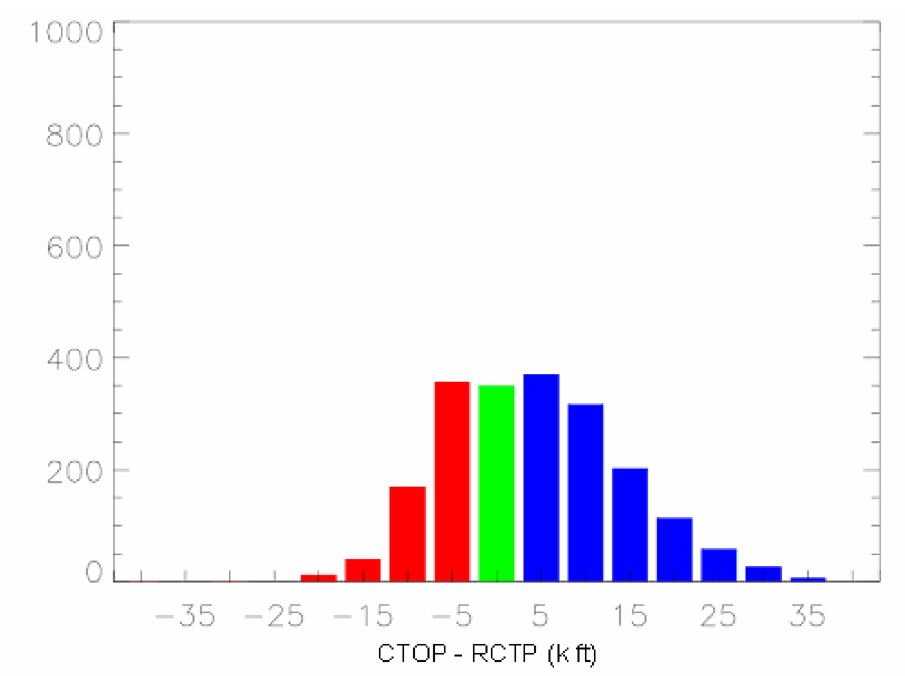


Figure 6. Histograms showing differences between CTOP and RCTP from 15,000 ft to 45,000 ft in 5,000 ft bins for the median CTOP height within a 6-km radius around the RCTP location. Red bars denote RCTP heights exceeding CTOP heights by more than 2,5 00 ft. Blue bars denote CTOP exceeding RCTP by more than 2,500 ft. The green bar indicates differences between -2,500 to 2,500 ft.

Figure 7 shows differences between CTOP and ET heights. In most cases, the cloud-top heights diagnosed by the CTOP are typically higher by nearly 10,000-15,000 ft. In a very small number of cases, the ET reports heights exceeding those observed by the CTOP. These slightly contrary cases are possibly due to the fact that at some sites, the ETs are not based on optimal elevation angle; in addition, topography may sometimes affect the radar observations (Brown et al. 2000).). This is consistent with the well-known relationship between radar echo tops and cloud-top heights.

A histogram showing the comparisons between the NCTP and CTOP is shown in Fig. 8. The results indicate that the NCTP usually reports higher heights than the CTOP when compared at station locations. However, the magnitudes of these differences are most commonly less than 5,000 ft.

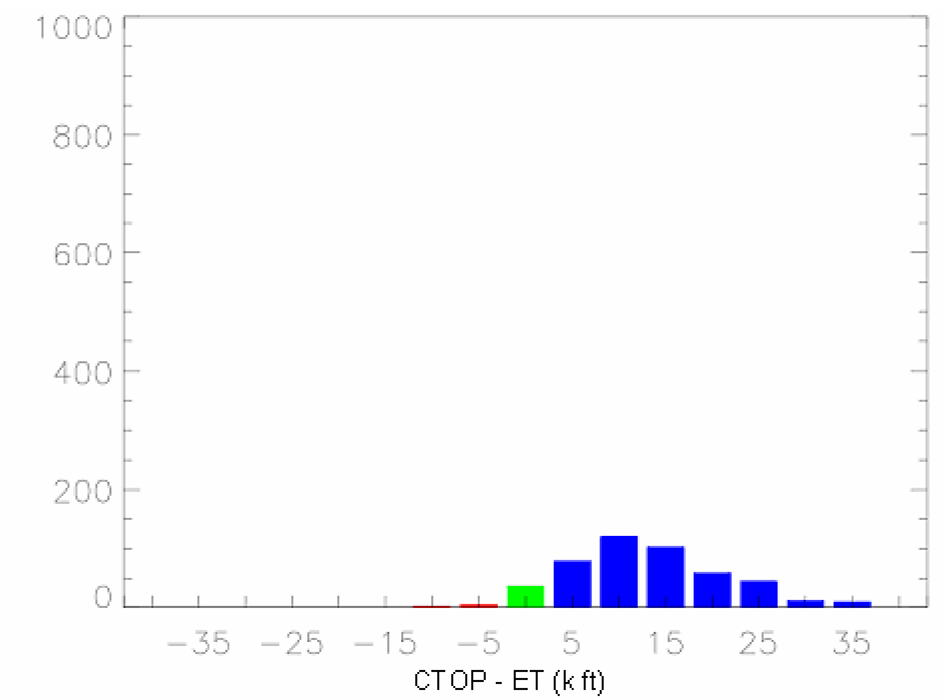


Figure 7. Same as Fig. 6, except for CTOP and ET.

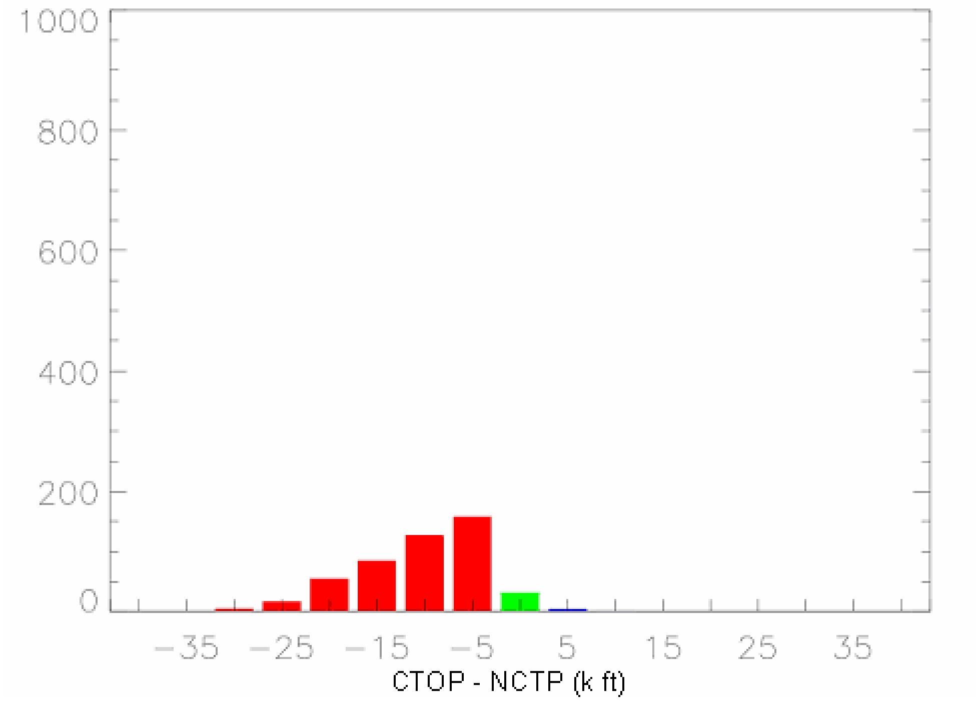


Figure 8. Same as Fig. 6, except for CTOP and NCTP (ECA = 100% excluded).

Figures 9-11 show differences between CTOP and RCTP as a function of CTOP height. At a given height between 15,000 to 45,000 ft, the number of CTOP/RCTP pairs, the median and mean difference between the derived cloud-top heights, and the 25^{th} and 75^{th} percentiles are plotted. The number of pairs at each height gives an idea of whether there are sufficient data to draw any conclusions from statistics (e.g., the mean is not very informative when based on only two values).

Figure 9 shows that at levels below 25,000 ft and for all clouds types, the differences between the cloud-top heights produced by CTOP and RCTP are typically less than 5,000 ft. However, above 25,000 ft, these differences steadily increase. At CTOP heights of less than 40,000 ft, the differences between CTOP and RCTP remain less than 12,000 ft. However, at heights above 40,000 ft, the differences reach nearly 20,000 ft. However, the variability as represented by the width of the IQR (i.e., the difference between the 25^{th} and 75^{th} quartiles), is relatively constant with height.

Figure 10 shows the comparison between the CTOP and ET derived cloud-top heights. Similarly to Fig. 8, differences in the cloud-top height values computed between the CTOP and ET increase from a difference of nearly 0 at 15,000 ft to a difference of nearly 20,000 ft at a height of 45,000 ft (Fig. 10). This result may be partially due to the decrease in sample size from 200 at 15,000 ft to nearly 0 at 45,000 ft. As expected, the CTOP is higher than the ET throughout all height ranges.

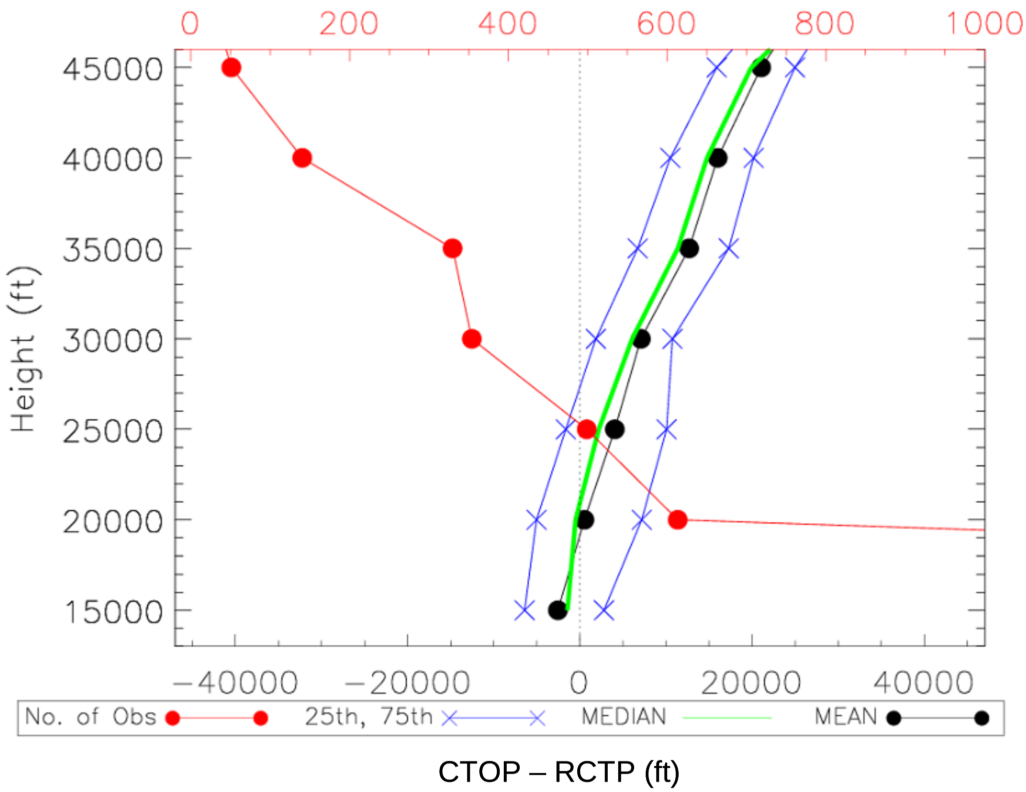


Figure 9. Differences between CTOP and RCTP heights stratified by CTOP height are shown in 5,000 ft intervals. Number of observations (red; circle, see upper scale), mean difference (black; circles to the right of zero), median difference (green; solid line), and 25th and 75th percentiles (blue;'x') are shown. The 15,000 ft interval includes all below it.

Contrary to the relationship shown between the RCTP and ET, differences between the CTOP and NCTP for all cloud types decrease with height (Fig. 11). NCTP usually reports higher heights than the CTOP particularly below 30,000 ft. The differences between CTOP and NCTP heights are largest below 20,000 ft where the NCTP may detect optically thin clouds at upper levels, but those clouds may be misclassified or undetected by CTOP.

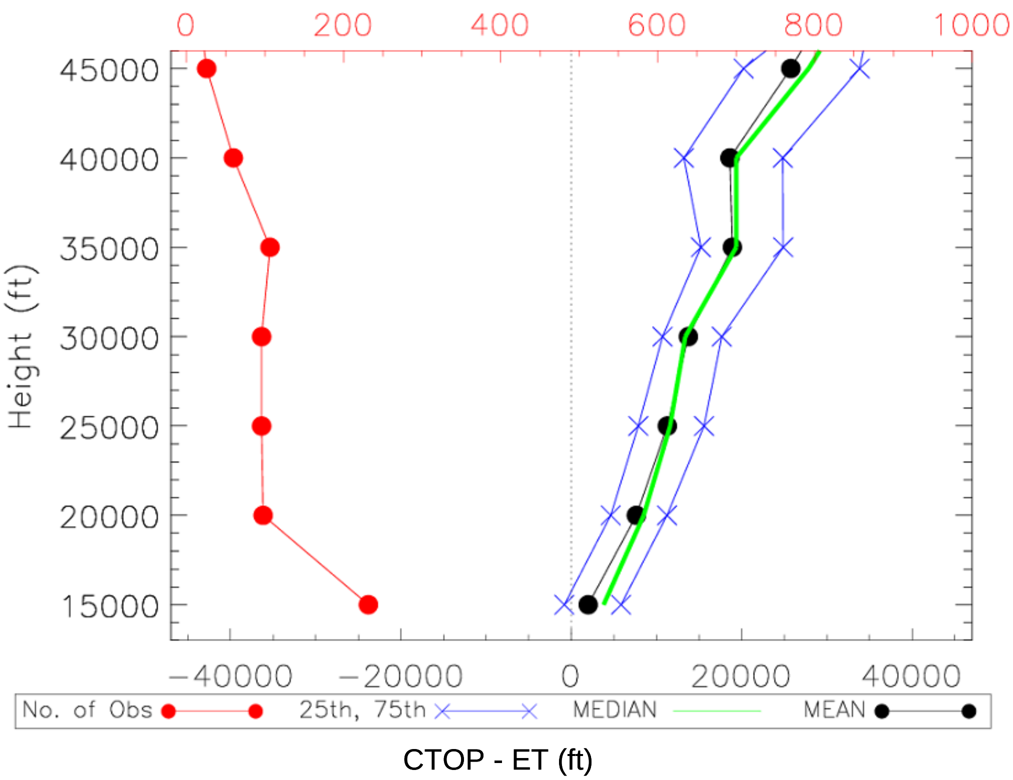


Figure 10. Same as Fig. 9, except for CTOP and ET.

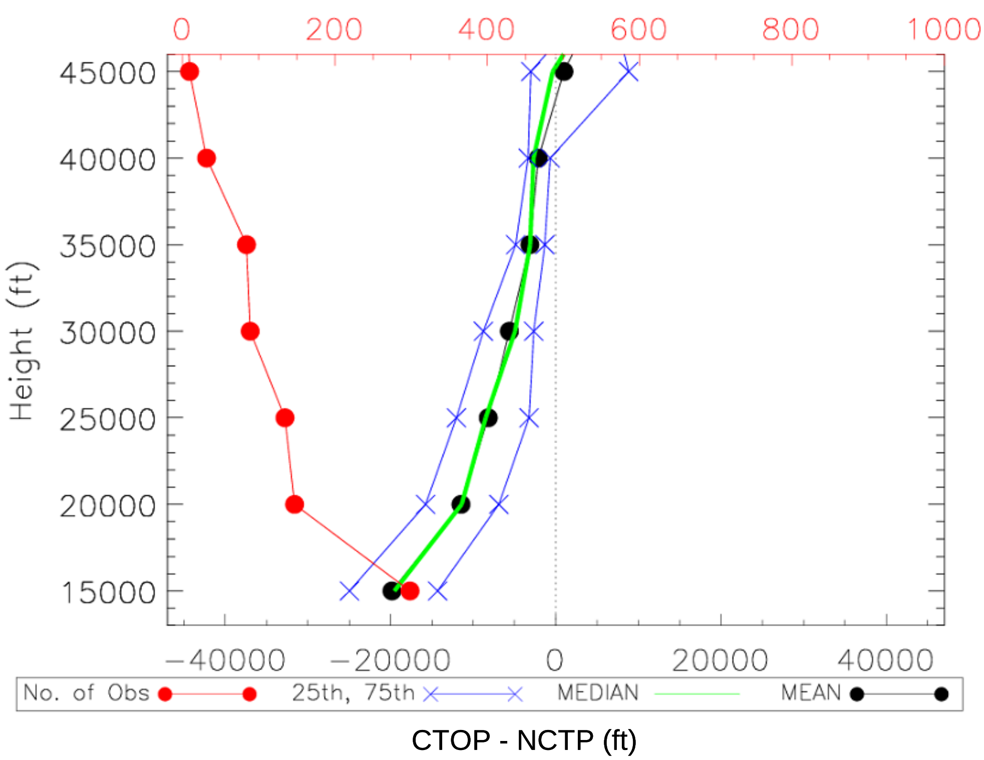


Figure 11. Same as Fig. 9, except for CTOP and NCTP (ECA = 100% excluded)

A basic comparison of cloud existence for all cloud types between CTOP and NCTP and ET was also completed. If a cloud was detected by the RCTP, then the existence or absence of cloud produced by the CTOP, NCTP, and ET was recorded. The results for CTOP vs. NCTP and CTOP vs.ET are presented in Tables 3 and 4, respectively.

Table 3. Contingency table showing, where CTOP and NCTP detect (**≠ 0)** or miss (= 0) a cloud, if a cloud was detected by the RCTP. These results are based on the single closest point to the rawinsonde location.

	NCTP = 0	**NCTP ≠ 0**
CTOP = 0	2,081 (37.2%)	225 (4.0%)
CTOP ≠ 0	1,254 (22.4%)	2,037 (36.4%)

Table 4. Same as Table 3, except for ET and CTOP.

	ET = 0	**ET ≠ 0**
CTOP = 0	1,735 (30.0%)	265 (4.6%)
CTOP ≠ 0	2,782 (48.1%)	1,003 (17.3%)

Based on Table 3, when RCTP detects a cloud, the CTOP was likely to diagnose a cloud 59% of the time, while the NCTP diagnosed a cloud 40% of the time. Also, 36% of the time the CTOP product detected a cloud when both the NCTP and RCTP also detected clouds. However, 37% of the time no clouds were diagnosed by the NCTP or CTOP when clouds were detected by the RCTP product.

The cloud no-cloud results for ET and CTOP are summarized in Table 4. The results in Table 4 indicate that when RCTP detects a cloud, the CTOP detects almost twice as many clouds as the ET. However, the CTOP misses clouds detected by ET 21% of the time.

Table 5 shows the differences between CTOP, Navy soundings (referred to as NRCTP in this Section), NCTP, and ET where Navy ship soundings exist. The NRCTP offers the only *in situ* observations over oceanic areas in our datasets and, although limited in coverage, provides a meaningful comparison for the CTOP for all types of clouds. In Table 5, because of the limited number of observations, the original NCTP values were used, including ECAs of 100% and CTOP reports of cloud-top heights less than 15,000 ft. In addition, the results for several radii and cloud-top estimates (peak, median, and best) are presented. As shown in Table 5, the NRCTP is consistently higher than the CTOP. Differences between NRCTP and CTOP are usually less than 2,000 ft in the February – April period, and increase in the August-September period.

The NRCTP is lower than NCTP during the Northern Hemisphere spring and higher than NCTP during the Northern Hemisphere summer. The NRCTP and NCTP generally show more variability in their differences than the NRCTP and the CTOP. The differences between NRCTP and both the NCTP and CTOP are both larger and more variable during August– September period than the February–April period.

Table 5. NRCTP v. NCTP and CTOP statistics over Navy soundings locations in the Pacific (values in ft). NCTP1 and CTOP1 refer to the cloud-top values spatially closest to the RCTP location. The NCTP2, NCTP3, NCTP4 and CTOP2, CTOP3, CTOP4 refer to values for the 6, 12, and 24 km radii, respectively (separated into best, mean, and peak categories)

NRCTP v.	**12 February – 23 April**				**15 August – 15 September**			
	N	Mean	Std. Dev.	Med	N	Mean	Std. Dev.	Med
NCTP1	43	-5814	10394	-786	60	2998	13968	3888
CTOP1	167	1486	7797	2631	147	7876	9652	6867
NCTP2 peak	31	-5623	10448	-448	48	3265	13862	3842
NCTP2 med	31	-5550	10376	-488	48	3443	13950	3842
NCTP2 best	31	-5550	10376	-488	48	3265	13862	3842
CTOP2 peak	167	634	7985	1994	147	6898	9885	6362
CTOP2 med	167	1783	7656	2740	147	8155	9656	7320
CTOP2 best	167	1507	7140	1994	147	7282	9277	6362
NCTP3 peak	40	-7582	10475	-8542	59	389	14696	3016
NCTP3 med	40	-3775	11443	-105	59	3016	13444	4025
NCTP3 best	40	-4711	9019	-345	59	1821	12563	3182
CTOP3 peak	167	-875	8104	573	147	5837	10284	5719
CTOP3 med	167	1516	7633	2353	147	8052	9650	7320
CTOP3 best	167	1142	6301	830	147	6871	8828	5719
NCTP4 peak	43	-13371	12764	-15896	60	-3962	16497	364
NCTP4 med	43	-4210	11287	0	60	4013	12794	4048
NCTP4 best	43	-3983	6893	-15896	60	3592	7392	2172
CTOP4 peak	167	-3244	8247	-1970	147	3631	11160	3368
CTOP4 med	167	1598	7659	2641	147	8175	9369	7201
CTOP4 best	167	436	5267	104	147	5880	7967	3368

5.1.2 Regional Results

This Section presents comparisons of CTOP and the other measures of cloud-top height for the CTOP regions. Because the initial investigation included the rawinsonde stations over coastal areas, the results for the coastal areas were weighted heavily in the overall statistics; more complete results for that region are shown in Appendix A. Results over coastal areas for the CTOP comparisons against RCTP, ET, and NCTP are consistent with the overall results, as are results over CONUS and the Pacific.

Comparisons of the cloud-top measures over each CTOP domain are shown in scatterplots in Figs. 12 and 13. As described in Section 4.2, scatterplots illustrate the joint distribution of two variables, and are useful displays for descriptive comparisons.

The RCTP and CTOP are compared in Fig. 12. Each diagram shows a general cluster of points, especially at heights less than 20,000 ft, around the 1:1 line. For the Pacific, the February-April period shows more scatter than the August-September period. Noting that the Pacific region contains significantly more data in the Southern Hemisphere, this result suggests that this region displays more variability during its convective season. The plots also indicate that there is little systematic difference between the RCTP and CTOP

When comparing the CTOP and ET (see Appendix A), the CTOP is consistently higher than the ET, showing more consistent differences over the Gulf of Mexico and less over CONUS.

In comparing the NCTP and CTOP (Fig. 13), in cases where the NCTP and CTOP detected clouds, the NCTP usually reported a higher height. The two products were most closely matched over the CONUS. There is less of a relationship between CTOP and NCTP over the Pacific. The 12 February–23 April period also shows less of a relationship between NCTP and CTOP than the 15 August–15 September period. The differences also appear more systematic, with very few instances where CTOP heights are lower than NCTP heights.

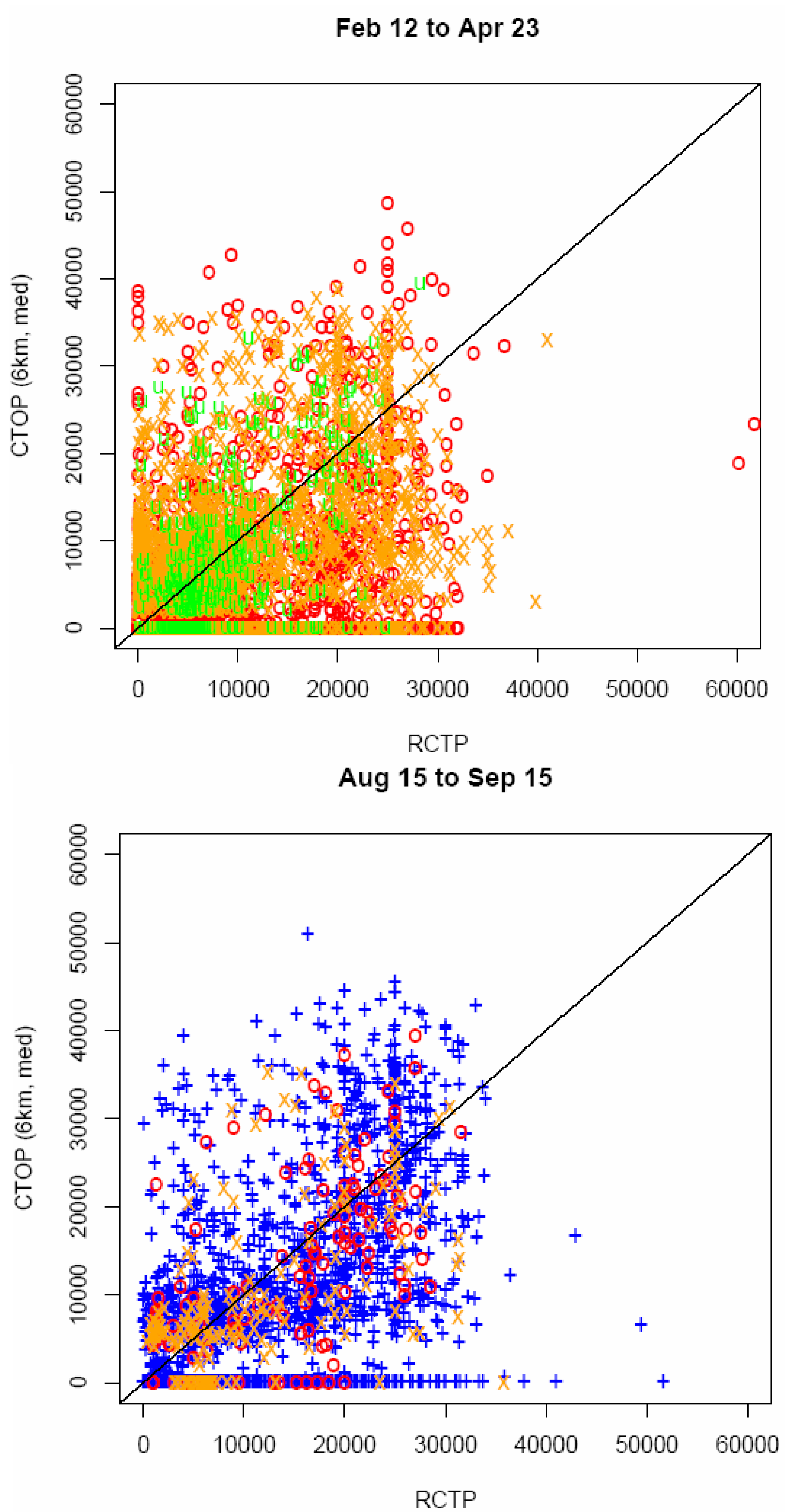


Figure 12. Scatterplot comparison of CTOP and RCTP for 12 February–23 April (top) and 15 August – 15 September (bottom) is shown. Different regions are shown using different symbols and colors; Pacific, 'o'; CONUS, '+'; Gulf of Mexico, 'x'; and North Pacific, 'u'.

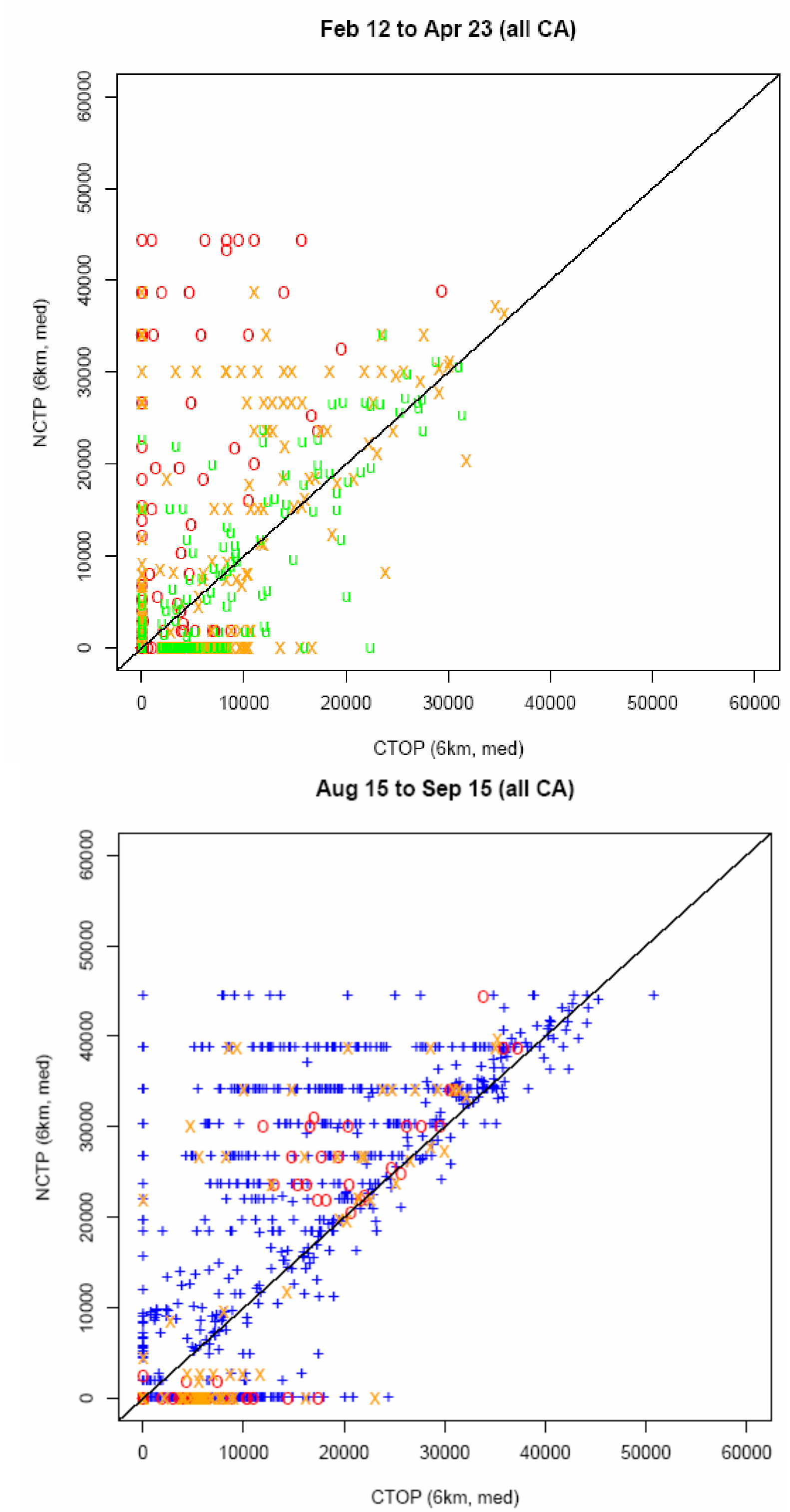


Figure 13. As in Fig. 12, for CTOP and NCTP. ECAs of 100% are included. Points with ECA values of less than 100% appear as scattered points between binned values of NCTP (since the NCTP values based on the CO_2-slicing method are binned; those based on the IR window technique are not). Symbols are as in Fig. 12.

5.1.3 Spatial Variations

Figure 14 shows a map of the medians of differences between CTOP and RCTP, and NCTP, and ET for CTOP heights greater than 15,000 ft. Variability is shown through the IQR in the second row of maps. In comparing RCTP and CTOP, CTOP is usually higher than RCTP, especially over the Gulf of Mexico, which suggests that RCTP underestimates cloud-top heights in areas of deep convection. Additionally, in areas with high low-level humidity, such as the Gulf of Mexico, a cloud may also be detected by RCTP, though none may be found (Wang and Rossow 1995). The relatively large IQR values in the Gulf of Mexico region indicate relatively large variability and may result from these factors.

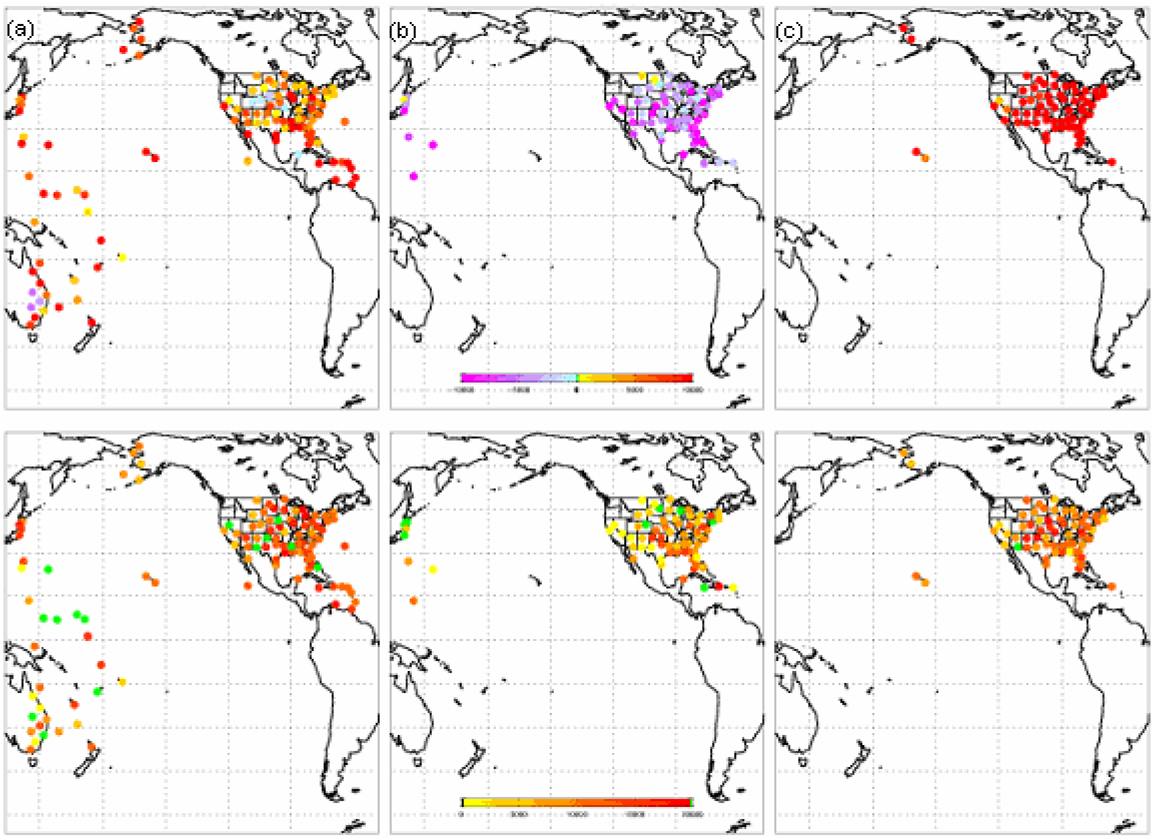


Figure 14. Maps showing median differences for comparisons with CTOP over 15,000 for (a) CTOP vs. RCTP, (b) CTOP vs. NCTP, and (c) CTOP vs. ET (top row). Green denotes areas where the median differences are near zero. Maps showing IQR for same differences (bottom row). Green in bottom row denotes areas where the IQR exceeds 20,000 ft.

This map shows that over most areas the RCTP heights are lower than the CTOP especially over the Gulf of Mexico and similar low-latitude regions. Also, particularly over CONUS, the NCTP is usually higher than the CTOP. This difference is likely due to the detection of high, thin clouds by NCTP. The differences are more variable over the Great Plains, Gulf of Mexico and surrounding areas, but less so over the southwestern U.S. The differences between CTOP and ET show that for nearly all stations, the

CTOP height exceeds that of the ET. However, some arid mountainous regions or regions with climatologically lower relative humidity show smaller differences between the two.

5.1.4 Stratification by NCTP Effective Cloud Amount

The results in this Section include only NCTP pixels with ECA values between 90 and 99% which are derived from the CO_2-slicing technique. The results for all regions and seasons combined for the NCTP and CTOP comparison are shown in Fig.15. Most frequently, the CTOP and NCTP are within 5,000 ft of each other and likely because the high, thin clouds diagnosed by NCTP and misclassified or undiagnosed by CTOP were removed from the dataset.

Examination of the differences between CTOP and RCTP when ECAs are 90-99%, as shown in Fig. 16, indicates there are fewer instances where the RCTP exceeds the CTOP. More commonly, the CTOP exceeds the RCTP by 5,000 ft. When focusing on ECAs of 90-99%, the differences between CTOP and ET (Fig. 17) change little from the case when all ECAs (excluding 100%) are included (Fig. 7). In this case, the sample size is reduced.

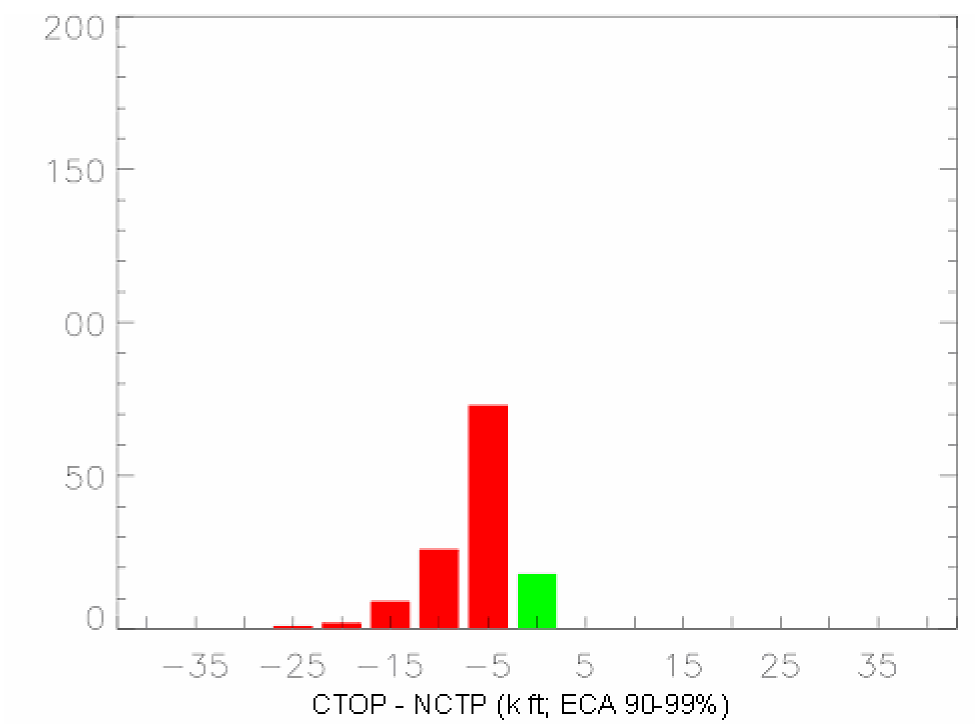


Figure 15. Histogram showing differences between CTOP and NCTP above 15,000 ft in 5,000 ft bins for the median CTOP height within a 6-km radius of the RCTP with ECA values between 90 and 99%.

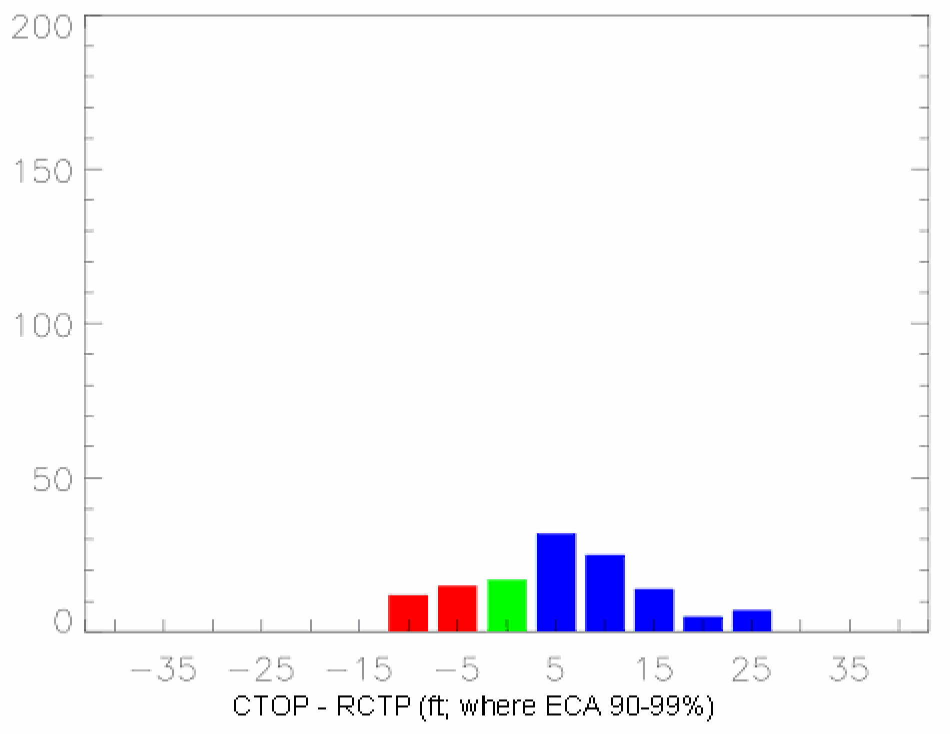


Figure 16. As in Fig. 15, except for CTOP and RCTP.

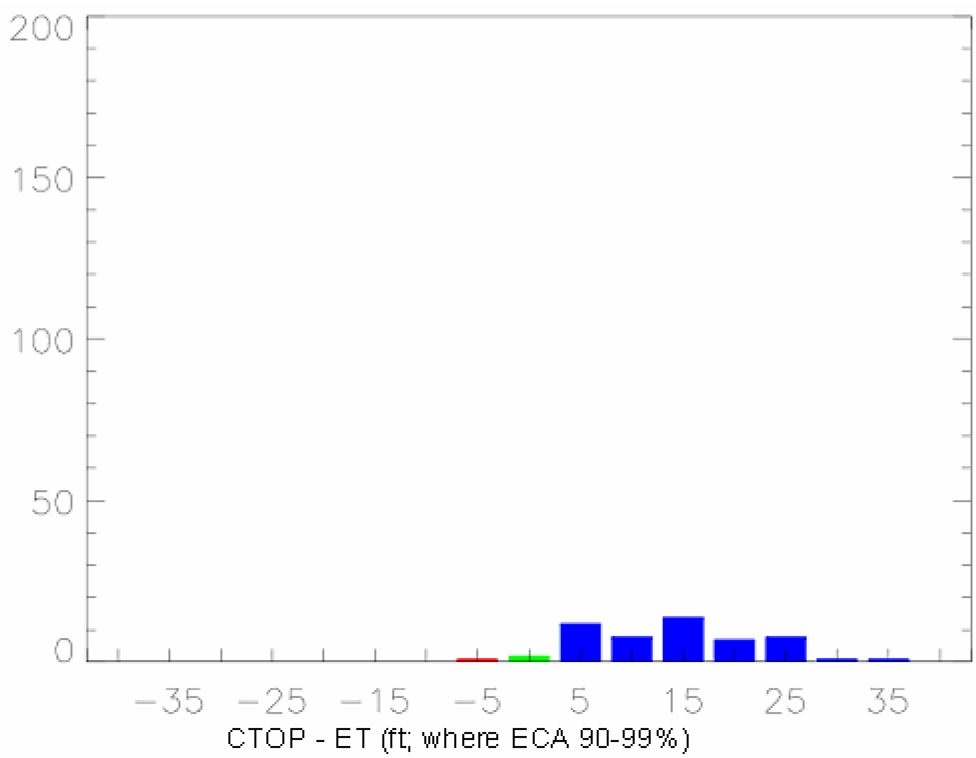


Figure 17. As in Fig. 15, except for CTOP and ET.

Figure 18 shows differences between the CTOP and the NCTP for each 5,000-ft level between 15,000 and 45,000 ft. When ECAs less than 90% and equal to 100% are excluded, NCTP and CTOP show mean and median differences of less than 5,000 ft at levels above 25,000 ft. Differences at the lower levels (below 25,000 ft) decrease from those presented for all cloud types (Fig. 11). However, the sample size also decreases possibly affecting the results.

By limiting the analysis to ECAs between 90-99% (Fig. 19), the differences between RCTP and CTOP become more exaggerated below 15,000 ft than compared to those presented for all cloud types (Fig. 9). The differences between 20,000 ft and 30,000 ft decreased, while those above 30,000 ft change little. At low levels, the RCTP heights are higher than those diagnosed by CTOP, while at mid-levels, they become more similar.

The results in Fig. 19 indicate that when the ECA values are restricted to 90-99%, the differences between ET and CTOP change little with height than compared to those presented for all cloud types (Fig. 10), although the sample size decreased greatly. (Note that below 15,000 ft, less than five observations meeting these criteria were included).

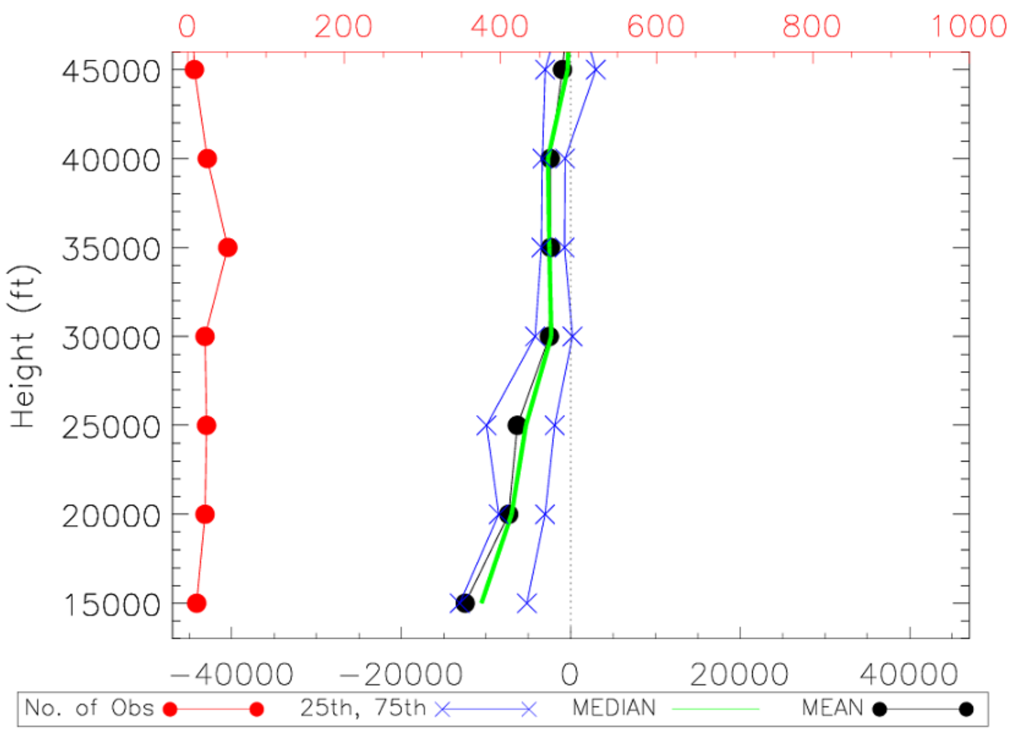
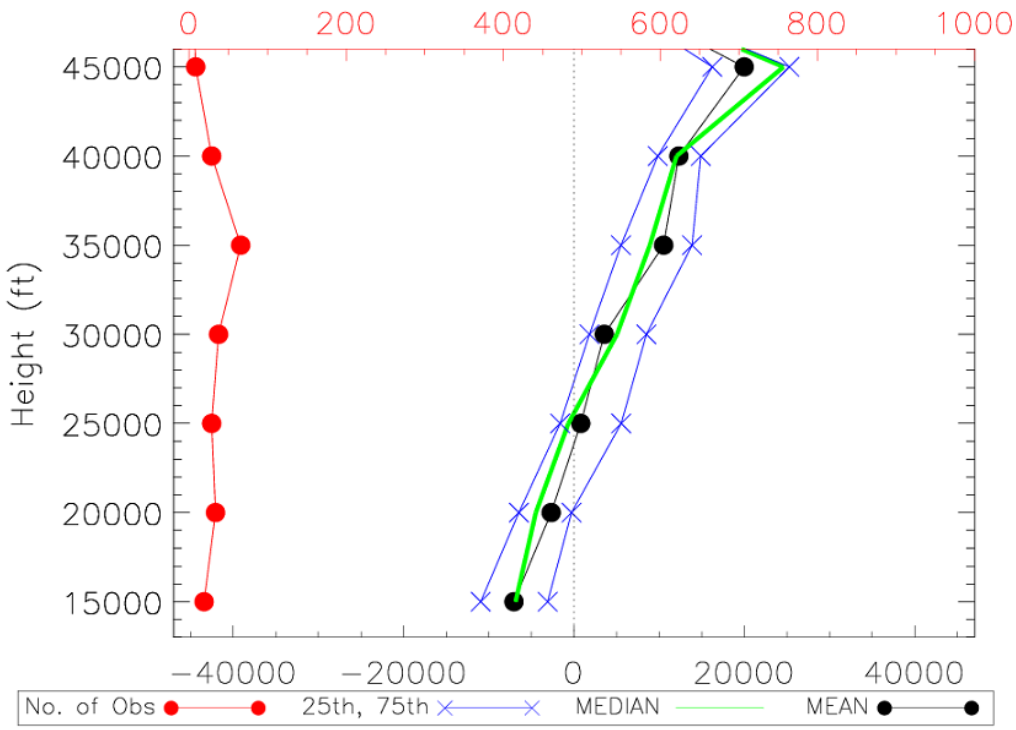
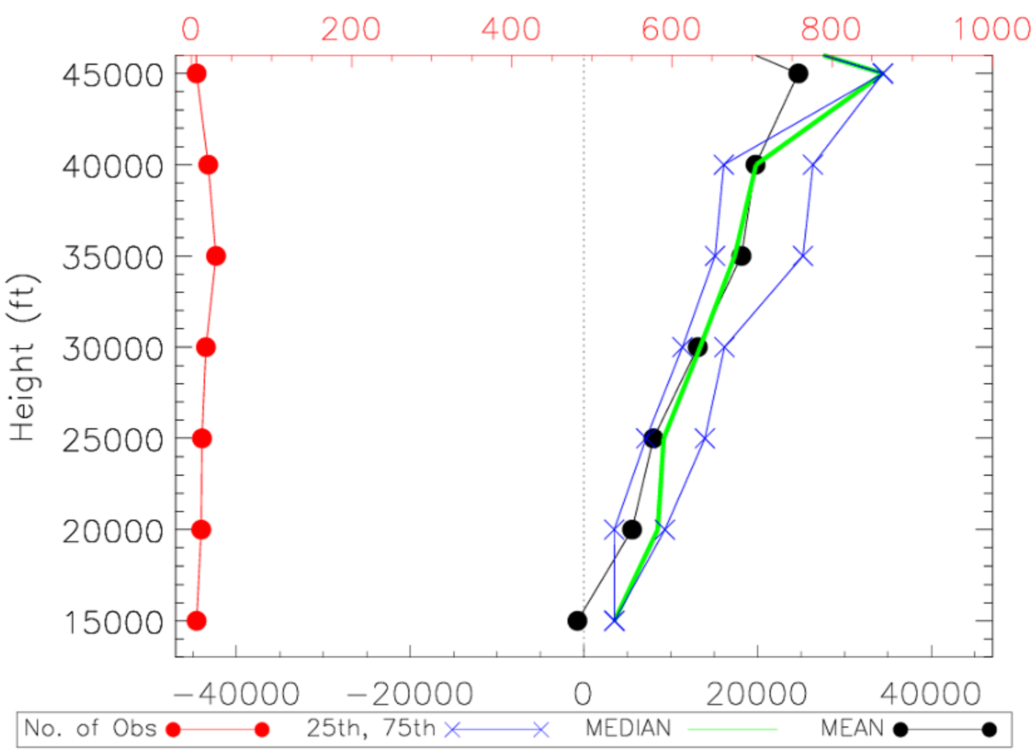


Figure 18. Differences between CTOP and NCTP heights, where the NCTP ECA is between 90-99%, stratified by CTOP height are shown in 5,000 ft intervals, valid below their plotted height. Number of matched pairs (red, see upper scale), mean difference (black), median difference (green), and 25th and 75th percentiles (blue) are shown. Each interval includes observations within 5,000 ft below it.



CTOP - RCTP (ft; where ECA 90-99%)

Figure 19. As in Fig. 18, except for CTOP and RCTP.



CTOP - ET (ft; where ECA 90-99%)

Figure 20. As in Fig. 18, except for CTOP and ET.

Figure 21 shows that the medians of the differences between CTOP and NCTP are smaller over the Gulf of Mexico as well as over CONUS when the gridpoints are restricted to those with ECA between 90 and 99%. These differences are smaller than the differences computed for all cloud types. In addition, the IQR is smaller, than for all cloud types, indicating the differences are less variable. The median differences between NCTP and CTOP are reduced when ECAs of 90-99% are examined. The IQR is also smaller for RCTP and ET when examining values associated with the 90-99% range of ECA than when all cloud types are included, but the sign of the differences does not change. Overall, when we restrict the analysis to areas where the NCTP ECA is between 90 and 99% (most likely opaque clouds), the differences between the cloud-top height values produced by RCTP and CTOP change little, as do the differences between ET and CTOP, possibly suggesting that the ECA has little effect in these comparisons on the RCTP and ET. However, the cloud-top height differences between NCTP and CTOP are reduced when the comparison is restricted to opaque clouds.

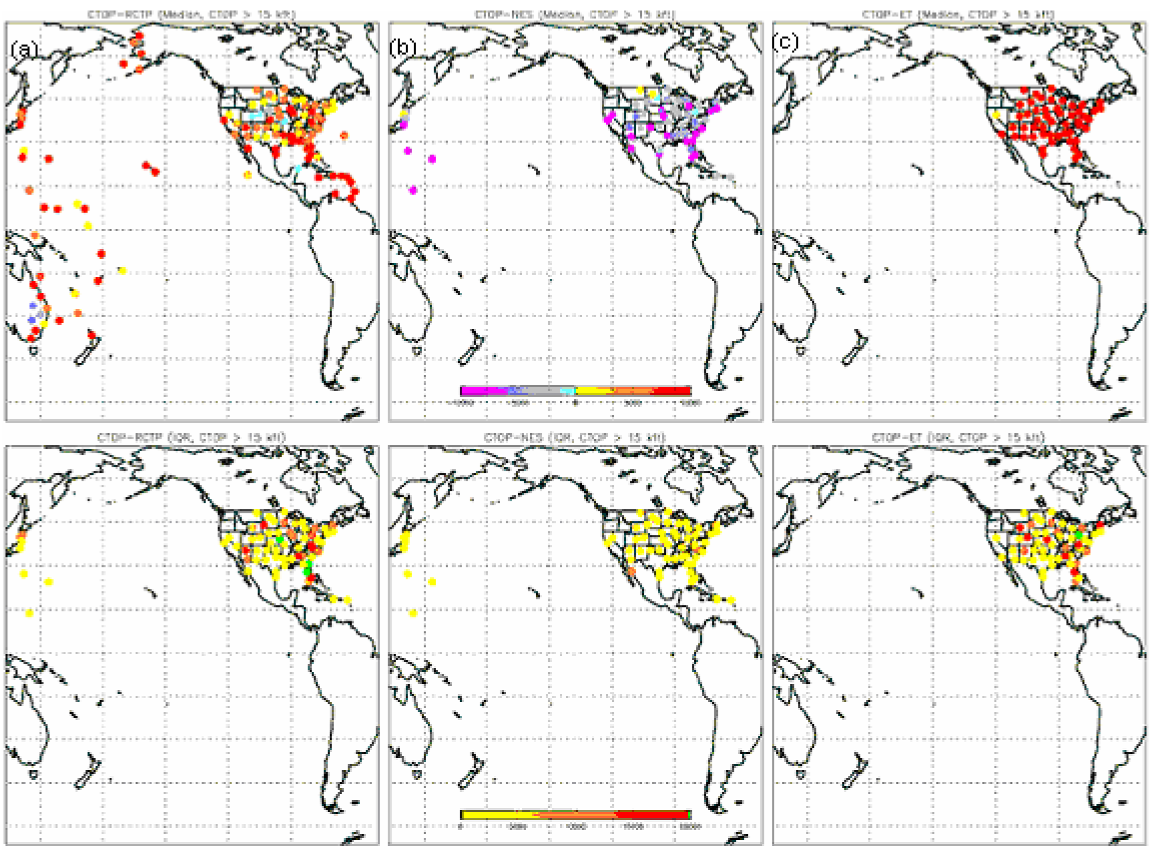


Figure 21. As in Fig.14, except for ECAs of 90-99%

5.1.5 Summary of Intercomparison Results

To summarize, the results from the intercomparisons, using the 6-km radius and median value, of CTOP with RCTP, ET, and NCTP, several relationships were identified:

Overall results for all cloud types and heights from 15,000 to 45,000 ft combined:

- CTOP typically had higher heights than RCTP with differences typically larger than 5,000 ft
- CTOP had heights that were consistently higher than ET, which was expected.
- CTOP heights were typically lower than the NCTP with differences ranging between 5,000 and 10,000 ft.

Overall results for all cloud types stratified by height:

- RCTP/CTOP showed good agreement below 25,000 ft. Differences above 25,000 ft increased to nearly 20,000ft.
- ET/CTOP differences were less than 5,000 ft at a height of 15,000 ft, but increased to nearly 20,000 ft at a height of 45,000 ft.
- NCT/CTOP differences were typically 10,000 to 20,000 ft at heights below 30,000 ft. Above 30,000 ft, the differences decrease to less than 5,000 ft, which are within the instrument error.

Overall results for ECA values between 90-99% stratified by height:

- NCTP/CTOP differences decrease when compared to all cloud amounts
- RCTP/CTOP and ET/CTOP differences were similar to those when all cloud types were included.

Other results to consider:

- CTOP and RCTP are more likely to agree than disagree on the existence of clouds

5.2 Grid-to-Grid Comparison of CTOP and NCTP

The evaluation is extended here to data-poor oceanic regions through comparisons of CTOP and NCTP using the grid-to-grid matching approach described in section 4.1.2. Results of the comparisons are presented and then interpreted in the context of previous cloud-top height validation studies.

5.2.1 NCTP CO_2-slicing Comparison

All results presented in this Section include only NCTP values inferred by the CO_2-slicing technique. The CTOP and NCTP values are independent in the sense that they are derived from different algorithms, the IR Window versus CO_2-slicing, utilizing data from different instruments, the GOES Imager versus the Sounder. The first pair of height plots presented in this Section (Fig. 22) compares biases, defined as the average CTOP value minus the NCTP value, for the two different CTOP spatial window values, the best match and the median. As described in Section 4.1.2, the best value probably underestimates the bias, while the median overestimates it. Together they bound a reasonable estimate of the measure. The results are stratified by ECA with each curve on the plots representing a different ECA range, as defined in Section 4.3.

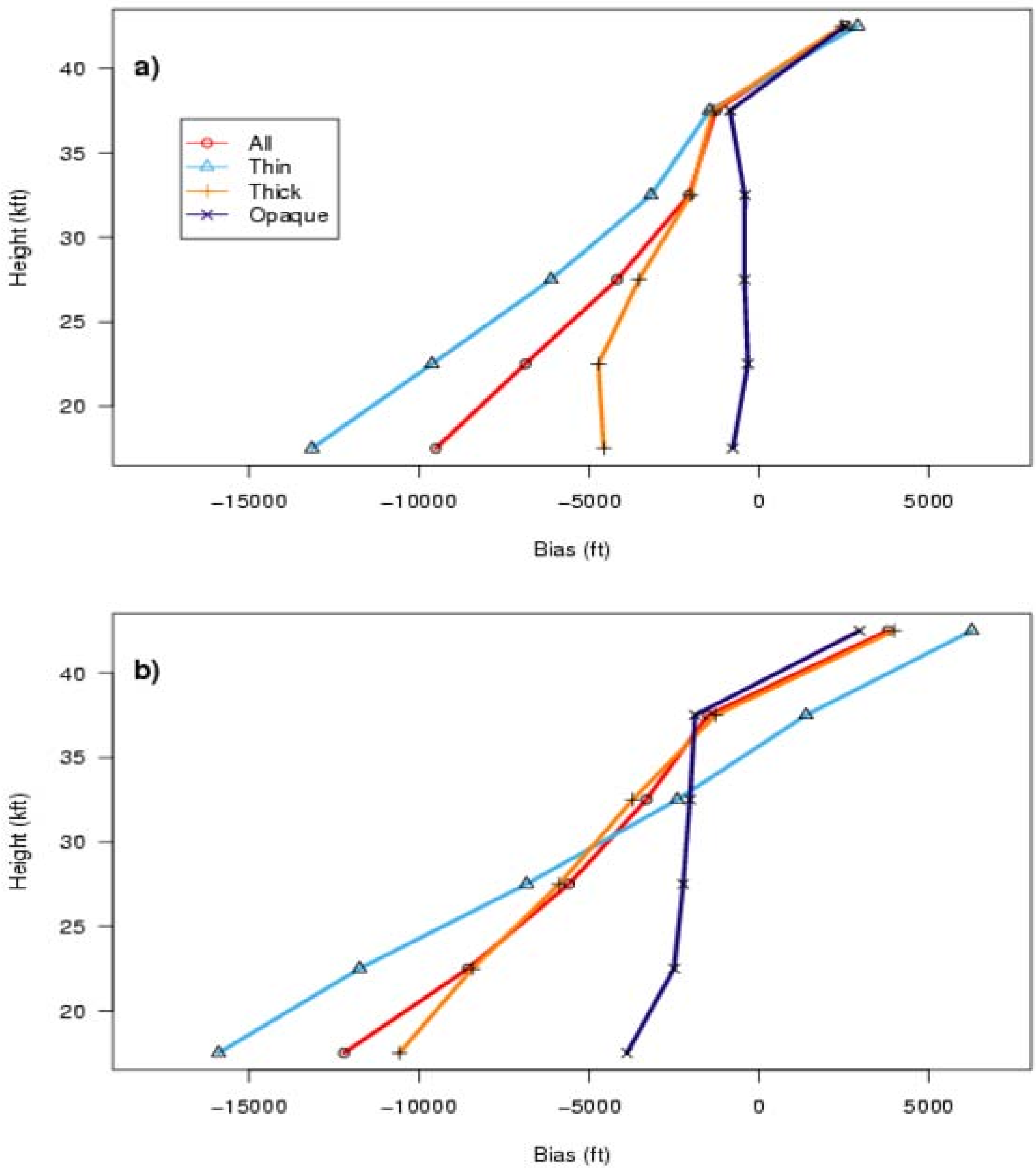


Figure 22. Height series of bias values stratified by effective cloud amount. CTOP comparison values determined (a) by 9x9 best (b) 9x9 median. Data points are plotted in the center of each 5,000 ft vertical bin.

Comparison of the curves in Fig. 22 (a) and (b) reveals that, in general, the best match has a bias lower in absolute value than the bias for the median. The plots show that for opaque clouds, the products agree well throughout the height levels; disagreement at the lower levels increases with decreasing ECA, a result predicted by theory. The following example illustrates the most likely cause for the divergence of the bias curves at lower altitudes. The NCTP CO_2-slicing technique would most likely correctly classify a high cloud with a moderate ECA. The CTOP IR Window technique, however, would "see through" the cloud, measure it as too warm, and misplace the cloud height too low in the atmosphere. The large difference contributing to the bias would then be associated with low clouds, as measured by CTOP.

The abrupt increase in differences at the highest level is mainly due to variation in the highest allowable cloud level of the NCTP product. For cases where the NCTP processing determines that the height of the tropopause is lower in the atmosphere than 150 hPa (~ 45,000 ft), the algorithm chooses the height of the tropopause, rather than 150 hPa as the maximum height value. No comparison is done for NCTP pixels set to the algorithm's maximum value of 150 hPa. Therefore, these differences do not contribute to the overall statistics in the uppermost height bin. The analysis approach does however include comparisons for which the NCTP pixel is set to the height of the tropopause. The NCTP dataset does not contain any additional information which may be used to determine whether the tropopause height at a given location is utilized. In situations where the algorithm chooses the tropopause height as its maximum value, boundary differences can contribute significantly to the overall statistics. For example, if CTOP determines a cloud height of 45,000 ft and NCTP agrees by determining the maximum cloud height equal to the tropopause height at 41,000 ft, then a difference of 4,000 ft is added to the statistics in the upper bin. Additionally, most of the NCTP values from 40,000-45,000 ft are determined by the IR Window technique and are not included in this analysis. This yields a much smaller sample size in the uppermost bin that may lead to greater variability in the values being compared.

Integrating the height curves provides the following overall values. For cloud heights of all ECA values, the best match bias is –5,470 ft while the median bias is –8,480 ft. With only opaque clouds included in the measure, the best match bias is –550 ft and the median bias is –2,270 ft. These bias values, when compared to the results of studies used to validate the NESDIS cloud-top product and the CO_2-slicing method, show very good agreement between the products for opaque clouds and marginal agreement for non-opaque clouds. It is notable that for cloud tops above 30,000 ft, as measured by CTOP, the products qualitatively agree for all ECA values.

The next pair of height series plots presented in this Section (Fig. 23(a) and (b)) compares MAD, for the two different CTOP spatial windows with values for the best match and median.

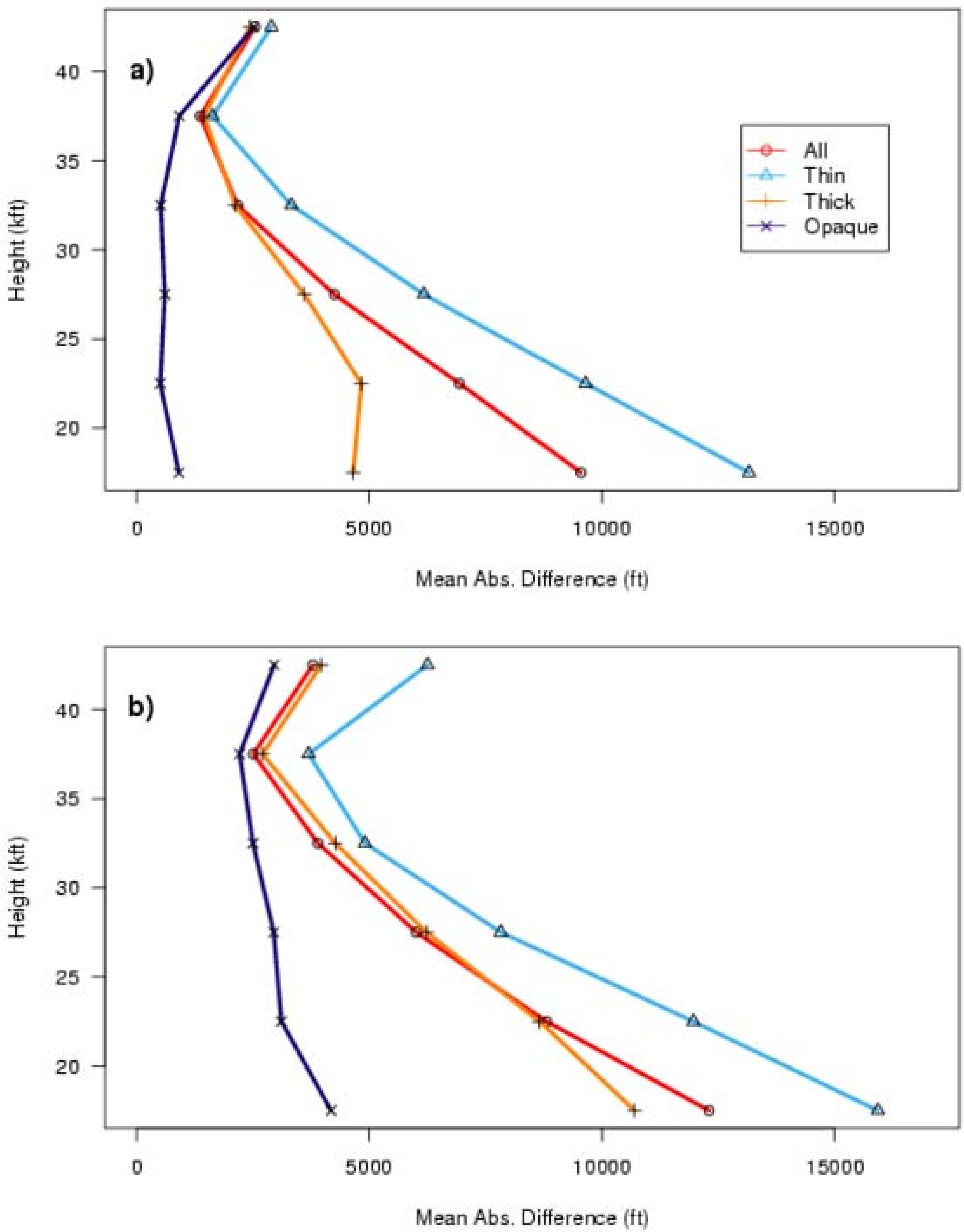


Figure 23. Height series of MAD values stratified by effective cloud amount. CTOP comparison values determined (a) by 9x9 best (b) 9x9 median.

Comparison of the curves in Fig. 23 (a) and (b) reveals that the best values have lower MADs than the medians, as would be expected. These plots of MAD exhibit characteristics similar to those in the bias plots of Fig. 22. According to the MAD, the products agree for opaque clouds throughout the height levels. Disagreement, as described and explained for the bias plots, is evident for the thin cloud categories.

Integrating the values over the 15,000-45,000 ft layer provides the following overall values. For cloud heights of all ECA values, the overall MAD for the best match is 5,550 ft while the MAD for the median is 8,790 ft. With only opaque clouds included in the measure, the overall MAD for the best match is 670 ft and the overall MAD for the median is 2,810 ft. As with the bias measures, these MAD values show very good agreement with the results of validation studies for the NESDIS and CO_2-slicing cloud-top measurement approaches.

Figure 24 compares the bias and MAD for two different ECA ranges 91% to 99% and 95% to 99%. Figure 24 is included to illustrate the compatibility of the “opaque” ranges as defined for the point analysis and for the grid analysis.

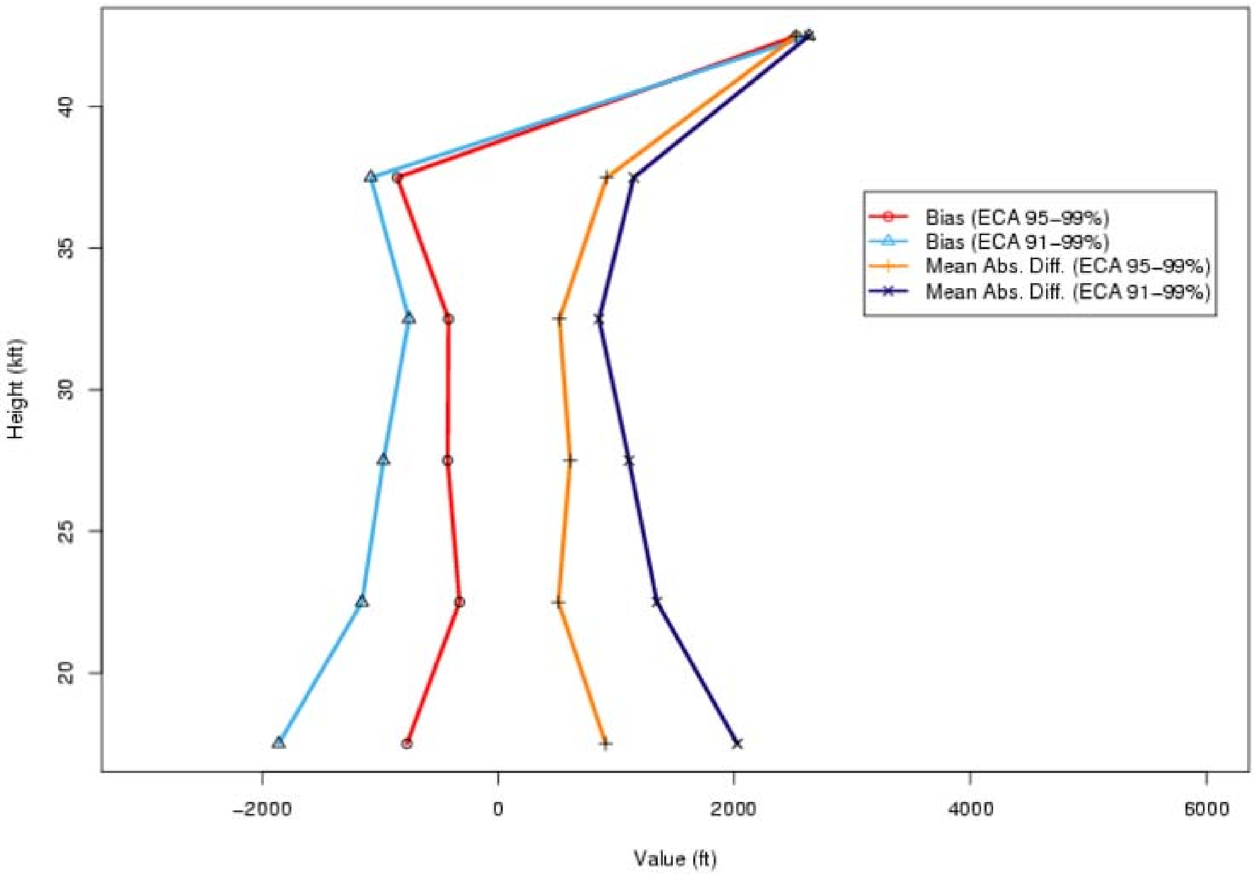


Figure 24 – Height series of bias and MAD values stratified by two different opaque ECA ranges 91% to 99% and 95% to 99%. The CTOP comparison values are determined by 9x9 best.

The curves demonstrate that the bias and MAD values differ by about 500 ft between the two ECA ranges with the smaller ECA range showing slightly better agreement between the products. This justifies use of the larger ECA range stratification in the point analysis, chosen to retain a significant sample size, as a proxy for comparison to results in the smaller ECA range of the grid-to-grid opaque stratification.

The next pair of height series plots presented in this section (Fig. 25) compares the bias stratified by CTOP region and separated for opaque and non-opaque clouds, which is the aggregate of thick and thin clouds as defined in Section 4.3.

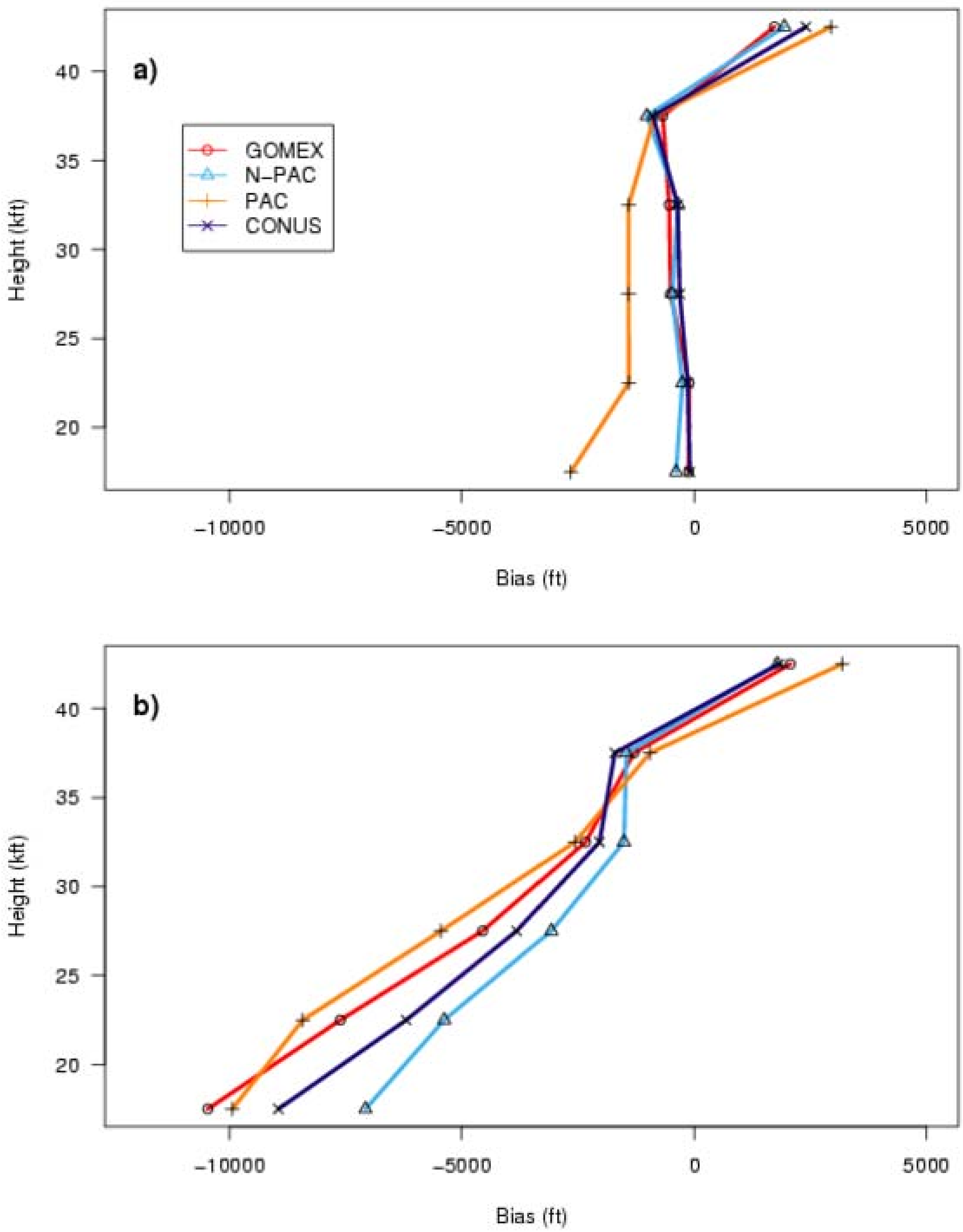


Figure 25. Height series of bias values stratified by CTOP region. Results are presented for (a) opaque (ECA >= 95%) and (b) non-opaque (ECA < 95%).

Figure 25 (a) demonstrates very good agreement between CTOP and NCTP for opaque cloud-top height values in all of the CTOP domains. The bias for the Pacific domain is notably different than the measures for the other domains. There are many possible explanations for the difference of the Pacific domain values. One of the NCTP product developers suggested that the algorithm quality control, geared to the domains in and around CONUS, might not be performing as well for the GOES-9 Sounder data. He also noted that the Sounder instrument on GOES-9, older than those on the other GOES satellites, appears to be noisier (Personal communication, A. Schreiner). Climatological differences between the Pacific domain and the others may also be contributing to the variation.

The bias difference for the non-opaque clouds, previously described for Figure 22, is evident in Figure 25 (b) for all of the CTOP regions. The two tropical regions, the Pacific and GOMEX domains, appear to have slightly higher overall biases. This result may be due to climatological differences. This plot also demonstrates marginal agreement between the products for upper level non-opaque clouds in all CTOP domains.

Comprehensive statistics for the CO_2-slicing comparison are presented in Tables 5 and 6. Values are included for all four of the comparison values as defined by the grid-to-grid mechanics. As described in Section 3.4.2, bias and MAD measures with absolute value less than ~3000 feet indicate very good agreement.

Table 5. Overall bias values (ft) for the various comparisons and ECA ranges.

Effective Cloud Amount	9x9 Best	9x9 Median	3x3 Best	3x3 Median
All	-5,470	-8,480	-7,110	-8,080
Opaque	-550	-2,270	-1,280	-2,020
Thick	-3,380	-7,800	-5,620	-7,410
Thin	-9,550	-14,300	-12,460	-14,080

Table 6. Mean absolute difference values (ft) for the various comparisons and ECA ranges.

Effective Cloud Amount	9x9 Best	9x9 Median	3x3 Best	3x3 Median
All	5,550	8,790	7,300	8,470
Opaque	670	2,810	1,640	2,680
Thick	3,480	8,120	5,840	7,800
Thin	9,605	14,480	12,550	14,320

The negative overall bias associated with all comparison mechanics and all ECA stratifications indicates that, in general, cloud-top heights from CTOP are lower in the atmosphere than those from NCTP. Both tables demonstrate that CTOP and NCTP strongly agree for opaque clouds when any of the four comparison values is used. There is also good agreement for thick clouds for the 9x9 best match comparison approach. The products strongly disagree with differences ranging from over 9,000 ft to nearly 15,000 ft, according to all comparison mechanics, for clouds in the thin ECA stratification.

5.2.2 NCTP IR Window Comparison

All results in this Section include only NCTP values inferred by the IR Window technique (i.e. NCTP pixels with an assumed ECA of 100%). The comparison values are not independent with respect to the algorithm used to identify the clouds, but they are independent with respect to the data source. This analysis provides a comparison in a regime where the products should strongly agree.

Figure 26 presents the bias and the MAD for the IR Window technique NCTP pixels only, with CTOP comparison values determined by a 9x9 spatial window. Statistics for both the median and best match mechanics are shown.

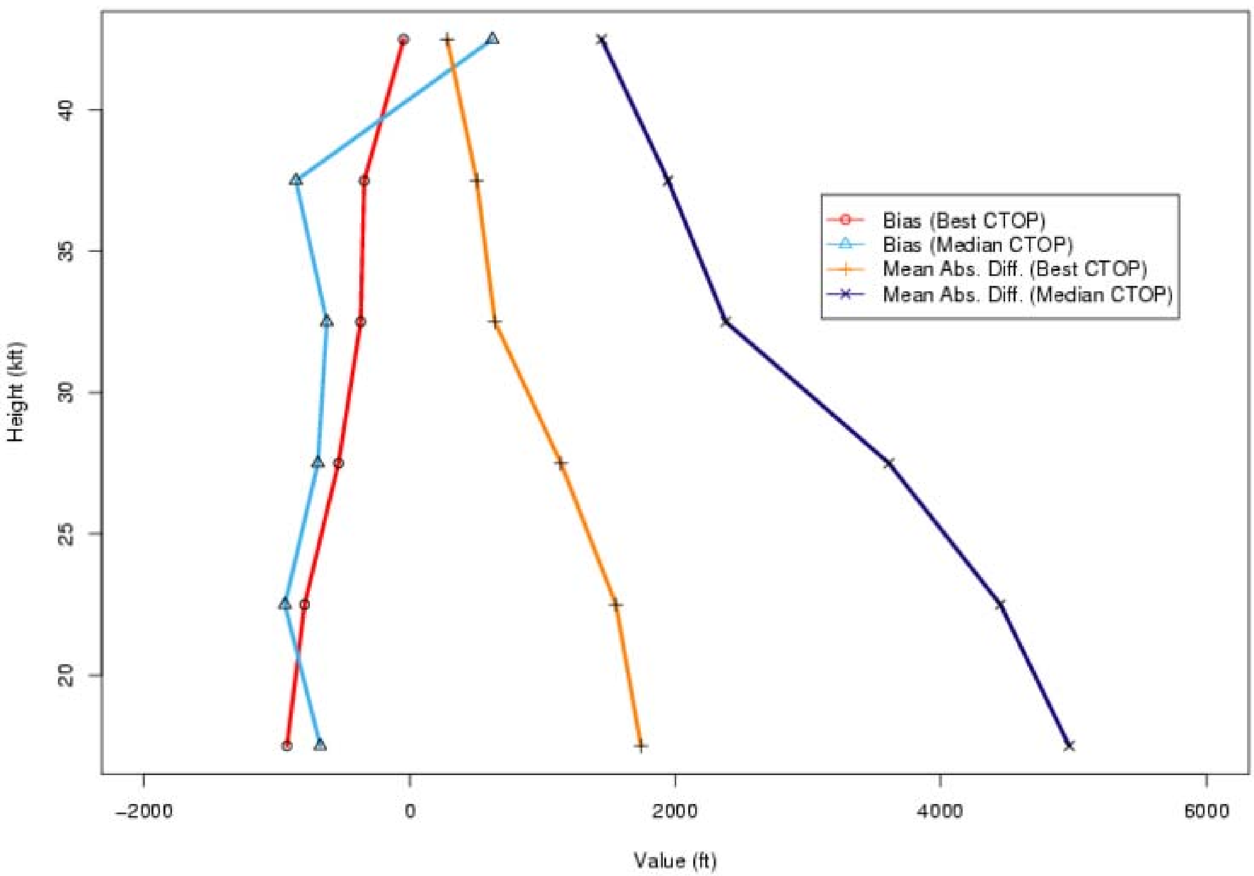


Figure 26. Height series of bias and MAD for IR Window technique NCTP pixels only, with CTOP comparison values determined by a 9x9 spatial window.

Figure 26 demonstrates very good agreement with a bias less than ~1000 feet when the two products infer cloud-top height values with the same technique. The MAD values for the 9x9 best match show very good agreement while the MAD values for the 9x9 median diverge slightly, although still showing marginal agreement at the lower levels. Differences in time and pixel resolution could account for the larger MAD values. Based on the comparisons shown in Table 5, the results for the 3x3 best and median values for the IR Window technique, not presented here, are expected to fall within the bounds presented in Fig. 26.

5.2.3 Summary of Grid-to-Grid Comparison and Interpretation

The results of the grid-to-grid comparison, bounded by the 9x9 best and 9x9 median values, demonstrate very good agreement (as determined by CTOP) with differences of ~ 3,000 ft between CTOP and NCTP for opaque and thick clouds, particularly at levels above 25,000 ft. These results are consistent with other cloud-top height validation studies as described in Sections 3.4.1 and 3.4.2 and with those revealed by the intercomparison. The statistics for the thin cloud comparison for all height levels below 40,000 ft and as determined by CTOP, show significant disagreement, an expected result given the theoretical strengths and weaknesses of the products.

6. Conclusions

This report has summarized results of a comparison of cloud-top heights produced by the OW PDT CTOP product and derived cloud-top heights from other measuring instruments to determine the relative quality of the CTOP product. The study used a point-to-point intercomparison approach to determine relative differences between four cloud-top height techniques. The intercomparison was then followed by an indepth grid-to-grid study of the relationship between the CTOP and the satellite-based cloud-top height product (NCTP). The investigation occurred for the periods 12 February-23 April and 15 August-15 September, 2004. The results of the study indicate that CTOP appears to have approximately the same capabilities as other methods for determining cloud-top heights over the oceans.

In particular, the intercomparison revealed that when all cloud types are considered from 15,000-45,000 ft the CTOP heights are typically higher than the radar-derived and echo top height products, but typically the differences were within the instrument error. When the CTOP was stratified by height, the differences between the 3 products increased with increasing height. However, when the CTOP was compared to the satellite-derived product (NCTP), the CTOP heights were typically lower than the satellite product. When the results were stratified by height and effective cloud amount, good agreement between the heights of the CTOP and NCTP products, as noted by the decrease in height differences, was evident.

Results from the grid-to-grid comparison indicated very good agreement between the CTOP and the satellite-derived heights for opaque and thick cloud from 25,000-40,000 ft (as measured by CTOP) with overall differences approximately 3,000 ft. The statistics for the thin cloud comparison showed significant disagreement for all heights below 40,000 ft, an expected result given the theoretical strengths and weaknesses of the two products. Therefore, the CTOP algorithm is particularly capable at estimating the heights of deep convective clouds that may be a hazard to aviation.

Acknowledgments

This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

We would like to thank the members of the Oceanic Weather Product Development Team for their support of the independent verification effort over the last several years and for providing guidance and direction with respect to the understanding of the remote sensing products. We also thank Jason Nachamkin, (NRL) for providing the Navy ship sounding data used in the analysis, Timothy J. Schmit, (NOAA/NESDIS/ORA, Madison, WI) and Anthony Schreiner (associate researcher at the Cooperative Institute for Meteorological Satellite Studies, at the University of Wisconsin-Madison Space Science and Engineering Center) for providing information regarding the creation and interpretation of the NESDIS Cloud Top Pressure product, R.M Steadham (NOAA Radar Operations Center) for providing information on the use of radars, and Robert M. Aune (NOAA/NESDIS) for providing information regarding the use of effective cloud amount (ECA) in the analysis. In addition, we would like to express our appreciation to other members of the NCAR and FSL verification teams who provided excellent support for this effort.

References

Brown, R.A., V.T. Wood, and D. Sirmans, 2000: Improved WSR-88D scanning strategies for convective storms. *Wea. Forec.*, **15**, 208-220.

Frey, R. A., b. A. Baum, W. P. Menzel, S. A. Ackerman, C. C. Moeller, and J. D. Spinhime 1999: A comparison of cloud-top heights computed from airborne LIDAR and MAS radiance data using CO_2 slicing. *J. Geophys. Res.*, **104**, 24,547-24,555.

Hansen, Anthony R., A. Sutera 1987: The probability density distribution of the speed and horizontal and vertical shear of the zonal-mean flow. *J. Atmos. Sci.*, **44,** No. 11, 1525-1533.

Hawkinson, James A., Wayne Feltz, Timothy J. Schmit, Anthony J. Schreiner, and Steven A. Acherman, 2001: A validation study of the GOES sounder cloud top pressure product. *Preprints of 11th Conference on Satellite Meteorology and Oceanography, Madison, WI, October 2001*. Amer. Meteor. Soc., Boston, MA, 348-350.

Landolt, S.D., R.E. Bateman, B.C. Bernstein, 2004: A comparison of satellite and sounding derived cloud top temperatures. *Preprints, 20th Conf. on WAF/16th Conf. on NWP*, 11-15 January, Seattle, WA, Amer. Meteor. Soc., Boston, MA.

Menzel, W.P., W. L. Smith, and T. R. Stewart, 1983: Improved cloud motion wind vector and altitude assignment using VAS. *J. Cli. and Appl. Meteor.*, **22,** 377-384.

Miller, S, J. Hawkins, T. Lee, 2002: Cloud-top heights for aviation near-real time application. Available from URL: http://www.rap.ucar.edu/projects/owpdt/documents/cloudtops2.html.

Murphy, A.H., What is a good forecast 1993: An essay on the nature of goodness in weather forecasting. *Wea. Forec.* **8,** 281-293.

Roy, B., J.B. Halverson, J. Wang, 2004: The influence of rawinsonde "age" on TRMM field campaign soundings humidity correction, *J. Atmos. And Oceanic Tech.*, **21**, 470-480.

Schreiner, A.J., Timothy J. Schmit, and W. Paul Menzel, 2001: Observations and trends of clouds based on GOES sounder data. *J. Geophys. Res*., **106**, 20,349-20, and 363.

Takacs, A., B. Brown, and J. Mahoney, 2004a: Verification of oceanic weather diagnoses and forecasts for aviation weather elements. *Preprints on the 84th AMS Annual Meeting.* Seattle, WA 11-15 January, Amer. Meteor. Soc., Boston, MA.

Takacs, A., B. Brown, R. Hueftle, L. Holland, S. Madine, J. Mahoney, and M. Kay, 2004b: Evaluation of the oceanic cloud-top height diagnostic product: Strategy of the verification methodology. *Preprints, 11th Conference of Aviation, Range, and Aerospace Meteorology*, 4-8 Oct 2004, Hyannis, Massachusetts, Amer. Meteor. Soc., Boston, MA.

Wang, J., and W.B Rossow, 1995: Determination of cloud vertical structure from upper-air observations. *J. Appl. Meteor.*, **34**, 2243-2258.

Wang, J., W.B. Rossow, and Y. Zhang, 2000: Cloud vertical structure and its variations from a 20-yr global rawinsonde dataset. *J. Climate*, **13**, 3041-3056.

Warren, S. G., C. J. Hahn, J. London, R. M. Chervin and R. L. Jenne, 1986: Global distribution of total cloud cover and cloud type amounts over land. NCAR Tech. Note NCAR/TN-273+STR, 228 pp.

Wielicki, B. A., and J. A. Coakley Jr., 1981: Cloud retrieval using infrared sounder data: Error analysis. *J. Appl. Meteor.*, **20**, 157-169.

Wylie, D.P. and W. P. Menzel, 1989: Two years of cloud cover statistics using VAS. *J. Climate*, **2**, 380-392.

Zhang, H., and W. P. Menzel, 2002: Improvement in thin cirrus retrievals using an emissivity-adjusted CO2-slicing algorithm. *J. Geophys. Res.*, **107**, 4235. .