P9.7

EVALUATION OF THE OCEANIC CLOUD-TOP HEIGHT DIAGNOSTIC PRODUCT: STRATEGY OF THE VERIFICATION METHODOLOGY

Agnes Takacs*, Barbara Brown, Robert Hueftle, Lacey Holland
Research Application Program, National Center for Atmospheric Research, Boulder, CO

Sean Madine, Jennifer Mahoney, Mike Kay
NOAA Research-Forecast Systems Laboratory, Boulder, CO

1. INTRODUCTION

This paper summarizes the strategy for an evaluation of the diagnoses produced by the Cloud-Top Height (CTOP) product being developed by the Oceanic Weather Product Development Team (OWPDT) of the Federal Aviation Administration's Aviation Weather Research Program (FAA/AWRP). The CTOP is based on a combination of infrared (IR) reflectance values from the geostationary (GOES) satellites with temperature and pressure profiles from the Global Forecast System (GFS) numerical weather prediction model. The CTOP value is converted from pressure to height using the U. S. Standard Atmosphere.

The greatest challenge associated with creating accurate and timely oceanic weather hazard information is the lack of data available for algorithm development and verification. To complicate the problem, the observation datasets used for verification should be independent of those used by the algorithm. In a data sparse environment, this independence is difficult to achieve. Therefore, it is necessary to develop creative approaches for extracting as much information as possible from the limited amount of global data available. The AWRP's Quality Assessment Product Development Team (QAPDT) has investigated approaches for evaluating the performance of the CTOP algorithm. The quality of the CTOP was evaluated from February through April 2004 by members of the QAPDT. An additional evaluation period has been chosen to increase coverage over land between August 1 and September 15, 2004. The evaluation approach is discussed in the next section. Techniques for measuring cloud-top heights are shown in Section 3, and the verification methodology is presented in Section 4.

**Corresponding author address:* Agnes Takacs, NCAR/RAP, Boulder, CO 80307-3000; email: agnes@ucar.edu

2. EVALUATION APPROACH

Cloud-top height is not observed directly, except by lidar observations which are usually only available for research purposes. Therefore, to obtain matching datasets for verification, cloud-top height values are being inferred using several different data sources that are not used by the algorithm to diagnose CTOP. Due to these observational uncertainties, it is not possible to verify the CTOP algorithm in the strict sense generally used for evaluating forecasts or diagnoses of variables that are more easily observable (e.g., surface temperature). Instead, the less ambitious focus of this evaluation is to ensure that the CTOP is consistent with other standard estimates of cloud-top height.

A primary dataset for this evaluation is the GOES sounder-based cloud-top pressure (CTP) product produced by the National Environmental Satellite, Data, and Information Service (NESDIS). Radiosonde and radar observations available over the CONUS, the coastal areas and islands are also used to create estimates of cloud-top height. These observations of cloud-top height are expected to provide at least the lower bounds on the expected CTOP values while pilot report observations of cloud tops, which will also be considered over the CONUS, are likely to provide more of an upper bound. For example, Schreiner et al. (2001) found that the NESDIS CTP altitude

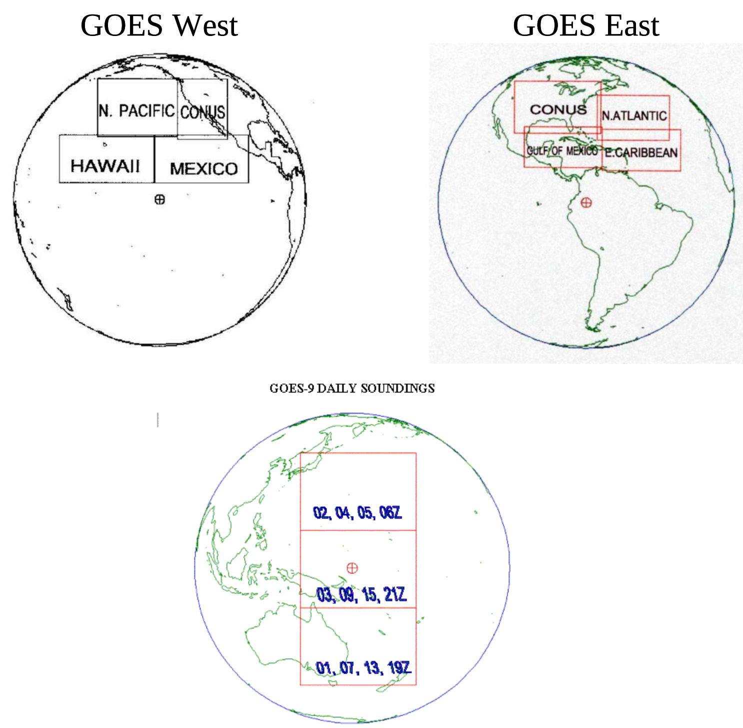

Figure 1. Domains over which CTOP is being evaluated; boxes indicate regions where NESDIS CTP product is available.

is often below the cloud-top level reported by pilots. As part of the study, the radiosonde-based cloud top (RCT) and the radar-based echo top (ET) observations are being compared to each other and to the NESDIS CTP and CTOP values, to understand their correspondence and relationships. It is important to understand that all of the cloud-top measurements have underlying uncertainties, and all of them pose difficulties for matching to the CTOP values. For example, the NESDIS CTP has coarser nominal horizontal resolution than CTOP (10 km vs. 4 km), which changes with latitude; radiosonde estimates of cloud top are inferred from temperature and relative humidity profiles; and radar-based estimates are based on echo top rather than cloud top.

Takacs et al. (2004) summarized the global observational datasets available for verification of a variety of types of oceanic weather forecasts. The datasets used for evaluation of CTOP are described in greater detail in the following section. The CTOP diagnoses are being evaluated over the Pacific, the Gulf of Mexico, Mexico, Hawaii, and Caribbean domains as shown in Fig. 1 using data collected during the spring of 2004. For the August/September evaluation, the eastern CONUS domain is also included.

3. TECHNIQUES FOR MEASURING CLOUD-TOP HEIGHT

3.1 CTOP diagnostic

The CTOP is designed to estimate the heights of optically thick clouds. Figure 2 shows an example of the product over the Pacific Ocean region. CTOP is based on an algorithm developed by the Naval Research Laboratory (NRL). The algorithm is based on brightness temperature measurements from the IR window channels (10.7 µm) from the imagers on the GOES satellites. These observations are combined with temperature profiles from a numerical weather prediction model to estimate the cloud top pressure (http://www.rap.ucar.edu/projects/owpdt/documents/ocndnrl.html/). In this case, temperature soundings from the National

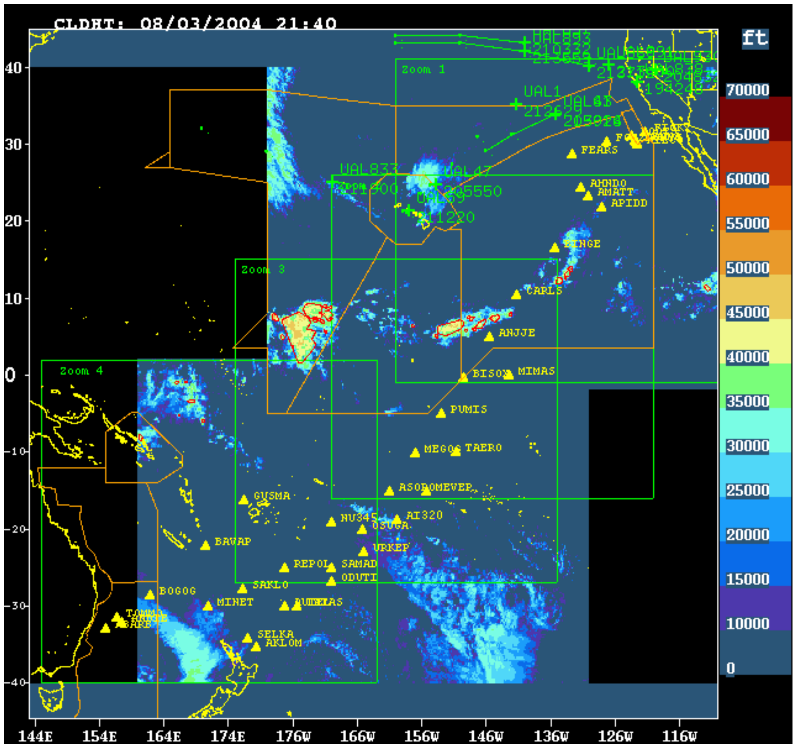

Figure 2. Example of cloud-top height product (CTOP) over the Pacific Ocean region (8/3/2004; 21:40 UTC)

Centers for Environmental Prediction (NCEP) GFS numerical weather prediction model are used as input to the CTOP for identifying the cloud top pressure. Altitude is then estimated using the U.S. Standard Atmosphere.

CTOP is provided on an approximately 4-km horizontal grid for cloud tops above 850 hPa, and is updated approximately every 20 minutes as new satellite data become available. Although CTOP is designed to be best at estimating cloud-top heights for optically thick clouds, the algorithm produces gridded estimates for all locations and types of clouds. To account for likely variations in performance, the statistical verification results will be stratified by the cloud opaqueness characteristics.

3.2 Description of verification datasets

3.2.1 Radiosonde observations

Radiosondes measure the vertical profiles of atmospheric variables from the surface to the stratosphere and transmit the data via radio to ground-based receiving systems. Radiosondes typically measure temperature, humidity, and pressure, as well as wind direction and speed. The observations are available through the Global Telecommunication System (GTS) gateway and NOAAPort. The University of Wyoming database of processed global upper-air data is being used as a source for these data. Observations for more than 150 locations over the CONUS, coastal areas and islands are obtained twice per day, at 0000 and 1200 UTC. In a few cases, 0600 and 1800 UTC soundings also are available. Using the radiosonde observations, several measures of cloud-top pressure/height values are being obtained:

- Radiosonde-based cloud-top height/pressure (RCT); determined using the revised analysis method of Wang and Rossow (1995). Using this technique, the cloud-top pressure is estimated using a top-down examination of the relative humidity with respect to water and ice (RH_w and RH_i, respectively). The cloud-top pressure is set to (a) the highest level where either RH_w or RH_i > 87% or (b) the highest level where RH_w or RH_i > 84% *and* RH_w or RH_i at the level above is at least 3% lower than the respective value at the level in question.

- Equilibrium level (EL; formerly called "expected cloud-top height"); EL is the level which a parcel from the lowest 500 m of the atmosphere reaches when it is lifted dry adiabatically to the lifted condensation level, and then moist adiabatically to a level where the temperature of the parcel is equal to the environmental temperature. If more than one EL exists, the highest one is chosen.
- Equilibrium level (ELV); computed in the same way as the EL but using virtual temperature.

It should be noted that the uncertainty associated with the RCT values can be quite large. Sounding analysis results can overestimate cloud-top pressures (i.e., underestimate cloud-top heights), especially in regions where cloud tops are frequently colder than -40^{o}C, such as in deep convection. This overestimation can reach as much as 170 hPa, especially in the tropics.

Radiosondes drift with the wind and it usually takes one and a half to two hours for the balloon and instrument package to ascend to the stratosphere and for the ground-based receiving system to obtain all the information transmitted. Due to the drift, the actual location and time of the cloud-top height measurement based on the radiosonde observations can be far from the place and time of the launch of the balloon. To reduce error due to drift, the time and location of the estimated cloud-top height are also determined. Based on past research, 5.5 ms^{-1} ascension rate is assumed for this study. To calculate the horizontal drift between any two levels of the sounding, the elevation at the lower level is subtracted from the elevation at the higher level to obtain the change in elevation. This value is divided by the ascension rate, 5.5 ms^{-1}, to get the time in seconds. The wind speed and wind direction at the upper level are considered as a velocity vector which, when multiplied by the time in seconds, gives the drift (horizontal displacement) of the radiosonde from the lower level to the higher level. The horizontal drift for a given level is determined by summing the drift from all levels below. In this way, the location of the estimated cloud-top height is calculated (Fig. 3).

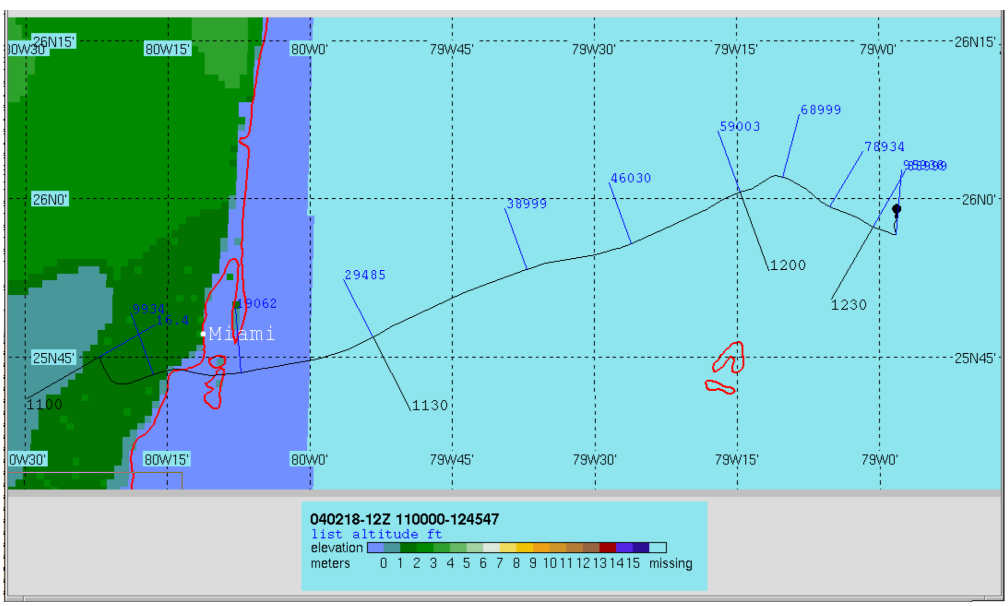

Figure 3: Example of the drift of a radiosonde launched from Miami at 1100 UTC on 2/18/2004. Black and blue numbers connected to the drift path indicate the time (UTC) and altitude (ft), respectively.

3.2.2 Radar observations

The radar echo top (ET) product shows the highest altitudes of precipitation echoes (i.e., reflectivities) across a 4-km grid. ETs are similar to cloud tops, but usually the top of a precipitating cloud will be somewhat higher than the top of the precipitation echo. Moreover, only the echo tops of precipitating clouds can be estimated by ET, whereas other methods are able to distinguish the tops of non-precipitating clouds. ET values are derived from reflectivity observations made by the National Weather Service (NWS) WSR-88D radars and are provided as a product for those radars in the NEXRAD Information Dissemination Service (NIDS) data stream. ET values are measured in thousands of feet above msl, and echo tops between 5,000 and 70,000 ft can be detected. The observations have a six-min temporal resolution and cover a range of 230 km around all radars. Nearly 30 radars are located inside the verification domain in coastal areas and on islands, and about 125 additional radars located across the CONUS.

The ET algorithm is constrained in several ways that reduce the accuracy of the echo top heights estimated by the ET product. For instance, the precision of echo top height measurement decreases with range due to beam broadening. At a range of 230 kilometers, the half-power beam width is 4,000 m. In addition, the ET algorithm uses a reflectivity threshold value of 18.5 dBZ to define echo tops. This threshold may not correspond to what a pilot sees as the cloud top. Further reduction in height accuracy stems from the resolution of the mapping routine. That is, the ET value for a particular 4x4 km box is defined as the highest of all (possible several) echo tops measured in that box.

Uncertainty in the ET values also results from the volume coverage pattern (VCP) in use. Since the different VCPs have gaps between various elevation angles, there is uncertainty regarding the 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 newer version of VCP is the best for estimating ET. Based on a study by Brown et al. (2000), the optimum distance range for estimating ET is between 45 and 120 km. Limiting ET computations to this range will reduce height estimation uncertainty that is related to the VCP. The beam at an elevation angle of 19.5^{o} would be likely to undershoot the tops of echoes that are closer than 45 km in range and at altitudes greater than 50,000 feet. Because the WSR-88D radars do not scan directly overhead, they also are unable to detect the true tops of echoes that are directly over the radar site. At ranges greater than 120 km, the lowest beam (0.5^{o}) is likely to overshoot the echo tops for many precipitating systems.

3.2.3 Pilot reports

Cloud-top height estimates are sometimes included in pilot's reports (i.e., PIREPs, AIREPs). These estimates will serve as a supplemental dataset for verifying the CTOP, especially over the CONUS where these observations are the most frequent. In addition, pilots frequently mention in the remarks section of their reports that the sky is "clear above." The altitudes of these reports will be used as an upper bound for the elevations of cloud tops in the region around the PIREP. These reports have been used effectively in previous verification studies for in-flight icing algorithms (e.g., Brown et al. 1997).

3.2.4 NESDIS Cloud Top Pressure Product (CTP)

Because of its wide spatial coverage, the NESDIS CTP is a very important observation dataset for evaluation of the OW CTOP product over the various domains. This product provides estimates of cloud-top pressure on a grid where the nominal resolution is 10 km, but varies with latitude (e.g., it is 14 km at 35^{o} latitude) across the domains shown in Fig. 1. The CTP estimates are updated approximately twice every hour. The NESDIS CTP product has good overall coverage, and has been extensively validated (Schreiner et al. 2001). Although CTP is based on observations from the GOES satellites (like CTOP), the CTP algorithm uses data from the GOES sounder rather than the imager, which is used to create CTOP diagnoses. In addition, the CTP utilizes the $C0_2$ absorption technique rather than the IR window technique that is used to derive the CTOP product. Because the sounder is used to estimate the cloud top pressure, CTP also does not directly depend on the output of a numerical weather prediction model.

4. VERIFICATION METHODOLOGY

4.1 Matching methods

Due to the scarcity, non-systematic nature, and general non-uniformity of most oceanic weather observations, the approach for matching forecasts and observations by necessity must be driven by the observations. That is, CTOP can only be verified at locations where there are observations. For each valid time of the CTOP product, a search is conducted to determine what observations are available.
Only a +/- 30-minute time difference is allowed. Two general approaches are being employed:

- A point-to-point approach is being used for cases where all of the observation types are available. For this analysis, the maximum and median CTOP, ET and NESDIS CTP values are being determined in a suitable region around the exact time and location of the radiosonde-derived cloud top (RCT). In addition, the CTOP values (within the relevant region and time period) that provide the *best* match to the other observations are being recorded. The "best" values provide an upper bound on the strength of correlation between the CTOP and the other sets of measurements.
- Grid-to-grid analysis will be used for the satellite-derived data (NESDIS CTP Fig. 4a and CTOP Fig.4b). The spatial and time scales are determined by the scales of the observations. The procedures applied in this case are similar to those used for the point-to-point matching process.

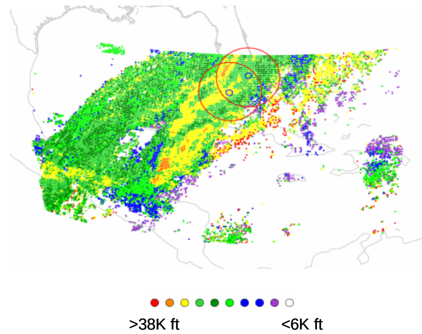

Figure 4a. NESDIS CTP over Miami (2/18/2004). Red circles represent radar, blue circles radiosonde stations.

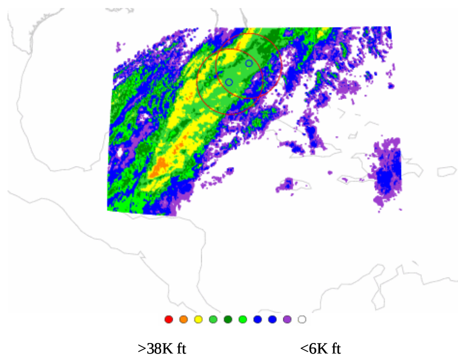

Figure 4b. As in Fig. 4a, but CTOP

4.2 Statistical evaluation

A variety of verification statistics, both categorical and continuous, will be computed for comparisons of the CTOP values with each type of observation (Table 1).

For the categorical statistics (e.g., probability of detection, POD), "Yes" and "No" categorical values of cloud-top height will be determined using suitable thresholds. The choice of these thresholds will be determined from the data collected. All of the statistics will be stratified by opacity, altitude, region, and other relevant factors. As noted earlier, because no direct measures of cloud-top height are available, the main focus will be to evaluate how consistent the CTOP values are with the other estimates.

Table 1. Statistical measures to be used to evaluate CTOP.

Statistic	***Definition***	***Description***	***Interpretation***	***Range***
PODy	YY/(YY+NY)	Probability of Detection of Yes observations	Proportion of Yes observations that were correctly forecasted	0-1 Best: 1 Worst: 0
PODn	NN/(YN+NN)	Probability of Detection of No observations	Proportion of No observations that were correctly forecasted	0-1 Best: 1 Worst: 0
Curve Area	Area under the curve relating PODy and 1-PODn	Area under the curve relating PODy and 1-PODn (i.e., the ROC curve)	Overall skill (related to discrimination between Yes and No observations)	0 to 1 Best: 1 No skill: 0.5
Bias(cont)	$\overline{y} - \overline{o}$	Difference between the average forecast and the average observation.	Measure of over or under forecasting	Unbiased: 0 Neg.: under forecasting Positive: over forecasting
Bias(dichot)	$\frac{a+b}{a+c}$	Ratio of the number of "yes" forecasts to the number of "yes" observations	Measure of over or under forecasting	Unbiased: 0 Neg.: under forecasting Positive: over forecasting
Mean Absolute Error	$\frac{1}{n}\sum_{k=1}^{n}\lvert y_k - o_k\rvert$	Arithmetic average of the absolute values of the differences between the members of each pair	Typical magnitude for the forecast error in a given verification dataset	Perfect MAE = 0
Mean Squared Error	$\frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2$	Average squared difference between the forecast and observation pairs	This measure is similar to the MAE except that the squaring function is used rather than the absolute value function to account for larger errors	0 to 1 Best: 0

5. SUMMARY AND CONCLUSION

Strategies for evaluating the diagnoses provided by the oceanic cloud-top height product have been outlined. The paper has described the efforts employed to obtain independent datasets for the evaluation. This process is made difficult because no direct cloud-top height observations of cloud-top height are generally available over the oceans. Cloud-top height values estimated using several different methods, based on as many different data sources as possible over the coastal areas and islands are being employed for the evaluation. The generated values are being compared to each other and to the CTOP product to find similarities and/or differences among them. In general, there has been relatively little experience measuring cloud-top height over the oceans. To gain more confidence in the verification of CTOP over the oceans, the verification is also being performed over the CONUS where some PIREP observations are available for the evaluation. In addition, greater knowledge about cloud occurrence and distribution exists for terrestrial regions. The relationships between the NESDIS CTP and the other cloud-top height estimates have special importance. Understanding these relationships will aid in the interpretation of the verification results over open ocean regions, where the evaluation is being extended using only the NESDIS CTP.

6. ACKNOWLEDGMENTS

This research is in response to requirements and funding by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy of the FAA. NCAR is sponsored by the National Science Foundation.

REFERENCES

Brown, B.G., G. Thompson, R.T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Weather and Forecasting*, **12**, 890-914.

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

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

Takacs, A., B. Brown, and J. Mahoney, 2004: Verification of oceanic weather diagnoses and forecasts for aviation weather elements. *Preprints, 84thAnnual Meeting of the American Meteorological Society*. Seattle, WA 11-15 January, American Meteorological Society (Boston).

Wang, J., and W.B Rossow, 1995: Determination of cloud vertical structure from upper-air observations. *Journal of Applied Meteorology,* **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. *Journal of Climate,* **13**, 3041-3056.