Submission to Atmospheric Measurement Techniques

Near Global Distributions of Overshooting Tops Derived from Terra and Aqua MODIS Observations

Yulan Hong, ~~Robert J. Trapp, ~~Stephen W. Nesbitt, -Robert J. Trapp, and Larry Di Girolamo

Department of Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, Illinois, USA

Corresponding author: Yulan Hong

Email: yulanh@illinois.edu

Abstract

Overshooting cloud tops (OT) form in deep convective storms when strong updrafts overshoot the tropopause. An OT is a well-known indicator for convective updrafts and severe weather conditions. Here, we develop an OT detection algorithm using thermal IR channels and apply this algorithm to about 20-year MODIS data from both Terra and Aqua satellites to form an extensive, near global climatology of OT occurrences. The algorithm is based on a logistic model which is trained using A-Train observations. We demonstrate that the overall accuracy of our approach is about 0.9 when the probability of the OT candidates is larger than 0.9. The OT climatology reveals a pattern that follows the climatology of deep convection, as well as shallow convection over the mid-latitude oceans during winter cold air outbreaks. OTs appear most frequently over the Intertropical Convergence Zone (ITCZ), central and southeast North America, tropical and subtropical South America, southeast and south Asia, tropical and subtropical Africa, and northern middle-high latitudes. OT spatial distributions show strong seasonal and diurnal variabilities. Seasonal OT variations shift with large-scale climate systems such as the ITCZ and local monsoonal systems, including the South Asian Monsoon, North American Monsoon and West African Monsoon, etc. OT diurnal variations agree with the known diurnal cycle of convection: Maximum OT occurrences are in the afternoon over most land area and around midnight over ocean; and the OT diurnal cycle is stronger and more varied over land than over ocean. OTs over land are usually colder than over ocean except around 10:30 am. The top 10 coldest OTs from both Terra and Aqua mostly occur over land and at night. This study provides OT climatology for the first time derived from two-decade MODIS data that represents the longest and stable satellite records.

1. Introduction

An overshooting cloud top (OT) forms when a convective-storm updraft penetrates the level of neutral buoyancy and thus extends into the upper troposphere-lower stratosphere (UTLS). OTs and their associated strong updrafts have been found to be an important transport mechanism for water vapor and other atmospheric constituents into the stratosphere, thus impacting the chemical composition and radiation budget of the UTLS (e.g. Gettelman et al., 2002, 2004). They are often used as indicators of hazardous weather conditions such as strong winds, large hail, flooding, and tornadoes at the Earth's surface (Bedka et al., 2018; Dworak et al., 2012; Marion et al., 2019). More generally, the characteristics of OTs express information about the characteristics of the related updrafts well below cloud top, including the convective mass flux through the troposphere, which is an important parameterized quantity used in global climate models.

In addition to the expectation of a connection between updraft strength and OT depth (Heymsfield et al., 2010), Trapp et al. (2017) has shown a strong link between updraft core area and OT area (OTA), indicating that a relatively intense and wide mid-tropospheric updraft core area will tend to have a large OTA. Given that the direct measurements of updrafts within intense convective environments are either from a few ground-based radars or several field campaigns, these studies suggest a pathway for characterizing global updraft and updraft-size distributions by quantifying the global OT distributions and characteristics from space.

Toward this end, the first step is to detect OTs. Geostationary satellite imagery provides the opportunity to study OT occurrence over a wide region with fine spatial and temporal resolutions. A series of OT detection algorithms have been developed based on geostationary satellite observations. A commonly used OT detection method utilizes the brightness temperature (T_b) difference (BTD) between Infrared (IR) water vapor (WV) and IR window channels (IRW) (Schmetz et al., 1997). The WV-IRW BTD method is based on the fact that water vapor transported into the lower stratosphere absorbs and emits more radiation at a water vapor channel (such as 6.7μm) compared to a window channel (such as 11 μm). Thus, positive BTD is usually observed in the OT regions. However, in convective anvils (e.g. Hong & Di Girolamo, 2020; Setvák et al., 2013) or in polar winter conditions when strong radiation inversions exist near the surface (Ackerman, 1996), positive BTDs are also observed, which pose challenges to differentiate OTs from these cases.

Another commonly used OT detection method is the IR Window (IRW) texture approach (Bedka et al., 2010). This method uses a threshold of 215 K T_b at IR window channel to first select OT candidates. These candidates are also colder than the tropopause temperatures. In the second step, surrounding anvil is sampled at a ~ 8 km radius in 16 directions. At each direction, pixels with T_{b11} colder than 225 K are included in calculating cirrus mean T_{b11}. The selected candidate is considered as an OT if the T_{b11} difference between the pixel and its surrounding cirrus is larger than a threshold of 6.5 K. The IRW texture approach has been widely applied for OT detections observed from space such as geostationary satellite imagery and Moderate Resolution Imaging Spectroradiometer (MODIS) (Bedka, 2011; Dworak et al., 2012; Griffin, 2017; Griffin et al., 2016; Monette et al., 2012; Proud, 2015). However, the strictly fixed thresholds of IRW texture method limit its ability to detect warm OTs that commonly occur in the mid-latitude regions, leading to seasonal and regional biases (Bedka & Khlopenkov, 2016). Based on the visible (VIS) and IR imagery, Bedka and Khlopenkov (2016) developed a new probabilistic OT detection algorithm to minimize the dependence of IRW texture method on thresholds. Khlopenkov et al. (2021) further updated this algorithm by incorporating the normalized tropopause temperature, surrounding anvil area and spatial uniformity. Improved accuracy is achieved with this probabilistic OT detection algorithm compared to the IRW texture method.

Observations from spaceborne active sensors have also been used for exploring OT detections. For instance, the cloud profiling radar (CPR) on CloudSat (Stephens et al., 2008) was used for validating the passive satellite-based OT detection methods (Bedka et al., 2010; Dworak et al., 2012; Rysman et al., 2017), calculating the heights of OTs (Griffin et al., 2016), and understanding WV-IRW BTD variability in OT regions (Setvák et al., 2013). The combined CloudSat-CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) data was also used for detecting OTs, which led to the creation of a 12-year OT database (Li et al., 2022). As demonstrated by these studies, the CloudSat-CALIPSO observations are powerful in detecting OTs and gauging OT depths, but they are only available in a narrow swath that leads to a lack of knowledge of three-dimensional (3-D) OT structures and large uncertainties in their coverage (Astin et al., 2001). The precipitation radar on Tropical Rainfall Measuring Mission (TRMM) or Global Precipitation Mission (GPM) can provide 3-D depictions of storm structures. The precipitation radar observations have been used to investigate OT climatology including their geodistributions, area and diurnal cycles in the tropical regions ($20°S$ – 20°N) (Alcala & Dessler, 2002; Liu & Zipser, 2005) and over broader areas ($60°S$ – 60°N) (Hourngir et al., 2021; Liu et al., 2020; Liu & Liu, 2016).

In addition, using three water vapor channels of the Advanced Microwave Sounding Unit B (AMSU-B), convective overshooting detection method was developed through the microwave technique (Hong et al., 2005). A seven-year OT climatology based on AMSU-B was derived in the tropical and subtropical areas that shows OT interannual to diurnal variations (Hong et al., 2008).

While many OT detection algorithms have been developed either using passive or active remote sensing techniques, their use toward quantifying OT occurrences and attributes from space are mostly from datasets with coarse ~~large~~ spatial resolutions, e.g. ≥ 2 km for geostationary satellites, 4-5 km for TRMM precipitation radar,5 km for GPM Ku radar, and 15 km for AMSU-B. Spatial resolution of observations significantly influences variations of WV-IRW BTD (Setvák et al., 2007) and thus influences the choice of T_b and BTD thresholds. Large spatial resolution also poses challenge in identifying OTs of small size and affects the accuracy of computing OT attributes such as OT area. Therefore, measurements from space with a higher spatial resolution will support a better characterization of OT climatology globally, which has not been derived so far.

The MODIS instrument (King et al., 1992) acquires data at a high spatial resolution (≤ 1 km) that allows to detect small OTs. This sensor has a wide view swath of 2330 km which is able to take a whole picture of a mesoscale system. It is operating on both Terra and Aqua satellites, overpassing the same latitude at four different times each day: around 1:30am/pm and 10:30am/pm equator-crossing time (ECT). In the last twenty years, both Aqua and Terra satellites have a consistent equator-crossing time, making the MODIS data the longest stable climate records from space.

To utilize these climate records, the main objective of this study is to show a near global climatology of OT occurrence derived from about 20-yr Aqua and Terra MODIS data. Owing to the relatively high spatial resolution of MODIS, this climatology includes OTs in small size that missed by GPM radar. It includes both the tropical and mid-latitude regions, and thus makes complementary to the climatology by Liu & Zipser (2005) and Hong et al., (2008) that were only focused on tropical and subtropical regions. It also provides OT diurnal information at four observation times. To achieve these objectives, we first develop an OT detection algorithm that is specifically designed for MODIS, works for both day and nighttime data, and is more flexible to thresholds compared to those used in Bedka et al. (2010) and Li et al. (2022). In sect. 2, we will present the details of data and methods used for developing the OT detection algorithm. Validation of the algorithm will be discussed in Sect. 3. Section 4 discusses the results produced from our OT detection algorithm. Finally, in sect. 5, we conclude the findings of this study.

2. Data and Methodology

In order to develop a method that can detect OTs during both daytime and nighttime, this study uses observations from multiple sensors onboard multiple platforms as well as a machine learning method – logistic regression. The OT detection algorithm is developed in two main steps. First, we manually identified a number of OT candidates from the combined CloudSat-CALIPSO data. The infrared radiative characteristics of these OTs extracted from the combined Aqua MODIS infrared data serve as inputs to train the logistic regression. Second, we applied the regressed model to the Terra and Aqua MODIS data for automatic OT detection. We call this method an *IR algorithm*.

2.1 Satellite and Reanalysis Datasets

2.1.1 CloudSat and CALIPSO

The CloudSat and CALIPSO satellites are two members of the afternoon constellation in a sun-synchronous orbit with an Equator-crossing time at 01:30/13:30 local time (LT). The cloud profiling radar (CPR) onboard CloudSat is a near-nadir-view radar operated at 94 GHz (~ 3.3 mm). Measuring radar reflectivity factor, the CPR probes the vertical structure of hydrometeors with a minimum sensitivity of about -30 dBZ (Stephens et al., 2002, 2008). The radar's footprint is 1.8 km along track and 1.4 km cross track. Its vertical resolution is 480 m with a resampled resolution of 240 m. The radar is able to penetrate thick clouds and therefore is suitable for OT identification as demonstrated by previous studies (Chung et al., 2008; Rysman et al., 2017; Setvák et al., 2013). The radar reflectivity factor from the 2B-GEOPROF (Version P1) product (Marchand et al., 2008) that shows time-height cross sections (curtains) of clouds and precipitation was used for manual OT identification.

The CALIPSO flew about 15 s after CloudSat during the time period of observations used in this work. The lidar onboard CALIPSO operates at 532 nm, having a vertical resolution of 30 m below 8.2 km and 60 m above 8.2 km (Winker et al., 2003). The lidar is sensitive to optically thin clouds and aerosols. The 2B-CLDCLASS-LIDAR product, provided by the CloudSat Data Processing Center, reports cloud top and base heights for up to five layers (Wang et al., 2012). This product utilizes the complementary features of the CloudSat radar and the

CALIPSO lidar, and thus includes thin cirrus clouds. The cloud top height of the topmost layer was used to aid identifying OTs. Two years of 2B-GEOPROF and 2B-CLDCLASS-LIDAR data (2007-2008) were used in this study.

2.1.2 MODIS

MODIS onboard both the Aqua and Terra platforms has 36 discrete spectral bands between 0.415 to 14.235 μm with spectral-dependent spatial resolutions varying between 250 m to 1 km at nadir (Barnes et al., 1998; King et al., 1992). The Aqua satellite launched in May 2002 is a member of A-Train satellite constellation. Terra was launched in December 1999 in a sun-synchronous orbit with an Equator-crossing time at 10:30/22:30 LT (Platnick et al., 2003).

To obtain OT radiative characteristics, the MODIS Collection 6.1 Level 1B calibrated radiance data, MYD021KM from Aqua and MOD021KM from Terra, were used. In this study, the bands selected have center wavelength at 6.715 and 11.03 μm for OT detection that are used for deriving brightness temperature. The uncertainties associated with these two bands are within 1% for both Terra and Aqua MODIS (Xiong et al., 2005, 2018). Navigation files with 1 km resolution (MYD03 and MOD03) were used for the geolocation information. The Aqua MODIS data from 2007-2008 were collocated to the CloudSat-CALIPSO data for selecting OT cases as a training dataset for the logistic regression model (Sect. 2.2). The Terra MODIS data from February 2000 – 2021 and the Aqua MODIS data from July 2002 -2021 were used for deriving the OT climatology presented in Section 4.

2.1.3 GPM

The Global Precipitation Monitor (GPM) core observatory, launched in February 2014, carries the first space-borne Dual-frequency Precipitation Radar (DPR) that includes a Ka-band (35.5 GHz) radar (KaPR) and a Ku-band (13.6 GHz) radar (KuPR) (Hou et al., 2014). The KuPR measures 3-D structures of convective systems with a vertical resolution of 250 m and a footprint of 5 km over a swath of 245 km. The GPM KuPR echoes have been demonstrated to be effective in the study of deep convection reaching to tropopause (Liu et al., 2020; Liu & Liu, 2016). To utilize the GPM as an independent detection of OTs, we collocated the Ku-band echoes to the OT candidates identified from Terra MODIS as a validation of our IR algorithm (Sect. 2.2). About six years (March 2014 - 2020) of data from the 2A.GPM.DPR product (V06) was used.

2.1.4 Reanalysis Data

Tropopause temperature is needed for our IR algorithm. We used the tropopause information output from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), instantaneous two-dimensional collections, hourly, single-level diagnostics (MERRA2_400.inst1_2d_asm_Nx) product (Bosilovich et al., 2016). The MERRA-2 parameter 'TROPT' is a blended estimate of tropopause temperature (T_p) based on a combination of the World Meteorological Organization (WMO) definition of the primary lapse-rate tropopause (Grise et al., 2010) and equivalent potential vorticity. The tropopause data has a spatial resolution of 0.625° x 0.5° longitude-latitude. The closest MERRA-2 grid in space and time was assigned to each MODIS observation.

2.2 OT Identification Algorithm

2.2.1 OT Selections from A-Train Observations

The first step of the IR algorithm is to generate an OT training dataset. We manually selected OT candidates around the world from 2007 and 2008 by visualizing the CPR reflectivity factor from 2B-GEOPROF, topmost cloud top height from 2B-CLDCLASS-LIDAR, tropopause information from MERRA-2 and the collocated T_{b11} from Aqua MODIS. For the CloudSat-MODIS collocation, the nearest Aqua MODIS pixels were assigned to the CloudSat track. The distance of the collocated CPR–MODIS pixels is usually less than 700 m, allowing these two sensors to observe nearly the same cloud within one minute (Hong & Di Girolamo, 2020).

Parallax correction was examined, but not employed in this work. We found that parallax correction produced nearly identical T_{b11} values as those without parallax correction. This is consistent with ~~(~~Wang et al., (2011), who stated that the effect of parallax correction will be small if the cloud object occupies a large enough area and is relatively homogeneous over the range comparable to the parallax correction, which is certainly the case for the OTs being studied here. Additionally, the parallax correction introduces a small amount of noise in the along-track T_{b11} values due to noise in the cloud top height used in the parallax correction. The noise is sufficient at times to create artifacts in our OT algorithm.

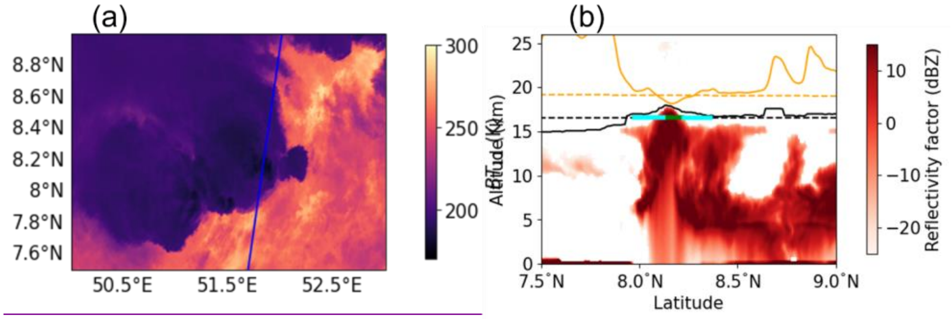
Formatted: Indent: First line: 0.5"

Field Code Changed

OTs were selected by visually inspecting the visualization rather than using a fix criterion. The OT selection basically followed four principles: T_{b11} colder than T_p, cloud top height above tropopause height, T_{b11} smaller than $T_{b6.7}$, and an obvious convective core. ~~.~~ For instance, Figure 1 shows an example of how we manually select OTs from this visualization. Figure 1a displays that CloudSat overpassed a strong convective system with T_{b11} as low as 180 K. Figure 1b shows the curtain of the radar reflectivity factor from CloudSat for this convective system, along with $T_{b11}/10$ (orange-solid line), cloud top height (black-solid line) and tropopause information (orange-dash for tropopause temperature (T_p) divided by 10, and black-dash for tropopause height~~-~~) along the transect. As Figure 1b shows, in the convective core, cloud top height is above the tropopause height, and the T_{b11} is colder than tropopause temperature (T_p). This case is identified as an OT. In total, we have selected 209 OTs from A-Train observations. ~~.~~ Additionally, 78 non-OTs (NOTs) were also selected for model training. The NOTs share very similar characteristics with OTs, i.e. T_{b11} is cold (close to or colder than T_p) and has a local minimum, but no obvious convective core ~~overshoot top~~ is observed from the visualization. Supplementary 1 displays four OT and three NOT cases. Figure 2 shows very similar OT and NOT T_{b11} distributions.

The total 287 samples were randomly distributed over four seasons and in different locations on Earth. The d~~The complexity of the dataset will help increasing model performance.~~ ~~D~~ata is available in Supplementary 2.

Formatted: Indent: First line: 0.5"



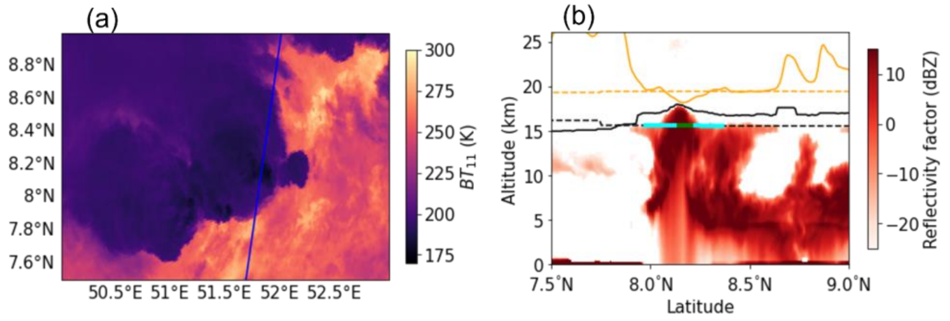


Figure 1. An OT case occurring at night over the Indian Ocean on June 1st, 2007: a) Brightness temperature at 11 μm from Aqua MODIS, with blue line indicating the CloudSat track; b) Vertical cross section of CloudSat radar reflectivity factor overlapped with MERRA-2 tropopause temperature divided by 10 (orange-dashed), tropopause height (black-dashed), topmost cloud top height from 2B-CLDCLASS-LIDAR (black-solid), and T_{b11} /10 (orange-solid). The green line in b) indicates the OT region along CloudSat track determined by the method from Marion et al. (2019), and the cyan line indicates the surrounding cirrus anv~~a~~il.

2.2.2 OT Edge and Cirrus Anvil

Once an OT was manually selected from the A-Train data, OT edges were determined using the method described in Marion et al. (2019). Briefly, the local minimum T_{b11} along the CloudSat track was set as the OT center. The 1-D second derivative along two radii along CloudSat track ($\frac{d^2T_b}{dr^2}$) was computed using three-point Lagrange interpolation after Gaussian filter was applied to smooth T_{b11} field. The OT edges along the two radii are defined as the first point where $\frac{d^2T_b}{dr^2} \leq 0$. With the OT edges determined, the diameter of the OT candidate can be obtained. As an example, Figure 1b shows the OT diameter in green, indicating that this method well catches the overshooting area.

The cirrus (Ci) anvil in this work was searched within 20 pixels around the OT center but with the OT area excluded. Pixels starting from the OT edge and having T_{b11}< 260 K contribute to the surrounding cirrus. A value of 260 K was used to screen cold clouds. This threshold has been commonly adopted for screening high clouds associated with deep convection (Chung et al., 2007; Tian et al., 2004). Figure 1b indicates the cirrus anvil in cyan. Once two edges of an OT and its cirrus anvil were determined, the OT center T_{b11}, the mean brightness temperature for the OT region ($\overline{T}_{b6.7}$ and $\overline{T}_{b11}$), the mean brightness temperature ($\overline{T}_{b11}$) for surrounding cirrus averaged over two radii and the tropopause temperature (T_p) for the OT case were recorded to construct the training dataset. Note that c~~C~~irrus anvil is only defined in~~as~~ a small area within 20 km from the OT center, which is sufficient to check Tb difference between the OT and its surrounding region because OT's size is usually smaller than 15 km (Bedka and Khlopenkov, 2016). Pixels outside the 20-km radii can also contain cirrus, but do not contribute to our calculation of cirrus T_b.

For the 209 OT candidates, all of them have their diameters less than 25 km, 180 OTs (86%) have their diameters less than 15 km, and the peak in the OT diameter distribution is about 10 km (Fig. 2a), being agreeable with Bedka & Khlopenkov, (2016) which states that OTs are typically less than 15 km in diameter. The T_{b11} of OT center along the CloudSat track is shown in Fig. 2b which displays an asymmetric U-shape distribution along latitudes. Tropical OTs tend to have their center T_{b11} less than 200 K, while mid-latitude OTs tend to have center T_{b11} colder than 230 K. The NOT candidates share a very similar T_{b11} distribution with OTs. We rarely found OTs outside the ±60-degree latitude range. The selected OTs and NOTs are available in Supplementary 2. In addition, ~~for all OT candidates, WV-IRW BTD ($T_{b6.7} - T_{b11}$) is found to be positive, and tropopause temperature is warmer than OT center T_{b11}. For NOTs, they also have positive BTD, but 16% of them are warmer than tropopause temperature.~~ WV-IRW BTD and T_p are two important variables used for our IR algorithm.

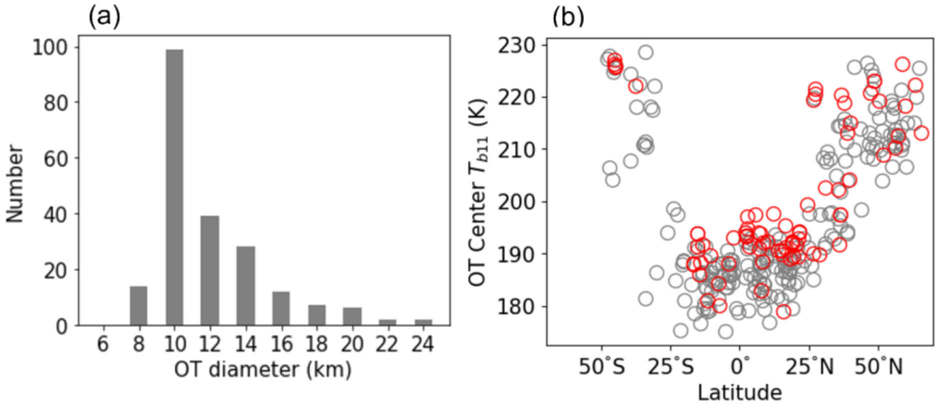


Figure 2. (a) OT diameter distribution of the 209 OT candidates selected from 2007 and 2008 A-train data, and (b) brightness temperature at 11 µm of OT (grey) and NOT (red) center along CloudSat track.

2.2.3 Logistic Regression

Similar to Bedka & Khlopenkov (2016), a probability was generated for an OT candidate. The 209 OTs and 78 NOTs selected from A-Train observations served as inputs for the logistic model. The logistic regression is a statistical model that is used to model a certain event through assigning a probability between 0 and 1 such as classification of OT and NOT. The logistic model depends on several variables or predictors, shown as

$$P = \frac{1}{1+e^{-(b_0+\sum_i^n b_i x_i)}} \quad , \quad (1)$$

where P is the probability of an OT candidate, b_0 is the constant, x_i is the variable and b_i represents the regressed coefficient.

Three MODIS-based variables were settled on after a series of tests to optimize the accuracy. They are x_1- the difference between Ci anvil mean T_{b11} and OT center T_{b11}, x_2- the difference of T_p and OT center T_{b11}, and x_3- the difference of mean $T_{b6.7}$ ($\bar{T}_{b6.7}$) and mean T_{b11} ($\bar{T}_{b11}$) of OT. 156 OTs and 48 NOTs were used to train the model and the regressed results are summarized in Table 1. The total accuracy is about 84% when probability > 0.6 is predicted to be an OT. 53 OTs and 30 NOTs were used to validate the regressed model with a total accuracy about 82%. ~~Note that we have obtain a stable model accuracy of ~ 0.84 when the total sample size is larger than 150. A total sample size of 287 cases should ensure the robustness of the regressed model.~~ We have tested the accuracy as a function of sample size and found a stable model accuracy of ~84% when total sample is larger than about 150, indicating that total samples of 287 should be sufficient for training a robust model in our study.

Table 1. A summary of the regressed coefficients ~~(significant at the 99% level)~~ for the variables selected for OT detection used in Equation 1.

Variables	Coefficients for the variables* (constant = -3.2397 ± 0.753)
~~b_0~~	~~-3.2397~~
x_1- difference between Ci mean T_{b11} and OT center T_{b11}	0.2075 ± 0.053
x_2- difference of tropopause T_p and OT center T_{b11}	0.3516 ± 0.085
x_3- difference of averaged OT $T_{b6.7}$ and averaged OT T_{b11}	0.4996 ± 0.188

* The Logistic Regression is based on python module – Statsmodels. The P values for the regressed coefficients are smaller than 0.008 based on z-test. The uncertainties for each coefficient are represented by standard errors.

2.2.4 Application of IR Algorithm to MODIS

The logistic regression in Sect. 2.2.3 forms the basis of our IR algorithm, which aims to automatically identify OTs from Terra and Aqua MODIS in the daytime and at nighttime. The application of the IR algorithm starts from pixel search with T_{b11} colder than T_p, and T_{b11} less than 200 K in the tropics (within 25° latitude) or less than 230 K in the midlatitudes (outside 25° latitude). These T_{b11} thresholds selected to ensure that all OTs identified in Fig. 2b would pass this first OT candidate selection criteria. If the pixel passed these thresholds and is a local minimum in T_{b11} field in a 41 km x 41 km window, we continued to find the OT edges in eight directions using the method by Marion et al. (2019), as mentioned in Sect. 2.2.2. ~~OT~~ $\bar{T}_{b11}$ and $\bar{T}_{b6.7}$ of the OT area are further computed over the pixels along eight radii once OT edges have been determined. $\bar{T}_{b11}$ of the surrounding cirrus is also computed in eight directions in the cirrus area as defined in Sect. 2.2.2. When the surrounding cirrus $\bar{T}_{b11}$ is warmer than OT center T_{b11} and this OT case shows positive WV-IRW BTD (i.e. $\bar{T}_{b6.7} - \bar{T}_{b11} > 0$), OT probability is calculated according to the logistic regression from Sect. 2.2.3. If one of the mentioned conditions does not satisfy, the algorithm will search for next pixel. The flowchart of the IR algorithm application is summarized in Fig. 3.

The window size of 41 km was adopted considering that 98% of the OTs (Fig. 2) have their diameters less than 20 km according to A-Train observations (Sect. 2.2.2). This window makes sure that two OT centers are at least 20 km apart and that enough pixels contribute to the cirrus anvils. If multiple OTs occurred in the same window, the one with the coldest T_{b11} was selected.

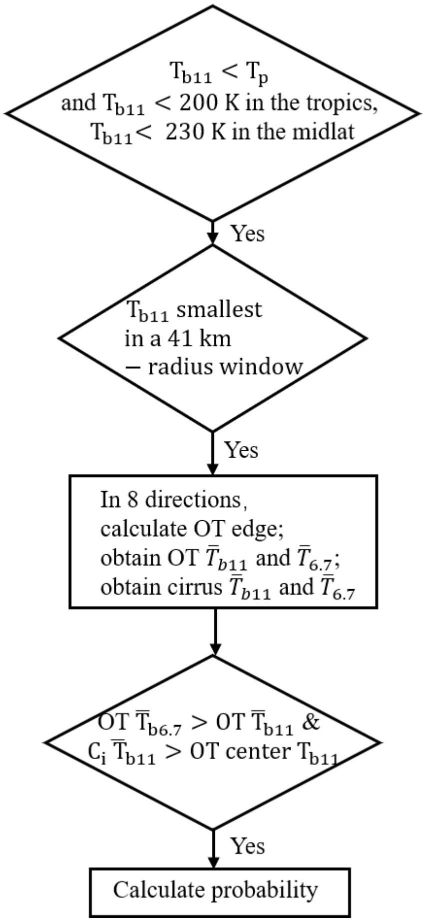


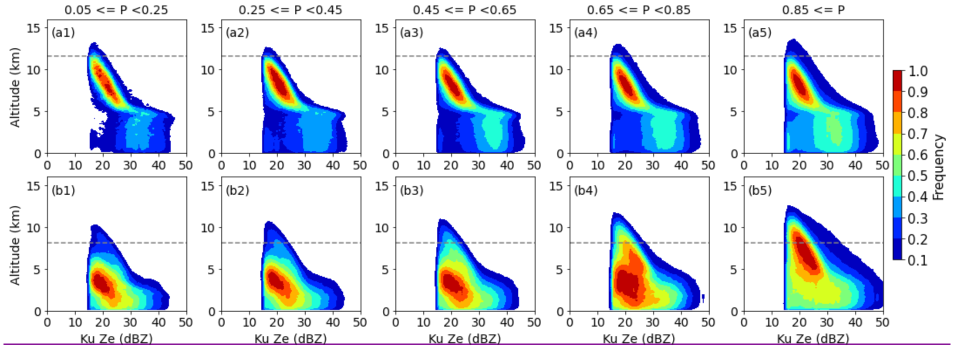
Figure 3. Flowchart for the application of IR algorithm to MODIS data. The $\bar{T}_b$ represents the mean of brightness temperature.

3. Validation of OT Detection Algorithm

3.1 Comparison with GPM

GPM has been demonstrated to be an effective tool in studying intense storms and overshooting top events (Hourngir et al., 2021; Liu et al., 2020; Liu & Liu, 2016). Here, we used the GPM observations for two purposes: To compare the performance of OT detection between GPM KuPR and Terra MODIS, and to investigate the cloud structure of detected OTs. The colocation between GPM KuPR and MODIS data was achieved when the time difference between them was within 5 minutes and the spatial difference between them was less than 10 km. A 5-minute time window was used because the life cycle of OTs can be as small as several minutes (Setvák et al., 2013). The collocating process was performed only when OT candidates were identified from Terra MODIS. We obtained 6949 colocations for the period of March 2014 – December 2020.

Ku-band radar reflectivity factor (Ze) in an area with a radius less than 40 km around the colocated radar pixel were collected to construct the contour frequency by altitude diagram (CFAD; Yuter & Houze, 1995). The parallax error between KuPR and MODIS could be more than 20 km according to the method described in Wang et al. (2011). Also, OT diameter is likely less than 20 km. An area with a 40 km radius for the collocated KuPR data is likely able to encompass the OT event identified by MODIS. Figure 4 shows the CFADs contributed by all (6949) collocated OT cases. The CFADs were segregated into 5 OT probability intervals for the tropical and mid-latitude areas. As shown, the largest frequency occurs above 5 km in tropical areas (Figs. 4a1-a5). As the OT probability increases, the frequency increases for large Ze (> 30 dBZ) below 5 km. In the midlatitudes (Figs. 4b1-b5), higher frequency of the Ze occurs below 5 km when OT probability is less than 0.85. For those OT cases with P > 0.85, large frequency is mostly above 5 km, and large Ze (> 30 dBZ) occurs more frequently below 5 km. With an analysis of DPR rain type product, we noticed that the large Ze (e.g. > 30 dBZ) below 5 km tend to associate with convective rain in both the tropics and midlatitudes. An increase of convective rain in the CFADs with larger OT probability indicates more likely OT occurrence. These CFADs demonstrate that the probability generated from our IR algorithm indicates storm intensity and a confidence level of OT detection.



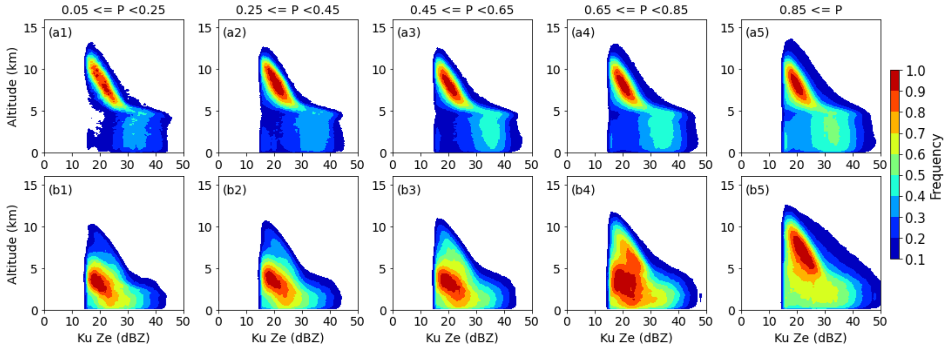


Figure 4. Contoured frequency by altitude diagram, showing the frequency normalized by the maximum bin of radar reflectivity. Data were binned at 1dBZ intervals at each level. The upper panels are for the tropics (within 25° latitude), and the lower panels are for the midlatitudes (between 25° and 60° latitude). ~~The dashed lines in upper and lower panels represent the mean tropopause height in the tropics and in the midlatitudes, respectively, derived from MERRA-2.~~

To compare the performance of OT detection between GPM and MODIS, we need to determine when GPM detects an OT. If the maximum altitude of 15 dBZ in the 40-km radius area was higher than 2 km below the MERRA-2 tropopause, an OT flag was assigned to the collocated GPM pixel. Previous studies also adopted a level below the tropopause as the OT reference considering the tropopause height variability (Sun et al., 2019; Zhuge et al., 2015) such as the noted double tropopause observed in deep convection (Vergados et al., 2014). Here, 2 km was selected due to an agreement of 67% between MERRA-2 and ERA-5 tropopause height (from ECMWF-AUX (Partain, 2007)) for the 287 OTs and NOTs cases used in Sect. 2. Once OT flags were assigned to the collocated GPM cases, agreement of OT detection between MODIS and GPM was calculated for a wide range of OT probability generated by the IR algorithm. The agreement is expressed as

$$Agreement = \frac{N(H > H_p - 2 \cap P_1 < P < P_2)}{N(P_1 < P < P_2)} \qquad (2)$$

where H is the maximum altitude (in km) of 15 dBZ in the 40-km radius aera, H_p is tropopause height from MERRA-2, and N is the OT numbers with OT probability between P1 and P2.

Figure 5 shows the agreement in OT detection between MODIS and GPM which increases with OT probability. In the tropics, the agreement is about 70% when P > 90% with enough samples, while in the midlatitudes, the agreement is larger than 80% when P > 0.90.

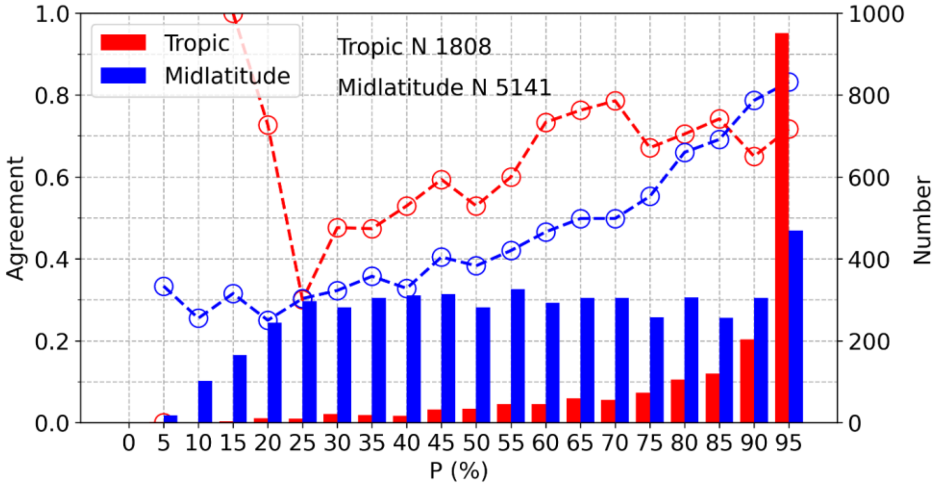


Figure 5. Comparison of OT detection between GPM and Terra MODIS. Curves represent agreement of OT detection between MODIS and GPM in various probability intervals, red for the tropics and blue for the midlatitudes; the numbers of potential OT candidates are shown in bars. N stands for sample number.

3.2 Manual Check

As a complement to GPM-MODIS comparison for assessing IR algorithm accuracy, we manually checked 1158 daytime OT candidates (selected randomly across the year) from Terra MODIS from 2018-2020 (data is available is Supplementary 3). These OT candidates are with a wide range of probability. OT and NOT flags were assigned to the candidates by visually inspecting the IR and visible images from the NASA Worldview website (https://worldview.earthdata.nasa.gov). The fraction of OT and NOT segregated at a 0.1 probability (generated from the IR algorithm) interval was calculated (Fig. 6). As displayed, the fraction of OT substantially increases when the probability is greater than 0.8 in both the tropics and midlatitudes. In the tropics, the fraction of NOT is about 30% when P is between 0.8 and 0.9, and it decreases to about 10% when $P \geq 0.9$. In the midlatitudes, when the P is small (e.g. < 0.8), NOT fraction is higher than OT fraction. Only when $P \geq 0.9$, NOT fraction drops to about 10%. With a manual check of about 900 OT candidates selected from July, 2018 Aqua MODIS, similar accuracy was obtained (~ 90% when $P \geq 0.9$). This manual check is consistent with the OT comparison with GPM as discussed in Sect. 3.1, i.e., higher OT probability gives higher confidence in our IR algorithm for OT detection.

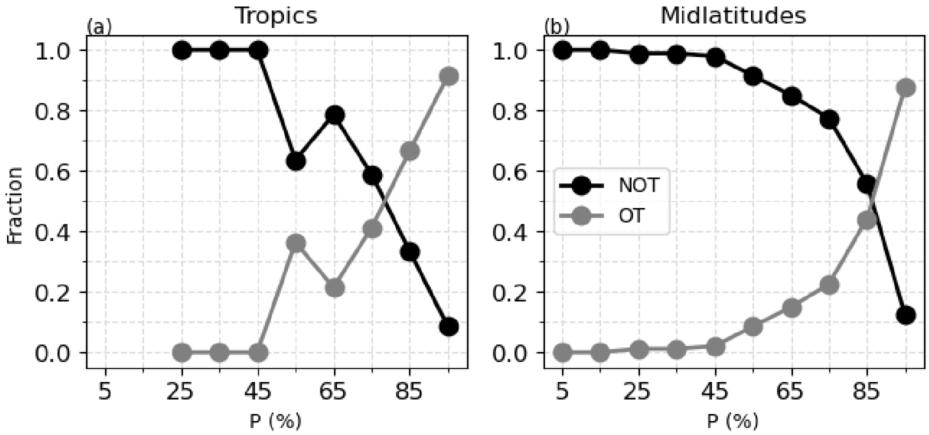


Figure 6. Fraction of OT candidates with a wide range of probability in the Tropics (a), and midlatitudes (b). X-axis shows in a probability interval of 0.1.

Overall, we choose a P threshold of 0.9 in both the tropical and mid-latitude regions, which en~~as~~sures a total detection accuracy of ~ 0.9 (better than 0.9 in the tropics and slightly lower than 0.9 in the midlatitudes) as demonstrated in Sect. 3.2. For the Terra MODIS data from February 2000 to December 2021 and Aqua MODIS data from July 2002 to December 2021, OT candidates that pass the probability threshold of 0.9 account for about 30% and 35%, respectively, of all candidates over regions within 60°S - 60°N. In the tropics, 58% (62%) of the candidates from Terra (Aqua) MODIS have P > 0.9, while in the midlatitudes, only 13% (16%) of the candidates were retained. Note that we do not consider polar regions because~~as~~ our manual selected OTs in Sect. 2.2 rarely occur outside 60° latitudes.

4. Results and Discussions

In this section, we show an OT climatology of those OT candidates with $P \geq 0.9$. Candidates with $P < 0.9$ were excluded due to a high fraction of NOTs as discussed in Sect. 3.

4.1 Case Analysis

Before showing the climatology, we first show four cases including all OT candidates with a ~~verity~~ variety of probabilities for a detailed view of the performance of our IR algorithm in different storm environments.

Figure 7 shows visible reflectance overlapped with OT centers, which are colored by OT probability. T_{b11} for each case is also shown overlapped with the pixels colder than tropopause and having positive WV-IRW BTD (marked in white). The rain type and precipitation rate averaged between 2-4 km from GPM are shown in the third and fourth columns.

Overshooting tops in tropical cyclones (TC) are common. They are found closely linked to intense convection and rapid intensification in TCs (Griffin, 2017; Monette et al., 2012; Tao and Jiang, 2013). Figures 7a1-7a4 displays a tropical cyclone over the north Indian Ocean on Nov. 08th, 2019. OTs are detected in the area with very cold T_{b11} associated with strong convection and precipitation as GPM identifies convective rain type near OT areas. Our algorithm usually generated high probability for OT candidates detected in TCs.

In the mesoscale convective system case (Figs. 7b1-b4), OTs are detected in the clusters that associate with cold T_{b11} and positive WV-IRW BTD. Strong precipitation is indicated by GPM. Our algorithm also usually produces high probability for OTs detected in mesoscale convective systems.

Cold air outbreaks can produce shallow convection when cold air blows from frozen surfaces to warmer ocean. The Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) found that these convective clouds are commonly lower than 5 km associated with updrafts of 4-5 m s^{-1} (Geerts et al., 2022). In the cold air outbreaks, the tropopause is low, which is often at a level below 500 hPa (Papritz et al., 2019; Terpstra et al., 2021), compared to the mid-latitude tropopause climatology of 200-300 hPa (Wilcox et al., 2012). Thus, updrafts in these shallow convective clouds are able to penetrate the tropopause and produce overshooting cloud tops. In the third case (Figs. 7c1-c4), overshooting tops from convective turrets over the north Atlantic within a cold air outbreak occur with high OT probability. GPM identifies convective precipitation surrounding by stratiform precipitation in these shallow convective clouds. Our method allows for the detection of these OTs that can occur in unstable conditions with shallow tropopauses.

Mid-latitude winter cyclones are associated with mostly stratiform cloud systems (Stewart et al., 1998), as also demonstrated by the GPM rain type that shows mostly stratiform precipitation (Figs. 7d1-d4). The tops of stratiform clouds associated with the fronts usually reach to tropopause without ~~no~~ strong convective cores. However, they can occur associated with lightning and heavy precipitation when fueled by potential instability, with updrafts of 6-8 m s^{-1} (Murphy et al., 2017; Rauber et al., 2014, 2015). Our algorithm detects OT candidates in this cloud system usually with low probability which will be excluded in our OT climatology analysis except for some rare situations with high OT probability.

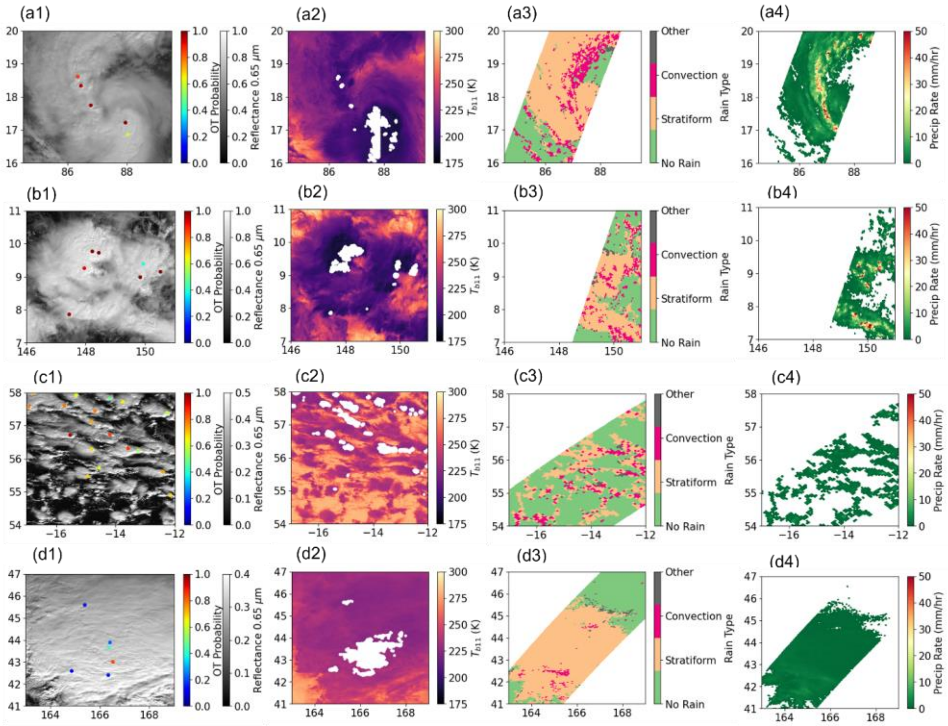


Figure 7. Four selected cloud systems with OTs detected by our IR algorithm. First column shows the reflectance at 0.65 μm (dots indicate OT probability), the second column shows the brightness temperature at 11 μm (white dots indicate pixels colder than tropopause temperature and having positive WV-IRW BTD). Columns 3 and 4 represent rain type and precipitation rate from GPM, respectively. Case 1 (a1-a4) is for the tropical cyclone over Bay of Bengal on Dec. 8th, 2019, case 2 (b1-b4) is for a mesoscale convective system over East of Philippines on Dec. 03rd, 2019, case 3 (c1-c4) (Mar. 10th, 2019 over the north Atlantic Ocean) is for shallow post-frontal convection, and case 4 (d1-d4) (Dec. 15th, 2018 over the north Pacific Ocean) is for the cloud system in the midlatitude cyclone.

4.2 Near Global OT Distributions

Figure 8 shows the seasonal distributions of OT occurrences contributed by those OT candidates with P ≥ 0.9, derived from Terra (February 2000 - 2021) and Aqua (July 2002 -2021) MODIS. As displayed, OT distributions and their seasonal variations follow the expected pattern based on the known climatology of convection (Alcala and Dessler, 2002; Funk et al., 2015). In JJA (Fig. 8b), as revealed by both Aqua and Terra MODIS, OTs primarily distribute over north of the equator in the intertropical convergence zone (ITCZ). A large population of OTs over India, Bay of Bengal, and southeast Asia are associated with the summer South Asian monsoonal system. Our algorithm also detects considerable OTs in Asia between 45°-60° latitudes and in Europe, where severe storms occur in local summer (Groenemeijer et al., 2017; Shikhov et al., 2021). These profound OTs agree with what GPM has found in the northern mid and high latitudes (Liu et al., 2020). However, a T_{b11} threshold of 215 K usually filter out these OTs (e.g. Li et al., 2022). Another hot spot of OTs occurs in central North America. In addition, we observe a narrow belt of large OT occurrences over the west Atlantic Ocean, which are associated with the location of tropical cyclones.

Aqua MODIS also shows frequent OT occurrences over the southeastern United States associated with the afternoon convection. In regions over the U.S. southwest and northwestern Mexico, OTs are detected associated with the summer North America Monsoon (Adams and Comrie, 1997).

During DJF (Fig. 8d), OT occurrences are about 44% at 10:30 LT (Terra equator crossing time) and 36% at 1:30 LT (Aqua equator crossing time) ((Nsummer-Nwinter)/ Nsummer) less than that in JJA. OTs are primarily located over the Southern Hemisphere as the ITCZ moves to the south of the equator. A large number of OTs are detected by Aqua MODIS over tropical and subtropical South America and Africa. In the Northern Hemisphere, OTs become infrequent over land. Note that ice clouds have an occurrence frequency about 70% over mid- and high-latitude Asia during winter (e.g. Hong and Liu 2015), which often pose challenges for OT identification. These cold ice clouds are rarely classified as OTs in our analysis, demonstrating the ability of our IR method to avoid the misclassification of cold ice clouds to OTs. In contrast, over the mid-latitude ocean in winter, we see some OT occurrences. These OTs are associated with isolated convective clouds occurring in the cold air outbreaks as discussed in Sect. 4.1. These OTs are also observed over Southern Ocean during JJA (Austral winter). We also notice a small number of OTs extending from northwest to southeast North America in DJF. These OTs are associated with the convection in winter mid-latitude cyclones as discussed in Sect. 4.1.

Convective activity over land is weak at Terra overpass time in the morning (~ 10:30 am) and it becomes more frequent and intense in the afternoon when Aqua satellite overpasses. This is revealed by the differences of OT occurrences between Terra and Aqua, indicating the variability of OT diurnal cycles.

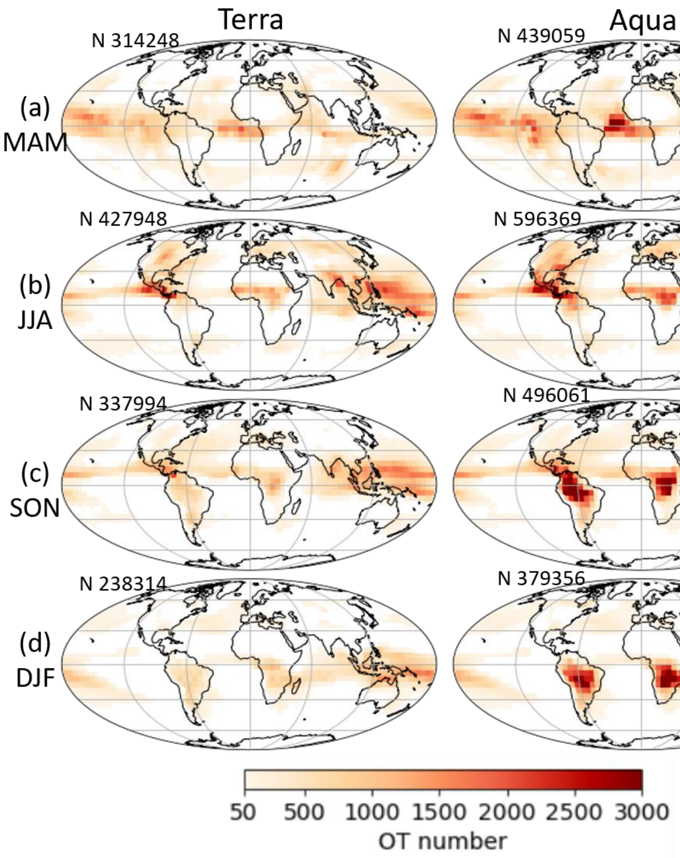


Figure 8. The global distributions of OT occurrences derived from Terra and Aqua MODIS in four seasons: (a) March-April-May (MAM), (b) June-July-August (JJA), (c) September-October-November (SON) and (d) December-January-February (DJF). Grid resolution is 5° longitude by 5° latitude. Samples in grids less than 50 are shown in white. N over the upper right corner in each panel stands for sample number.,

4.3 OT Diurnal Cycle

This section discusses OT diurnal cycles based on the four observation times by Aqua and Terra MODIS. The OT occurrences in the daytime (~10:30 am and ~1:30 pm) and at night (~10:30 pm and ~1:30 am) are displayed in Fig. 9. According to previous studies on the diurnal cycle of convection (Alcala and Dessler, 2002; Nesbitt and Zipser, 2003), convective activity over land is generally more frequent and intense in the afternoon and evening compared with early morning. Over oceans near the coastlines, morning convection is more intense (Johnson, 2011). In agreement with previous studies, we observe the most OT occurrences at about 1:30 pm from Aqua MODIS, primarily contributed by land areas including tropical South America, tropical Africa, the Maritime continent and the southern foothills of Himalayas. Over Bay of Bengal, South China Sea, Gulf of Guinea, Gulf of Mexico, Panama and its surrounding regions, OTs away from coastlines have been observed, commencing in the morning (~ 10:30 am) and continuing into afternoon (~ 1:30 pm). Over the west Pacific Ocean, OTs occur the most around midnight at ~ 1:30 am.

To better view the OT diurnal cycles, Figure 10 shows when maximum and minimum OT occurrences occur in the four-observation time. Diurnal cycle intensity defined by the difference of maximum and minimum OT numbers normalized by the mean is shown in Figs.10 e and f. As expected (Figs. 10a and 10b), the largest OT occurrences over land occur at about ~ 1:30 pm except for central North America and west Africa where have a midnight maximum in convection during JJA (Janiga and Thorncroft, 2014; Nesbitt and Zipser, 2003; Tian et al., 2005). Ocean areas consistently have maximum OT occurrence at ~ 1:30 am (Figs.10a and 10b). The minimum OT occurrence over land usually occurs at ~ 10:30 am except for some regions over North America and Asia where the minimum OT occurrence is at ~1:30 am during JJA (Fig. 10c). The time for minimum OT occurrence over ocean has a large variability.

The diurnal cycles of OT occurrences over ocean are generally weak (Figs. 10e and 10f), being consistent with previous convection diurnal cycle analysis (Alcala & Dessler, 2002; Liu & Zipser, 2005; Nesbitt & Zipser, 2003). In contrast, the OT diurnal cycles over land are much stronger than over ocean. Strong regional variations are also discovered over land areas. Relatively strong OT diurnal cycles are found during JJA over southwest North America, southeast United States, Tibetan High, tropical South America, and during DJF over southeast Australia, tropical and subtropical South America and subtropical Africa. Relatively weak diurnal cycles over land are observed in central North America and west Africa in JJA. Strong regional variations in OT diurnal cycle over land are consistent with previous studies based on convection and precipitation that demonstrate the diurnal cycles are complicatedly modulated by land-sea contrast, topography, coastline curvature and response to solar heating to surface (Janiga and Thorncroft, 2014; Tian et al., 2005).

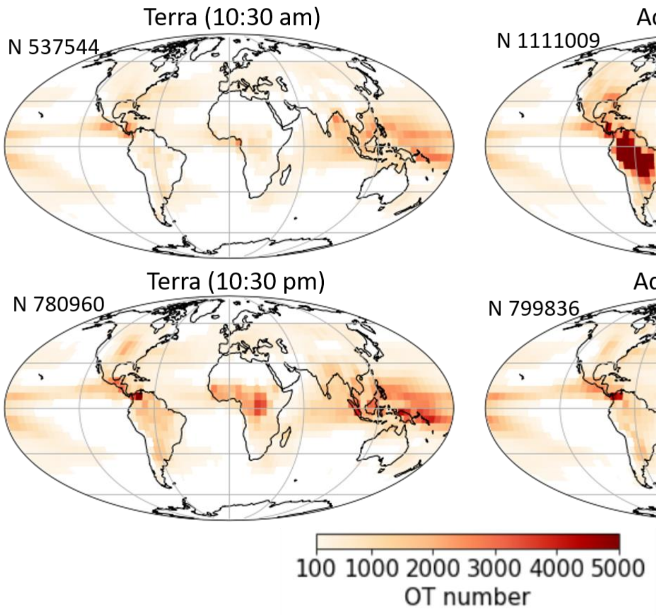


Figure 9. The global distributions of OTs at four observation times. Grids with OT number < 100 are shown in white. N stands for sample number.

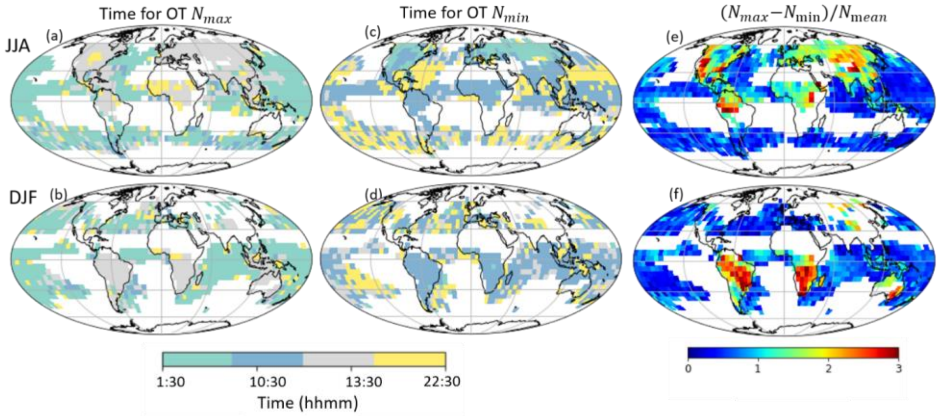


Figure 10. Panels a-d are for the time when maximum and minimum OT occurrence occurs across the four-observation time. Panels e-f are for diurnal intensity of OT occurrences, defined as the difference of maximum and minimum OT occurrences, normalized by the mean. Only when the minimum OT occurrences > 10 in each 5°x5° grid, data is shown.

4.4 Land-Sea Contrast

From the diurnal cycle analysis in Sect. 4.3, we have noticed some land-sea contrast in OT characteristics. For instance, OTs occur more frequently in the afternoon over land, whereas they are more frequent at midnight over ocean, and OT occurrence diurnal cycle is stronger over land than over ocean. In this section, attention is placed on OT center T_{b11}, which indicates storm intensity. By checking the geospatial distributions of OT center T_{b11}, we observe extremely cold OT center T_{b11} (e.g. < 180 K) appearing over the tropical regions, including regions near northern Australia, east of Papua New Guinea, India and nearby Arabian sea, tropical and subtropical Africa, and tropical and subtropical South America, derived from both Aqua and Terra MODIS (Figs. 11a and 11b). The locations of cold OTs are also aligned with the places where occur intense convection based on TRMM (Zipser et al., 2006).

Particularly, the first 10 coldest OTs (marked in red triangles and summarized in Table 2) from Aqua and Terra MODIS nearly occur in Sothern Hemisphere with more cases over land than over ocean. The top 10 coldest OTs from Aqua are colder than 167 K with the coldest OT of 165.6 K over east of Papua New Guinea, whereas Terra shows the coldest OT of 167.2 K occurring in northern Australia. This finding agrees with the cold OT distributions discussed in Proud & Bachmeier, 2021, which states that an extremely cold tropopause coupled to an energetic overshooting top produced such a cloud top temperature.

Additionally, Figs. 11a and 11b reveal colder OTs over land than over ocean at the same latitudes. By checking the probability density distributions (PDFs) of OT center T_{b11}, we find that land-sea contrast in OT T_{b11} also relies on diurnal cycle. In the daytime morning (~ 10:30 am) when convection over land is weak, T_{b11} over land is slightly warmer than over ocean in both the tropics and midlatitudes (Figs. 11c and 11e). Land-sea contrast in T_{b11} is small at this time. At ~1:30 pm as convection becomes stronger over land, T_{b11} over land is on average 0.8 K and 2.3 K colder than over ocean in the tropics and midlatitudes, respectively (Figs. 11c and 11e). At nighttime (Figs. 11d and 11f), land-sea contrast in T_{b11} becomes stronger than in the daytime. In the tropics, T_{b11} over land is about 1 K on average colder than that over ocean, whereas in the midlatitudes, it is about 2 K colder over land than over ocean.

Our findings indicate that OTs over land are more intense than over ocean except for the early morning (~ 10:30 am) when convection over land is weak. These findings agree with previous studies that have shown more intense convection over land area, associated with stronger updrafts than the oceanic counterpart (Jeyaratnam et al., 2021; Liu & Zipser, 2005).

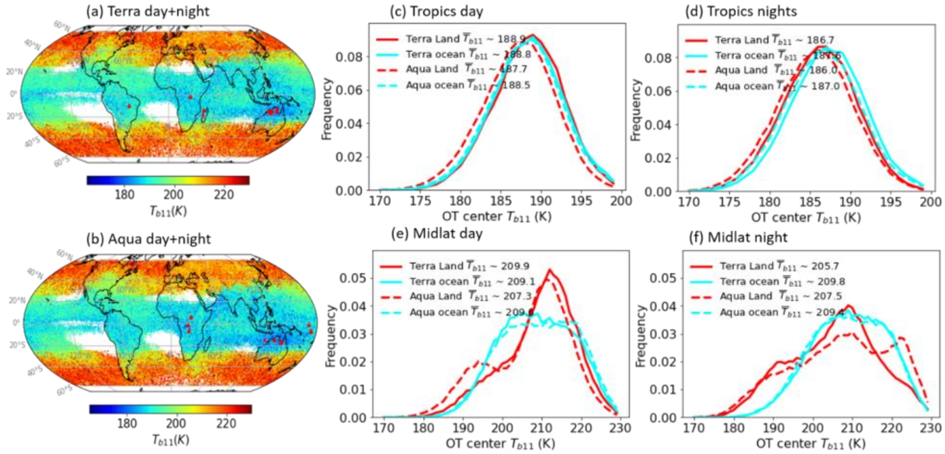


Figure 11. a and b are for spatial distributions of OT center T_{b11}. Red triangles indicate the locations of top ten coldest OTs. Panels c-f are for OT center T_{b11} PDFs in the tropics and midlatitudes, segregated in day and nighttime.

Table 2. Summary of the top 10 coldest OTs from Terra and Aqua, respectively.

	Terra				Aqua			
	$T_{b11}(K)$	Location(lon,lat)	Time[1]	D/N[2]	$T_{b11}(K)$	Location(lon,lat)	Time	D/N
1	167.2	129.75,-14.54	2016365.1350	N	165.6	172.65,-6.66	2018365.1415	N
2	167.6	128.62,-15.76	2006023.1410	N	166.4	169.81,-0.99	2018365.1410	N
3	167.6	125.47,-14.62	2014004.0205	D	166.5	21.34,-0.74	2015082.0030	N
4	167.8	27.28,-1.46	2013023.2020	N	166.6	22.29,-6.15	2020053.2355	N
5	167.8	-48.54,-9.98	2013029.0130	N	166.7	172.54,-7.35	2018365.1415	N
6	167.8	136.12,-14.72	2019003.1320	N	166.7	138.75,-17.12	2018332.0420	D
7	167.8	44.58,-19.05	2006074.1935	N	166.8	138.71,-15.78	2012080.1625	N
8	168.1	44.63,-13.74	2004023.1925	N	166.8	129.75,-14.64	2003020.0500	D
9	168.3	135.39,12.79	2008006.1405	N	166.9	118.45,-14.93	2016359.0600	D
10	168.3	128.49,-16.31	2018001.1405	N	166.9	24.93,5.63	2008123.1140	D

1.Time in the format of year.day.hhmm

2.D for day and N for night

5. Conclusions

To utilize about two-decade MODIS records in study of convective overshooting tops, we developed an IR algorithm to detect OTs from MODIS. The resultant OT climatology was used to understand OT regional and seasonal distributions, OT diurnal cycles and OT land-sea contrast.

The approach to detect OTs uses IR radiances from MODIS water vapor (6.7 μm) and window (11 μm) channels. This approach was built upon the logistic regression which was trained and validated with ~ 287 OT candidates identified from the combined CloudSat-CALIPSO-MODIS (CCM) data. As demonstrated by six-year collocated GPM observations, the OT probability generated by the IR algorithm indicates storm intensity and represents a confidence level of OT detection. When OT probability is higher than 0.9, the accuracy for OT detection is better than about 0.9 as validated by manual check.

The global and seasonal distributions of OT occurrences follow the expected pattern based on the known climatology of deep convection and precipitation, shifting with the ITCZ and monsoonal systems. Frequent OTs are also observed over central North American, Europe, northern Asia and the northwest Atlantic Ocean in summer. Our OT climatology also includes those OTs observed in the shallow convection over the mid-latitude ocean during spring-winter cold air outbreaks.

MODIS observations at four different times were used to derive part of the OT diurnal cycle. The diurnal cycle follows the known diurnal cycle of convection: The most OT occurrences are observed at about 1:30 pm (ECT) over most land area, including tropical and subtropical South America, tropical and subtropical Africa, southeast North America, foot of Himalayas and Maritime continent, etc. Over ocean, maximum OT occurrences are usually at around midnight (~1:30 am) except for offshore ocean. OT occurrences in the morning (~ 10:30 am) over coastal ocean are apparent which continue to the afternoon at ~ 1:30 pm. Minimum OT occurrences are usually at ~10:30 am over land. Over ocean, however, minimum occurrences can be at any time except 1:30 am. Also, the OT diurnal cycle is stronger and more varied over land than over ocean.

Jeyaratnam et al., (2021) indicated that tropical convection is deeper than mid-latitude convection. This is also revealed by the midlatitude-tropics contrast in OT center T_{b11} shown in this study, i.e. tropical OTs are colder than mid-latitude OTs. In the tropics, the ~~the~~ OT center T_{b11} tends to be colder over land than over ocean ex~~a~~c~~e~~ept at ~ 10:30 am when convection over land is weak. Also, the top 10 coldest OTs from either Terra or Aqua mostly occur over land. These results agree with previous studies that have confirmed that tropical land areas exhibit more intense overshooting convection than the tropical oceans (Alcala & Dessler, 2002; Liu & Zipser, 2005). Mid-latitude OTs have stronger land-sea contrast in T_{b11} than in the tropics with OTs over land being 2.3, 4.1 and 1.9 K colder than that over ocean at about 1:30 pm, 10:30 pm and 1:30 am, respectively.

This study has displayed a comprehensive analysis of OT occurrences over near the globe for the ~~first time~~first -time using MODIS data. As MODIS~~that~~ has a fine~~better~~ spatial resolution

(1 km) and provides about two-decade stable climate records~~covers the longest time period than previous OT climatologies that were derived from either GPM, GOES or AMSU-B~~, results in this study are an important complementary to the current OT climatology in literature derived from GPM, GOES and AMSU-B. This study also lays a foundation to understand near global climatological distributions of hazardous thunderstorms,~~ Our ongoing work seeks to use this OT climatology to quantify OT aera, which will~~ leading to valuable insights into intense updraft size distributions in deep convection over the globe.

Data availability

CloudSat data including 2B-GEOPROF, 2B-CLDLASS-LIDAR and ECMWF-AUX, were downloaded from https://www.cloudsat.cira.colostate.edu/.

GPM radar data is available at https://disc.gsfc.nasa.gov/datasets/GPM_2ADPR_07/summary.

MODIS data is available at https://ladsweb.modaps.eosdis.nasa.gov/.

MERRA-2 data can be downloaded at https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/M2I1NXASM.5.12.4/.

The dataset used for training and cross-validating Logistic Regression is available in Supplementary 2.

The dataset used for manually validating Logistic Regression is available in Supplementary 3.

Author contribution

YH, JT, SN and LDL conceived this study. YH performed the analysis, collected data, and wrote the manuscript. SN collected data, helped with data analysis and edited the manuscript. JT helped with interpretation of results and edited the manuscript. LD joined result discussions and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was mainly supported by the NASA award 80NSSC20K0902. The authors would like to acknowledge Dr. Guangyu Zhao for his help in downloading the Terra MODIS data. We thank the CloudSat Data Processing Center for providing CloudSat products, including 2B-GEOPROF, 2B-CLDCLASS-LIDAR and ECMWF-AUX. We thank the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) for offering MODIS data (LAADS DAAC, 2022). We also acknowledge the NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC) for archiving MERRA-2 data and GPM data (Iguchi and Meneghini 2021).

References

Ackerman, S. A.: Global satellite observations of negative brightness temperature differences between 11 and 6.7 µm, J. Atmos. Sci., 53(19), 2803–2812, doi:10.1175/1520-0469(1996)053<2803:GSOONB>2.0.CO;2, 1996.

Adams, D. K. and Comrie, A. C.: The North American Monsoon, Bull. Am. Meteorol. Soc., 78(10), 2197–2213, doi:10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2, 1997.

Alcala, C. M. and Dessler, A. E.: Observations of deep convection in the tropics using the Tropical Rainfall Measuring Mission (TRMM) precipitation radar, J. Geophys. Res. Atmos., 107(24), doi:10.1029/2002JD002457, 2002.

Astin, I., Di Girolamo, L. and Van De Poll, H. M.: Bayesian confidence intervals for true fractional coverage from finite transect measurements: Implications for cloud studies from space, J. Geophys. Res. Atmos., 106(D15), 17303–17310, doi:10.1029/2001JD900168, 2001.

Barnes, W. L., Pagano, T. S. and Salomonson, V. V.: Prelaunch characteristics of the moderate resolution, IEEE Trans. Geosci. Remote Sens., 36, 1088–1100, 1998.

Bedka, K., Brunner, J., Dworak, R., Feltz, W., Otkin, J. and Greenwald, T.: Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients, J. Appl. Meteorol. Climatol., 49(2), 181–202, doi:10.1175/2009JAMC2286.1, 2010.

Bedka, K. M.: Overshooting cloud top detections using MSG SEVIRI Infrared brightness temperatures and their relationship to severe weather over Europe, Atmos. Res., 99(2), 175–189, doi:10.1016/j.atmosres.2010.10.001, 2011.

Bedka, K. M. and Khlopenkov, K.: A probabilistic multispectral pattern recognition method for detection of overshooting cloud tops using passive satellite imager observations, J. Appl. Meteorol. Climatol., 55(9), 1983–2005, doi:10.1175/JAMC-D-15-0249.1, 2016.

Bedka, K. M., Allen, J. T., Punge, H. J., Kunz, M. and Simanovic, D.: A long-term overshooting convective cloud-top detection database over Australia derived from MTSAT Japanese Advanced Meteorological Imager Observations, J. Appl. Meteorol. Climatol., 57(4), 937–951, doi:10.1175/JAMC-D-17-0056.1, 2018.

Bosilovich, M. G., Lucchesi, R. and Suarez, M.: MERRA-2: File Specification, Earth, 9(9), 73 [online] Available from: http://gmao.gsfc.nasa.gov/pubs/office_notes., 2016.

Chung, E. S., Sohn, B. J., Schmetz, J. and Koening, M.: Diurnal variation of upper tropospheric humidity and its relations to convective activities over tropical Africa, Atmos. Chem. Phys., 2489–2502, doi:www.atmos-chem-phys.net/7/2489/2007/, 2007.

Chung, E. S., Sohn, B. J. and Schmetz, J.: CloudSat shedding new light on high-reaching tropical deep convection observed with Meteosat, Geophys. Res. Lett., 35(2), 1–5, doi:10.1029/2007GL032516, 2008.

Dworak, R., Bedka, K., Brunner, J. and Feltz, W.: Comparison between GOES-12 overshooting-top detections, WSR-88D radar reflectivity, and severe storm reports, Weather Forecast., 27(3), 684–699, doi:10.1175/WAF-D-11-00070.1, 2012.

Funk, C., Verdin, A., Michaelsen, J., Peterson, P., Pedreros, D. and Husak, G.: A global satellite-assisted precipitation climatology, Earth Syst. Sci. Data, 7(2), 275–287, doi:10.5194/essd-7-275-2015, 2015.

Geerts, B., Giangrande, S. E., McFarquhar, G. M., Xue, L., Abel, S. J., Comstock, J. M., Crewell, S., DeMott, P. J., Ebell, K., Field, P., Hill, T. C. J., Hunzinger, A., Jensen, M. P., Johnson, K. L., Juliano, T. W., Kollias, P., Kosovic, B., Lackner, C., Luke, E., Lüpkes, C., Matthews, A. A., Neggers, R., Ovchinnikov, M., Powers, H., Shupe, M. D., Spengler, T., Swanson, B. E., Tjernström, M., Theisen, A. K., Wales, N. A., Wang, Y., Wendisch, M. and Wu, P.: The COMBLE Campaign: A Study of Marine Boundary Layer Clouds in Arctic Cold-Air Outbreaks, Bull. Am. Meteorol. Soc., 103(5), E1371–E1389, doi:10.1175/bams-d-21-0044.1, 2022.

Gettelman, A., Salby, M. L. and Sassi, F.: Distribution and influence of convection in the tropical tropopause region, J. Geophys. Res. Atmos., 107(9–10), doi:10.1029/2001jd001048, 2002.

Gettelman, A., Forster, P. M. de F., Fujiwara, M., Fu, Q., Vo ̈mel, H., Gohar, L. K., Johanson, C. and Ammerman, M.: Radiation balance of the tropical tropopause layer, J. Geophys. Res., 109(D7), D07103, doi:10.1029/2003JD004190, 2004.

Griffin, S. M.: Climatology of tropical overshooting tops in North Atlantic tropical cyclones, J. Appl. Meteorol. Climatol., 56(6), 1783–1796, doi:10.1175/JAMC-D-16-0413.1, 2017.

Griffin, S. M., Bedka, K. M. and Velden, C. S.: A method for calculating the height of overshooting convective cloud tops using satellite-based IR imager and CloudSat cloud profiling radar observations, J. Appl. Meteorol. Climatol., 55(2), 479–491, doi:10.1175/JAMC-D-15-0170.1, 2016.

Grise, K. M., Thompson, D. W. J. and Birner, T.: A global survey of static stability in the stratosphere and upper troposphere, J. Clim., 23, 2275–2292, doi:10.1175/2009JCLI3369.1, 2010.

Groenemeijer, P., Púcik, T., Holzer, A. M., Antonescu, B., Riemann-Campe, K., Schultz, D. M., Kühne, T., Feuerstein, B., Brooks, H. E., Doswell, C. A., Koppert, H. J. and Sausen, R.: Severe convective storms in Europe: Ten years of research and education at the European Severe Storms Laboratory, Bull. Am. Meteorol. Soc., 98(12), 2641–2651, doi:10.1175/BAMS-D-16-0067.1, 2017.

Heymsfield, G. M., Tian, L., Heymsfield, A. J., Li, L. and Guimond, S.: Characteristics of deep tropical and subtropical convection from nadir-viewing high-altitude airborne doppler radar, J. Atmos. Sci., 67(2), 285–308, doi:10.1175/2009JAS3132.1, 2010.

Hong, G., Heygster, G., Miao, J. and Kunzi, K.: Detection of tropical deep convective clouds from AMSU-B water vapor channels measurements, J. Geophys. Res. D Atmos., 110(5), 1–15, doi:10.1029/2004JD004949, 2005.

Hong, G., Heygster, G., Notholt, J. and Buehler, S. A.: Interannual to diurnal variations in tropical and subtropical deep convective clouds and convective overshooting from seven years of AMSU-B measurements, J. Clim., 21(17), 4168–4189, doi:10.1175/2008JCLI1911.1, 2008.

Hong, Y. and Di Girolamo, L.: Cloud Phase Characteristics Over Southeast Asia from A-Train Satellite Observations, Atmos. Chem. Phys., (20), 8267–8291, doi:https://doi.org/10.5194/acp-20-8267-2020, 2020.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T.: The global precipitation measurement mission, Bull. Am. Meteorol. Soc., 95(5), 701–722, doi:10.1175/BAMS-D-13-00164.1, 2014.

Hourngir, D., Panegrossi, G., Casella, D., Sanò, P., D'adderio, L. P. and Liu, C.: A 4-year climatological analysis based on gpm observations of deep convective events in the mediterranean region, Remote Sens., 13(9), 1–21, doi:10.3390/rs13091685, 2021.

Janiga, M. A. and Thorncroft, C. D.: Convection over tropical Africa and the East Atlantic during the West African monsoon: Regional and diurnal variability, J. Clim., 27(11), 4189–4208, doi:10.1175/JCLI-D-13-00449.1, 2014.

Jeyaratnam, J., Luo, Z. J., Giangrande, S. E., Wang, D. and Masunaga, H.: A Satellite-Based Estimate of Convective Vertical Velocity and Convective Mass Flux: Global Survey and Comparison With Radar Wind Profiler Observations, Geophys. Res. Lett., 48(1), 1–11, doi:10.1029/2020GL090675, 2021.

Johnson, R. H.: Diurnal cycle of monsoon convection, in The global monsoon system: Research and forecast, pp. 257–276, Singapore: World Scientific., 2011.

Khlopenkov, K. V., Bedka, K. M., Cooney, J. W. and Itterly, K.: Recent Advances in Detection of Overshooting Cloud Tops From Longwave Infrared Satellite Imagery, J. Geophys. Res. Atmos., 126(14), 1–25, doi:10.1029/2020jd034359, 2021.

King, M. D., Kaufman, Y. J., Menzel, W. P. and Tanré, D.: Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS), IEEE Trans. Geosci. Remote Sens., 30, 2–27, doi:10.1109/36.124212, 1992.

Li, H., Wei, X., Min, M., Li, B., Nong, Z. and Chen, L.: A Dataset of Overshooting Cloud Top from 12-Year CloudSat/CALIOP Joint Observations, Remote Sens., 14(10), 2417, doi:10.3390/rs14102417, 2022.

Liu, C. and Zipser, E. J.: Global distribution of convection penetrating the tropical tropopause, J. Geophys. Res. Atmos., 110(23), 1–12, doi:10.1029/2005JD006063, 2005.

Liu, N. and Liu, C.: Global distribution of deep convection reaching tropopause in 1 year GPM observations, NJournal Geophys. Res. Atmosperes, 121, 3924–3842, doi:10.1002/ 2015JD024430, 2016.

Liu, N., Liu, C. and Hayden, L.: Climatology and Detection of Overshooting Convection From 4 Years of GPM Precipitation Radar and Passive Microwave Observations, J. Geophys. Res. Atmos., 125(7), 1–14, doi:10.1029/2019JD032003, 2020.

Marchand, R., Mace, G. G., Ackerman, T. and Stephens, G.: Hydrometeor detection using Cloudsat - An earth-orbiting 94-GHz cloud radar, J. Atmos. Ocean. Technol., 25(4), 519–533, doi:10.1175/2007JTECHA1006.1, 2008.

Marion, G. R., Trapp, R. J. and Nesbitt, S. W.: Using overshooting top area to discriminate potential for large, intense tornadoes, Geophys. Res. Lett., 46(21), 12520–12526, doi:10.1029/2019GL084099, 2019.

Monette, S. A., Velden, C. S., Griffin, K. S. and Rozoff, C. M.: Examining trends in satellite-detected tropical overshooting tops as a potential predictor of tropical cyclone rapid intensification, J. Appl. Meteorol. Climatol., 51(11), 1917–1930, doi:10.1175/JAMC-D-11-0230.1, 2012.

Murphy, A. M., Rauber, R. M., McFarquhar, G. M., Finlon, J. A., Plummer, D. M., Rosenow, A. A. and Jewett, B. F.: A microphysical analysis of elevated convection in the comma head region of continental winter cyclones, J. Atmos. Sci., 74(1), 69–91, doi:10.1175/JAS-D-16-0204.1, 2017.

Nesbitt, S. W. and Zipser, E. J.: The diurnal cycle of rainfall and convective intensity according to three years of TRMM measurements, J. Clim., 16, 1456–1475, doi:10.1175/1520-0442-16.10.1456, 2003.

Papritz, L., Rouges, E., Aemisegger, F. and Wernli, H.: On the Thermodynamic Preconditioning of Arctic Air Masses and the Role of Tropopause Polar Vortices for Cold Air Outbreaks From Fram Strait, J. Geophys. Res. Atmos., 124(21), 11033–11050, doi:10.1029/2019JD030570, 2019.

Partain, P.: Cloudsat ECMWF-AUX auxiliary data process description and interface control document, Coop. Inst. Res. Atmos. Color. State Univ. [online] Available from: http://129.82.109.192/ICD/ECMWF-AUX/ECMWF-AUX_PDICD_3.0.pdf, 2007.

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C. and Frey, R. A.: The MODIS cloud products : Algorithms and examples from Terra, IEEE Trans. Geosci. Remote Sens., 41, 459–473, doi:10.1109/TGRS.2002.808301, 2003.

Proud, S. R.: Analysis of overshooting top detections by Meteosat Second Generation: A 5-year dataset, Q. J. R. Meteorol. Soc., 141(688), 909–915, doi:10.1002/qj.2410, 2015.

Rauber, R. M., Wegman, J., Plummer, D. M., Rosenow, A. A., Peterson, M., McFarquhar, G. M., Jewett, B. F., Leon, D., Market, P. S., Knupp, K. R., Keeler, J. M. and Battaglia, S. M.: Stability and charging characteristics of the comma head region of continental winter cyclones, J. Atmos. Sci., 71(5), 1559–1582, doi:10.1175/JAS-D-13-0253.1, 2014.

Rauber, R. M., Plummer, D. M., Macomber, M. K., Rosenow, A. A., McFarquhar, G. M., Jewett, B. F., Leon, D., Owens, N. and Keeler, J. M.: The role of cloud-top generating cells and boundary layer circulations in the finescale radar structure of a winter cyclone over the great lakes, Mon. Weather Rev., 143(6), 2291–2318, doi:10.1175/MWR-D-14-00350.1, 2015.

Rysman, J. F., Claud, C. and Delanoe, J.: Monitoring Deep Convection and Convective Overshooting from 60° S to 60° N Using MHS: A Cloudsat/CALIPSO-Based Assessment, IEEE Geosci. Remote Sens. Lett., 14(2), 159–163, doi:10.1109/LGRS.2016.2631725, 2017.

Schmetz, J., Tjemkes, S. A., Gube, M. and Van De Berg, L.: Monitoring deep convection and convective overshooting with METEOSAT, Adv. Sp. Res., 19(3), 433–441, doi:10.1016/S0273-1177(97)00051-3, 1997.

Setvák, M., Rabin, R. M. and Wang, P. K.: Contribution of the MODIS instrument to observations of deep convective storms and stratospheric moisture detection in GOES and MSG imagery, Atmos. Res., 83(2-4 SPEC. ISS.), 505–518, doi:10.1016/j.atmosres.2005.09.015, 2007.

Setvák, M., Bedka, K., Lindsey, D. T., Sokol, A., Charvát, Z., Šťástka, J. and Wang, P. K.: A-Train observations of deep convective storm tops, Atmos. Res., 123, 229–248, doi:10.1016/j.atmosres.2012.06.020, 2013.

Shikhov, A., Chernokulsky, A., Kalinin, N., Bykov, A. and Pischalnikova, E.: Climatology and Formation Environments of Severe Convective Windstorms and Tornadoes in the Perm Region (Russia) in 1984-2020, Atmosphere (Basel)., (12), 1407, doi:https://doi.org/ 10.3390/atmos12111407, 2021.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'Connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A. and Mitrescu, C.: The cloudsat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation, Bull. Am. Meteorol. Soc., 83, 1771–1790, doi:10.1175/BAMS-83-12-1771, 2002.

Stephens, G. L., Vane, D. G., Tanelli, S., Im, E., Durden, S., Rokey, M., Reinke, D., Partain, P., Mace, G. G., Austin, R., L'Ecuyer, T., Haynes, J., Lebsock, M., Suzuki, K., Waliser, D., Wu, D., Kay, J., Gettelman, A., Wang, Z. and Marchand, R.: CloudSat mission: Performance and early science after the first year of operation, J. Geophys. Res., 113, D00A18, doi:10.1029/2008JD009982, 2008.

Stewart, R. E., Szeto, K. K., Reinking, R. F., Clough, S. A. and Ballard, S. P.: Midlatitude cyclonic cloud systems and their features affecting large scales and climate, Rev. Geophys., 36(2), 245–273, doi:10.1029/97RG03573, 1998.

Sun, L. X., Zhuge, X. Y. and Wang, Y.: A Contour-Based Algorithm for Automated Detection of Overshooting Tops Using Satellite Infrared Imagery, IEEE Trans. Geosci. Remote Sens., 57(1), 497–508, doi:10.1109/TGRS.2018.2857486, 2019.

Tao, C. and Jiang, H.: Global distribution of hot towers in tropical cyclones based on 11-Yr TRMM data, J. Clim., 26(4), 1371–1386, doi:10.1175/JCLI-D-12-00291.1, 2013.

Terpstra, A., Renfrew, I. A. and Sergeev, D. E.: Characteristics of cold-air outbreak events and associated polar mesoscale cyclogenesis over the north Atlantic region, J. Clim., 34(11), 4567–4584, doi:10.1175/JCLI-D-20-0595.1, 2021.

Tian, B., Soden, B. J. and Wu, X.: Diurnal cycle of convection, clouds, and water vapor in the tropical upper troposphere: Satellites versus a general circulation model, J. Geophys. Res. D Atmos., 109(10), 1–16, doi:10.1029/2003JD004117, 2004.

Tian, B., Held, I. M., Lau, N. C. and Soden, B. J.: Diurnal cycle of summertime deep convection over North America: A satellite perspective, J. Geophys. Res. D Atmos., 110(8), 1–10, doi:10.1029/2004JD005275, 2005.

Trapp, R. J., Marion, G. R. and Nesbitt, S. W.: The regulation of tornado intensity by updraft width, J. Atmos. Sci., 74(12), 4199–4211, doi:10.1175/JAS-D-16-0331.1, 2017.

Vergados, P., Luo, Z. J., Emanuel, K. and Mannucci, A. J.: Observational tests of hurricane intensity estimations using GPS radio occultations, J. Geophys. Res. Atmos., 119, 1936–1948, doi:10.1002/2013JD020934, 2014.

Wang, C., Luo, Z. J. and Huang, X.: Parallax correction in collocating CloudSat and Moderate Resolution Imaging Spectroradiometer (MODIS) observations: Method and application to convection study, J. Geophys. Res. Atmos., 116(17), 1–9, doi:10.1029/2011JD016097, 2011.

Wang, Z., Vane, D. and Staphens, G.: Level 2 Combined Radar and Lidar Cloud Scenario Classification Product Process Description and Interface Control document. [online] Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Level+2+Combined+Radar+and+Lidar+Cloud+Scenario+Classification+Product+Process+Description+and+Interface+Control+Document#1, 2012.

Wilcox, L. J., Hoskins, B. J. and Shine, K. P.: A global blended tropopause based on ERA data. Part I: Climatology, Q. J. R. Meteorol. Soc., 138(664), 561–575, doi:10.1002/qj.951, 2012.

Winker, D. M., Pelon, J. R. and McCormick, M. P.: The CALIPSO mission: Spaceborne lidar for observation of aerosols and clouds, Lidar Remote Sens. Ind. Environ. Monit. III, 4893, 1–11, doi:10.1117/12.466539, 2003.

Xiong, X., Sun, J., Wu, A., Chiang, K.-F., Esposito, J. and Barnes, W.: Terra and Aqua MODIS calibration algorithms and uncertainty analysis, Sensors, Syst. Next-Generation Satell. IX, 5978(November 2015), 59780V, doi:10.1117/12.627631, 2005.

Xiong, X., Angal, A., Barnes, W. L., Chen, H., Chiang, V., Geng, X., Li, Y., Twedt, K., Wang, Z., Wilson, T. and Wu, A.: Updates of Moderate Resolution Imaging Spectroradiometer on-orbit calibration uncertainty assessments, J. Appl. Remote Sens., 12(03), 1, doi:10.1117/1.jrs.12.034001, 2018.

Yuter, S. E. and Houze, R. A.: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and differential reflectivity, Mon. Weather Rev., 123, doi:https://doi.org/10.1175/1520-0493(1995)123<1941:TDKAME>2.0.CO;2, 1995.

Zhuge, X. Y., Ming, J. and Wang, Y.: Reassessing the use of inner-core hot towers to predict tropical cyclone rapid intensification, Weather Forecast., 30(5), 1265–1279, doi:10.1175/WAF-D-15-0024.1, 2015.

Zipser, E. J., Cecil, D. J., Liu, C., Nesbitt, S. W. and Yorty, D. P.: Where are the most intense thunderstorms on Earth?, Bull. Am. Meteorol. Soc., 87(8), 1057–1071, doi:10.1175/BAMS-87-8-1057, 2006.