Comparing Tropopause-Penetrating Convection Identifications Derived From NEXRAD and GOES Over the Contiguous United States

John W. Cooney1,2, Christopher M. Bedka2, Kenneth P. Bowman3, Konstantin V. Khlopenkov4, and Kyle Itterly4

1NASA Postdoctoral Fellowship Program at NASA Langley Research Center, Hampton, VA, USA, 2NASA Langley Research Center, Hampton, VA, USA, 3Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA, 4Science Systems and Applications Inc., Hampton, VA, USA

Abstract
Overshooting tops (OTs) are a well-known indicator of updrafts capable of transporting air from the troposphere to the stratosphere and generating hazardous weather conditions. Satellites and radars have long been used to identify OTs, but the results have not been entirely consistent due to differences in sensor and measurement characteristics. OT detection approaches based on satellite infrared (IR) imagery have often been validated using human-expert OT identifications, but such datasets are time-consuming to compile over broad geographic regions. Despite radar limitations to detect the true physical cloud top, OTs identified within multi-radar composites can serve as a stable reference for comprehensive satellite OT analysis and detection validation. This study analyzes a large OT data set compiled from Geostationary Operational Environmental Satellites (GOES)-13/16 geostationary IR data and gridded volumetric Next-Generation Radar (NEXRAD) reflectivity to better understand radar and IR observations of OTs, quantify agreement between satellite and radar OT detections, and demonstrate how an increased spatial sampling from GOES-13 to GOES-16 impacts OT appearance and detection performance. For nearly time-matched scenes and moderate OT probability, the GOES-13 detection rate (~60%) is ~15% lower than GOES-16 (~75%), which is mostly attributed to coarser spatial resolution.

NEXRAD column-maximum reflectivity and tropopause-relative echo-top height as a function of GOES OT probability were quite consistent between the two satellites however, indicating that efforts to account for differing resolution were largely successful. GOES false detections are unavoidable because outflow from nearby or recently decayed OTs can be substantially colder than the tropopause and look like an OT to an automated algorithm.

1. Introduction and Background

Updrafts that elevate cloud tops above the tropopause are common within deep convective storms over the contiguous US (CONUS), with tens of thousands of these overshooting tops (OTs) occurring in a given year (Bedka et al., 2010; Cooney et al., 2018; Liu & Zipser, 2005; Liu et al., 2020; Solomon et al., 2016). OTs in the mid-latitudes are closely connected to the seasonal evolution of convection and the tropopause and thus, most frequently occur during the early summer months when the tropopause altitude is lower than later summer months and convective activity is higher than during the winter (Cooney et al., 2018; Dai et al., 1999; Solomon et al., 2016). Convective storms that produce OTs are often associated with extremely hazardous weather on Earth. Convective storms that produce OTs often generate hazardous weather such as heavy rainfall, lightning, damaging straight-line winds, large hail, and tornadoes (Bedka et al., 2018; Dworak et al., 2012; Fujita, 1989; Liu & Zipser, 2005; Negri, 1982; Reynolds, 1980, and references therein).

Aside from the relationship between OTs and severe weather, the boundaries of updrafts, including the OT, are turbulent, which leads to irreversible mixing of air between the stratosphere and troposphere, altering the composition of both the Upper Troposphere and Lower Stratosphere (UTLS) (Herman et al., 2017; Liu & Zipser, 2005; Pan et al., 2014; Smith et al., 2017). Aircraft observations of OTs show the downward transport of ozone-rich stratospheric air into the upper troposphere (Pan et al., 2014). Within a storm system containing updrafts that surpassed the local tropopause by as much as 3 km, Pan et al. (2014) measured the upper troposphere mixing ratio of ozone 50–140 ppbv higher than in areas external to the storm system. In addition to downward transport, OTs inject water into the lower stratosphere, predominantly in the form of...
Funding acquisition: Kristopher M. Bedka, Kenneth P. Bowman
Investigation: John W. Cooney, Kristopher M. Bedka, Konstantin V. Khlopenkov, Kyle Itterly
Methodology: John W. Cooney, Kristopher M. Bedka, Konstantin V. Khlopenkov
Resources: Kristopher M. Bedka, Kenneth P. Bowman, Konstantin V. Khlopenkov
Software: John W. Cooney, Konstantin V. Khlopenkov, Kyle Itterly
Supervision: Kristopher M. Bedka, Kenneth P. Bowman
Validation: John W. Cooney
Visualization: John W. Cooney, Kyle Itterly
Writing – original draft: John W. Cooney
Writing – review & editing: John W. Cooney, Kristopher M. Bedka, Kenneth P. Bowman, Konstantin V. Khlopenkov, Kyle Itterly

To further study how overshoots impact stratospheric composition, we need to know precisely when and where they are occurring, especially in regions outside the range of the U.S. Next-Generation Radar (NEXRAD) network, where satellites often provide the only observations of overshooting convection. Within the last two decades, remote sensing observations from satellites have been used extensively to study deep convective storm precipitation characteristics and vertical structure. Some methods are essentially geometric (radar, lidar, and stereoscopy), while others are indirect (passive microwave, infrared temperature, visible, and near-infrared reflectance imaging). The Tropical Rainfall Measuring Mission (TRMM) (Kummerow et al., 2000) measured the precipitation structure within tropical deep convection with its Precipitation Radar (PR) and microwave imager (TMI) instruments (Alcala & Dessler, 2002; Liu & Zipser, 2005; Zisper et al., 2006). Its successor, the Global Precipitation Measurement Mission (GPM) (Hou et al., 2014), collected passive microwave imager data with higher spatial detail and multi-frequency radar profiles extending to a 65° latitude, compared to the ~35° latitude extent observed by TRMM. Precipitation echo tops at low reflectivity thresholds, such as a range between 5 and 20 dBZ, are a proxy for cloud top height that can be compared with various reference levels such as the equilibrium level (commonly referred to as the level of neutral buoyancy), the tropopause altitude, or the 380 K isentrope to identify OTs (Liu & Zipser, 2005; Liu et al., 2020). Convective updrafts with large ice particles or high concentrations of ice scatter microwave radiation before it reaches the satellite sensor, generating prominent brightness temperature depressions (Bang & Cecil, 2019) that have been used to infer the presence of OTs (Liu et al., 2020). Both TRMM and GPM can only observe the same storm system a few times per day which limits their capability of observing how OTs, which can have lifetimes as short as 5–15 min (Elliot et al., 2012), evolve and redevelop throughout a storm lifetime.

Two other cloud detecting satellites, CloudSat (Stephens et al., 2002) and CALIPSO (Winker et al., 2003), are sun-synchronous, nadir-pointing, low Earth-orbiting satellites that are part of the National Aeronautics and Space Administration (NASA) “Afternoon” constellation of satellites (A-Train). The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), aboard CALIPSO, and the Cloud Profiling Radar (CPR), aboard CloudSat have high spatial resolution and are very sensitive to cloud particles, which makes them ideal for identifying cloud tops. The narrow beam width of the instruments, however, require the satellites to pass directly above an overshooting event which is extremely rare (Bedka et al., 2012; Takahashi & Luo, 2014). In addition, A-Train satellite overpasses at 01:30 a.m. and p.m. local time do not coincide with the late afternoon and evening peak in deep convective activity over the continents, so the vast majority of OTs are not observed by those satellites (Cooney et al., 2018; Homeyer, Pan, Dorsi, et al., 2014).

To provide better spatial and temporal coverage, Cooney et al. (2018) identified OTs using data from the NEXRAD network over the CONUS using methods developed in Homeyer and Kumjian (2015) and Solomon et al. (2016). To assemble a 10 years OT climatology, Cooney et al. (2018) used horizontally polarized radar reflectivity ($Z_\theta$) to estimate echo-top altitude ($z_e$). OTs are identified by comparing NEXRAD $z_e$ with the primary lapse-rate tropopause altitude, as defined by the World Meteorological Organization (WMO) (WMO, 1957). NEXRAD radars, however, have a limited sensitivity to small particles (Lautenbach et al., 2005), so the echo-top altitude is generally below the cloud top altitude, which may lead to missing tropopause-overshooting events. A comparison of 10 dBZ echo-top altitudes from NEXRAD with high vertical resolution CloudSat CPR shows that NEXRAD $z_e$ are consistent with the satellite observations (Cooney et al., 2018). The two estimates are essentially unbiased with respect to each other, with mean absolute

ice, where it can sublimate and moisten the dry stratosphere. Based on aircraft measurements and a back trajectory analysis, Smith et al. (2017) found that a Mesoscale Convective System (MCS) in Minnesota and Wisconsin with persistent OT activity irreversibly injected approximately 6.6–13.5 kilotones of water into the stratosphere, which corresponds to ~0.0007%–0.0013% of the stratospheric water vapor reservoir (Smith et al., 2017). In another study, Herman et al. (2017) analyzed JPL Laser Hygrometer (JLH Mark2) observations of stratospheric water vapor mixing ratio downwind of OTs and found a 4–6 ppmv enhancement above the typical <5 ppmv background of the stratospheric overworld (above the ~380 K isentrope level) 1–7 days after the OT events. Although these numbers are small and localized, changes in the stratospheric composition of this magnitude can have a significant impact on chemistry and Earth’s radiation balance (Anderson et al., 2012, 2017; Dessler & Sherwood, 2004; Forster & Shine, 1999; Holton et al., 1995; Stohl et al., 2003).
differences ranging from 0.5 to 1 km. The largest differences between NEXRAD and CloudSat typically occur in areas of sparse radar coverage (Cooney et al., 2018). In these regions, the vertical resolution of echoes near the tropopause is inadequate, which can also result in missed OT detections.

In addition to uncertainties in the altitude assignment of echo tops, global tropopause altitude estimates from reanalyses are also imperfect. Large uncertainties can significantly alter OT detections because, as shown by Cooney et al. (2018), the number of OTs detected by NEXRAD drops logarithmically as a function of altitude above the tropopause. Thus, if the tropopause is estimated too high or too low, this can impact whether or not an OT is detected. In a study of reanalysis estimated tropopauses, Xian and Homeyer (2019) compared tropopause altitudes derived from JRA-55, ERA-Interim, MERRA-2, and CFSR with radiosonde measurements and found altitude biases were typically less than ±150 m and the error was comparable to the model resolution. The largest biases were observed in the subtropics. Xian and Homeyer (2019) also found that MERRA-2 tropopause altitude estimates were more often biased high while the other reanalysis estimates were more often biased low, relative to the radiosondes. Despite these limitations and uncertainties, due to the broad spatial coverage of the NEXRAD network, which observes storm morphology and intensity with high spatio-temporal sampling, we feel that NEXRAD-based OT detections provide reliable reference data for validating satellite-derived OT detections.

The infrared (IR) imagers aboard the Geostationary Operational Environmental Satellites (GOES) have collected brightness temperature ($T_b$) observations with nadir resolutions of 4 km/pixel for GOES-8–15 and 2 km/pixel for GOES-16 and -17, comprising a data record extending 25+ years across the Americas and adjacent oceanic regions where tropopause-penetrating updrafts frequently occur. Bedka and Khlopenkov (2016) introduced an IR-based method for OT detection that improved upon a previous method described in Bedka et al. (2010). This method has been further developed and is described in Khlopenkov et al. (2021), which serves as a companion to this paper. The algorithm attempts to quantify how much a cluster of pixels resemble an OT on the basis of spatial $T_b$ patterns in GOES ∼11 μm IR images and their relationship with a tropopause temperature forecast or reanalysis. Anomalously cold $T_b$ are caused by persistent moist-adiabatic ascent into the lower stratosphere, allowing updrafts to be much colder than the temperature in the surrounding environment (Bedka & Khlopenkov, 2016). IR $T_b$ colder than the tropopause temperature likely represent an intrusion into the lower stratosphere. Changes in spatial resolution throughout the GOES data record affect how prominently OTs appear and how much colder they are relative to the tropopause (Khlopenkov et al., 2021). Satellite-based OT detection approaches have often been validated using manual OT identifications from human experts because the algorithms are designed to detect OT-like patterns in the imagery itself, independent of whether or not they are detected with weather radars or other methods (Bedka & Khlopenkov, 2016; Khlopenkov et al., 2021; Kim et al., 2017). These manual identifications are able to provide a glimpse into the accuracy of satellite-based detections, however, they are time-consuming, so previous studies only analyze detections from a few OT storm events. Despite radar limitations with detection of the true physical cloud top, NEXRAD databases of OT identifications can serve as a stable reference to enable comprehensive satellite OT detection validation and quantification of how IR-based detection methods are impacted by changes in image resolution throughout the satellite imager data record.

Though both NEXRAD and GOES methods seek to identify tropopause-penetrating OT updrafts, the results generated across multi-year data records are not entirely consistent. Khlopenkov et al. (2021) show significant differences between cloud top IR temperature patterns and the apparent locations of updraft cores through comparison of Visible Infrared Imaging Radiometer Suite (VIIRS), GOES-13, and GOES-16 imagery as well as NEXRAD precipitation echoes near the tropopause. A GOES climatology from Bedka et al. (2010) and NEXRAD climatology from Cooney et al. (2018) find large regional differences in the geographic distributions of OTs. A majority of GOES OTs were detected in the Southeast United States while NEXRAD detected a maximum in the Central Plains, with fewer OTs in the Southeast. Recent GOES OT detection algorithm improvements have sought to improve agreement with NEXRAD, but the statistical agreement between the two datasets has not yet been quantified. The goal of this study is to analyze the degree of agreement between OT detection methods, and demonstrate how the 4x increase in spatial sampling from GOES-13 to GOES-16 impacts OT algorithm detection performance.
2. Data and Methods

2.1. Meteorological Analyses

Both the GOES and NEXRAD OT detections use meteorological data from the hourly Modern-Era Retrospective analysis for Research and Applications Version 2 (Bosilovich & Lucchesi, 2016) tropopause analysis, contained in the 2 days, 1 H, Time-Averaged, Single-Level, Assimilation, Single-Level Diagnostics V5.12.4 (“MERRA2_400.tavg1_2 days_slv_Nx”) collection. MERRA-2 data, originally at ∼0.625° × ∼0.5° spatial resolution, are interpolated in space and time to the GOES IR longitude-latitude grid from the OT detection product, described in Section 2.3. The MERRA-2 “TROPT” and “TROPPB” parameters are used in this analysis, which represent a blended estimate of the tropopause temperature and pressure, respectively, based on a combination of the WMO definition of the primary lapse-rate tropopause (WMO, 1957) and equivalent potential vorticity. Tropopause geopotential height is derived from the tropopause pressure and temperature with the hypsometric equation. In this paper we will refer to the tropopause temperature and height as $T_{MERRA2}$ and $Z_{MERRA2}$, respectively. As shown by Khlopenkov et al. (2021), tropopause estimates directly from reanalyses or model forecasts depict high frequency variability which may not be realistic or meaningful for determining OT locations. Khlopenkov et al. (2021) smooths the tropopause analysis to minimize artifacts when compared with IR $T_b$ to enable reliable OT detection. These smoothed tropopause analyses are also used in this study to derive NEXRAD OT detections (described in Section 2.2).

2.2. NEXRAD OT Analysis

The NEXRAD OT detection algorithm uses 5-min GridRad (V4.1) analyses, which are based on WSR-88D radar data obtained from the National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Information (NCEI) (Cooney et al., 2018; Homeyer & Bowman, 2017; Homeyer & Kumjian, 2013; Solomon et al., 2016). GridRad data in the UTLS region are stored on a regular longitude-latitude-altitude grid with a resolution of ∼0.02° × 0.02° × 1 km [−2 × 2 × 1 km], respectively. The analysis domain is the rectangular region from 125°W to 66°W and 24°N to 50°N (shown in Figure 1), which covers most of the continental United States. Subsets of this domain are analyzed, as shown in Figure 2. Four of the study dates were chosen because GridRad data were already on hand for the Bedka et al. (2018) study, while the other dates were randomly selected due to the presence of widespread and long-lived convection across the US.
Each analysis uses all available radar azimuth scans (elevation angles) within a ±3.8 min time window centered on the 5 min GridRad analysis times. Observations from multiple radars are merged using weights that depend on the distance from the radar to the analysis location and on the time difference from the nominal analysis time (Homeyer & Bowman, 2017). The data include weighted-average horizontal reflectivity factor (\(Z_H\)) from multiple observations, averaging weights, number of azimuth scans observing each volume, and the number of azimuth scans with echo in each volume.

Cooney et al. (2018) identified tropopause-penetrating convection by comparing the radar echo-top height (\(z_e\)) to the primary lapse-rate tropopause height derived from the ERA-Interim reanalysis (\(Z_{ERA}\)). A NEXRAD OT is defined as a contiguous group of overshooting GridRad columns. To allow for a consistent comparison with the GOES OT algorithm, this study identifies NEXRAD overshoots by comparing \(z_e\) with MERRA-2 tropopause altitudes interpolated to the GridRad grid (\(Z_{MERRA2}\)) and is defined as a contiguous region with \(z_r = z_e - Z_{MERRA2} > 0\) km, denoted in this paper as OTN.

NEXRAD radars, which transmit at a 10 cm wavelength, are only able to observe precipitable hydrometeors. The nominal reflectivity threshold for the existence of a valid echo is \(\sim 5\) dBZ, but at the maximum range used for the merging of multiple radars, the minimum detectable signal is \(\sim 7.5\) dBZ (Cooney et al., 2018). A lower echo top reflectivity threshold can provide a better estimate of the true top altitude but is also more susceptible to noise. Cooney et al. (2018) applied additional echo-top criterion in order to avoid false positives that occur when the reflectivity threshold nears the noise limit of the NEXRAD radars. The echo-top height, which is discretized to a 1 km vertical grid in the UTLS region, is defined by Cooney et al. (2018) as the highest altitude in each column with \(Z_H \geq 10\) dBZ, with the additional constraint that the reflectivity in the altitude bin containing the tropopause be greater than or equal to 20 dBZ. This additional criterion serves to filter spurious above-tropopause echoes (Cooney et al., 2018). These thresholds remove potentially false identifications but could also inhibit detection of actual OTs, as described in Cooney et al. (2018). In our analysis, we tested other criteria, such as 40 dBZ echo-top altitudes (not shown) and found it to be too restrictive and removed many valid OTs and thus are not used in this analysis.

We use the same criteria in this paper to define OTN, however, to evaluate the sensitivity of the detections to the choice of reflectivity and tropopause-relative echo-top height thresholds, we also derive two additional measures of OTN that we will refer to as possible-OTN: (a) \(\geq 20\) dBZ pixels with \(z_r > -1\) km and (b) \(\geq 20\) dBZ pixels with \(z_r > -2\) km. At the tops of strong updrafts, smaller particles, which have little or no reflectivity signal, are often located above larger or more abundant particles that produce \(\geq 20\) dBZ reflectivity. Cloud tops with weak to no detectable reflectivity from NEXRAD (i.e., very small cloud particles) could induce a notable IR \(T_b\) perturbation while the 10 dBZ echo top is beneath the tropopause. NEXRAD
echo-top altitudes have uncertainties (standard deviation) of approximately ±1 km when compared with CloudSat CPR measurements (Cooney et al., 2018). In addition, uncertainties in the estimated tropopause altitude can also bias OT detection by ~1 km (Cooney et al., 2018; Homeyer, Pan, & Barth, 2014; Solomon et al., 2016). Therefore, comparisons of GOES OTs with the less restrictive OT criteria provide a reasonable “best case scenario” for OT detection validation. This is tested with sensitivity experiments described below.

The GOES satellite can image the entire study region with relatively uniform coverage, but the NEXRAD network has a complex coverage pattern. Figure 1 shows the average number of contributing NEXRAD radar azimuth scans per analysis time within each grid column in the GridRad domain. Coverage is good in the Plains, Midwest, and Southeast regions of the United States. Coverage in the western US, offshore, and outside international borders, is generally poorer; and the likelihood of NEXRAD detecting an overshooting event is lower. Homeyer (2014) shows that the achieved vertical resolution is ≤1 km if at least three radars observe a column. When precipitation is present, this corresponds to ~40 radar azimuth scans. Therefore, the GOES and NEXRAD OT comparison is restricted to locations with ≥40 NEXRAD instantaneous radar azimuth scans in a column.

### 2.3. GOES OT Analysis

GOES OT detections are performed using the methods described in Khlopenkov et al. (2021). Similar to the NEXRAD analysis, an individual GOES OT is defined as a contiguous group of GOES pixels identified as overshooting the tropopause. Information about each overshoot includes (a) estimated parallax correction, (b) brightness temperature (Tb), (c) tropopause-relative IR brightness temperature (ΔT_{MERRA} = Tb - T_{MERRA}), (d) difference between the minimum cloud top Tb and the anvil-mean Tb (IR-anvil BTD), and (e) OT probability. Parallax corrections, required to match satellite observations with ground-based observations like NEXRAD, are derived by first matching the GOES IR temperature to the MERRA-2 profile to derive a cloud top height. For pixels with Tb colder than the tropopause, the stratospheric portion of the MERRA-2 temperature profile is modified to cool at a constant lapse rate, 6 K/km for GOES-16 and 4.5 K/km for GOES-13, to enable a height assignment (Griffin et al., 2016). This height is used in combination with the pixel position relative to GOES satellite nadir (75°W for GOES-13 and 89.5°W for GOES-16) to compute the correction. The GOES-13 data are produced on a ~4 km (28 pixels/degree) horizontal grid for each GOES-13 scan. GOES-16 data are produced on a ~2 km (56 pixels/degree) horizontal grid. Both GOES-13 and GOES-16 were analyzed over the spatial domains and days denoted in Figure 2.

Rather than applying fixed weights to a set of parameters, as is the case in Bedka and Khlopenkov (2016), Khlopenkov et al. (2021) applies an approach similar to fuzzy logic to assign an OT probability. Fuzzy logic is a process that is based on assigning variables values or “degrees of truth”, ranging from 0, not likely, to 1, extremely likely (Wilks, 2005). Likelihood scores from four parameters, (a) ΔT_{MERRA}, (b) IR-anvil BTD, (c) anvil area surrounding a pixel, and (d) spatial coherence of the anvil, are combined in Khlopenkov et al. (2021), their Figures 10 and 11) by a set of sensitivity functions based on empirical and statistical analyses of human-identified OTs. The shape of the sensitivity functions for ΔT_{MERRA} and IR-anvil BTD were found to agree well with cumulative frequency distributions derived from GOES observations of NEXRAD identified OTs. Khlopenkov et al. (2021) weights ΔT_{MERRA} the heaviest because an event with Tb much colder than the tropopause is almost certainly an OT, despite what other parameters might suggest. The remaining three parameters are multiplied together and thus are all weighted the same.

Khlopenkov et al. (2021) validates OT detections from the algorithm by using human OT identifications based primarily on visible channel texture, with some contribution from IR temperature in the identification process. OT identification can be quite subjective, so Khlopenkov et al. (2021) derives two sets of OT “truth,” a conservative OT mask containing very prominent OT signatures, and a liberal mask OT mask containing all image features that a set of human analysts thought “might be” OTs. Khlopenkov et al. (2021) found that the algorithm was optimized at a probability of 0.61 for the conservative mask and 0.23 for the liberal mask. Due to this wide range of optimal values, we decided to only plot OTG with OT probability ≥0.5 in the case study examples that are shown and discussed below. An OT probability ≥0.5 implies the GOES algorithm has a ≥50% confidence that the GOES detection is a true OT. As will be shown later, low probability thresholds result in a higher rate of falsely identified OTs, while using a threshold that is too high causes the algorithm to miss OTs that are warmer or less prominent relative to the anvil.
2.4. Temporal Matching of NEXRAD and GOES Analyses

During the study days listed in Figure 2, GOES-13 was fully operational while GOES-16 was still in its pre-operational testing phase. GOES-13 collected data every 5 to 30 mins depending on whether the satellite was in rapid scan mode to monitor severe storms (up to 5 min frequency), in nominal operational scan mode (15 min frequency), or in “full disk” scan mode. The “full disk” scans occur every 3 hrs, with scans beginning at 0245, 0545,…2345 UTC, and take approximately 30 mins to complete, thus GOES-13/16 and GridRad data matches are not possible during these gaps. GOES-16 data was collected at 1 min intervals over 1000 × 1000 km Mesoscale Domain Sectors (April and May 2017, and June 28–29, 2017) and at 5 min intervals over 60°W–115°W (other remaining June and July 2017 dates). We only use GOES-13 and GOES-16 scans nearest to the NEXRAD 5 min analysis times to compare the NEXRAD OT detections. NEXRAD analyses are compared with 1157 GOES-13 and 2601 GOES-16 scenes, respectively.

2.5. Spatial Matching of NEXRAD and GOES Analyses

Spatial matches between individual GOES and NEXRAD overshoots are identified by mapping parallax-corrected GOES data to the GridRad grid. The GOES-13 IR data grid is approximately half the resolution of the GridRad grid in each dimension, so each parallax-corrected GOES-13 pixel is mapped to the nearest four GridRad grid boxes. If multiple GOES-13 pixels are mapped to the same GridRad grid box, the data values are averaged. For regions with pixels identified as GOES OTs, only the OT pixels are included in averaging $T_b$. The largest magnitude OT probability and IR-anvil BTD are retained in each grid box and are not averaged. This is done to preserve the OT data and avoid contamination of OTs by surrounding pixels that are not part of an OT but are parallax-corrected into the same grid box. The GOES-16 IR data grid has approximately the same horizontal resolution as GridRad, so each parallax-corrected GOES-16 pixel is mapped into the nearest GridRad grid box.

Because the GOES OT algorithm is applied to the original (non-parallax corrected) GOES images, the parallax corrections cause some spatially distinct GOES OTs to be contiguous with one another when mapped to the GridRad grid. In this case, the GOES OTs are combined into a single overshoot. Contiguous GOES overshoots on the GridRad grid with OT probability $\geq$0.5 are denoted OTG in this paper.

A GOES OT with any OT probability was considered a match with a NEXRAD OTN or possible-OTN if the two were within 10 km of each other for GOES-13 and -16. This match radius compensates for storm motion in the time between the GOES and NEXRAD scan, imprecision with GOES pixel navigation, and uncertainty of OT cloud top height retrievals leading to errors in parallax corrections. An OTN can vary in size depending on updraft width and proximity to each other. Rather than deriving an “OT center” coordinate corresponding to the peak OTN height which can be misleading in large overshoots where multiple OTs are adjacent to each other, the outer periphery of the OTN is used as the basis for computing distance to OTG (see Figure S1 for a graphical representation of the matching). It is important to note that more than one OTG can be within 10 km of an OTN and vice versa. Therefore, the total number of OTG matching with OTN is not always equivalent to the total number of OTN matching with OTG.

Due to lower horizontal spatial resolution, GOES-13 is unlikely to be able to detect very small OTs that may be observable by NEXRAD and GOES-16. To provide a fairer comparison between GOES-13 and -16 OT detection capability, as well as to further quality control NEXRAD OT identifications, only OTG, OTN, and possible-OTN that contain at least five GridRad grid boxes are considered. This threshold should also assist in filtering spurious above-tropopause echoes. For possible-OTN regions, there must be at least five contiguous GridRad pixels, analogous to what is done with OTN, to identify a distinct OT region and compute the probability of OT detection (POD) statistic, defined in Equation 1. On the other hand, it is possible that a cluster of less than five anvil- or tropopause-penetrating echo top pixels could impact GOES image appearance to trigger an OTG detection. Therefore, an OTG is considered to be a "hit" if just one possible-OTN pixel is within 10 km in order to avoid unrealistically penalizing the GOES false alarm ratio (FAR), defined in Equation 2, and provide a best-case-scenario perspective of algorithm performance. Note, Equations 1 and 2 can apply to either OTN, as shown in the equations, or possible-OTN, not shown.
3. Results

3.1. Analysis of Dependence on NEXRAD Overshooting Criteria

As an introduction to some of the successes and challenges associated with OT detection, Figure 3 shows an example of GOES-16 and NEXRAD OT detections in a squall line over the Southeast United States. The black contours represent GOES-16 OT identifications with OT probability $\geq 0.5$ (i.e., OTG). The yellow arrows on the map in Figure 3a point to OTG that are in regions in which echo is observed above the tropopause but no OTN is identified. OTN are not identified for these locations because there are not at least five contiguous pixels with $\geq 20$ dBZ echo nearest to the MERRA-2 tropopause altitude level (Figure 3f), though some 10 dBZ echo was above the tropopause (Figure 3e). The GOES data in Figures 3a–3c show the OTG were present in regions where $T_b$ was colder than the tropopause and at least 4 K colder than the surrounding anvil, often where visible texture, indicative of updrafts and gravity waves, was present. Other GOES cold region detections with OT probability $<0.5$ had warmer $\Delta T_{MERRA}$ and/or lower IR-anvil BTD (Figure 3c). Therefore, the OTG identifications, indicated by the yellow arrows, are likely OTs that were missed by the OTN method due to stringent identification criteria.

Figure 3d shows the OTG identifications are all located within, or very near, high $Z_{H}$ regions, indicative of localized strong updraft regions. Figure 3c shows five of seven OTG regions were not co-located with an OTN, including the OTG identified by the yellow arrows. Six of the seven OTG agree with the $z_r > -2$ km possible-OTN criteria (blue contours in Figure 3c), demonstrating how narrow regions of radar echo near the tropopause can generate cloud top perturbations, and the value of using multiple validation criteria for understanding GOES detection performance. Figure 4 provides the vertical cross-section through some of the storm updraft regions, as observed by NEXRAD. The vertical structure of the updrafts verifies the OTG identifications along the magenta line in Figure 3c. On the other hand, the small area with $z_r > -2$ km identified by the red arrow has no GOES detection at all, due to very small anvil and $\Delta T_{MERRA} > 5$ K, a value present throughout the entire anvil for storms to the east. It is possible that the $T_b$ is biased warm or the possible-OTN does not surpass the tropopause. Thus, not all possible-OTN regions have cloud tops above the tropopause; a point that will be reinforced with other examples below.

3.2. Geographic Distribution of Overshoots

Figures 5a and 5c display OTN and GOES-16 OTG, respectively, accumulated across all 15 study days. OTG are identified more frequently than OTN, particularly in the Southeast US. An OTN, possible-OTN, and GOES-16 OTG contain on average 36.4, 40.0, and 40.2 GridRad grid boxes (not shown), respectively. This indicates that differences between the maps can be mostly attributed to differences in the frequency of detections, and not the areal coverage of detections. The spatial distributions of OTG and OTN detections generally agree well across the domain with the best agreement occurring in the Central Plains. In order to quantify the spatial agreement, OTG and OTN detections were binned into 0.5° x 0.5° grid boxes and grid boxes with both an OTG and an OTN were considered matches (not shown). Across the entire domain, 75% of 0.5° x 0.5° grid boxes with an OTG detection also contained an OTN detection, while 94% of grid boxes with an OTN detection were located in the same grid box as an OTG detection. For the Central Plains, the match rate increases for both OTG and OTN to 90% and 97%, respectively. Figure 5b shows the geographic distribution of NEXRAD possible-OTN with $z_r > -2$ km. There are nearly 3x more possible-OTN (Figure 5b) detections as OTN (59428 vs. 22514), but we see a better agreement in the spatial distribution of detections relative to what is depicted by OTG. Grid boxes with an OTG detection matched with possible-OTN
increases from 75% to 95% across the entire domain. In particular, over the Southeast US, the OTG distribution matched possible-OTN far better than the OTN (70% vs. 90%). This may be due to higher tropopause altitudes at lower latitudes where higher-reaching updrafts are required to satisfy the tropopause-penetrating echo top OTN criteria. Another possibility is that the equilibrium level, and altitude of the broader anvil, is more often well below the tropopause when tropopause altitudes are high (>15 km), which could yield larger IR-anvil BTD for storms that nearly, or marginally, exceed the altitude of the tropopause in the Southeast US (Solomon et al., 2016).

Figure 5d shows the difference between GOES-16 OTG occurrences when the radar azimuth scan threshold of 40 is not applied and Figure 5c overshoot detections. The remaining points correspond to OTs that were removed from analysis by applying the radar azimuth scan threshold. Over the CONUS, OTG differences are very small, indicating that radar coverage is not a reason for the smaller number of OTN relative to
Figure 4. Vertical cross-section through updraft regions in Figure 3. White line represents estimated MERRA-2 tropopause altitude through the region. Yellow (black) circles represent OTN (z_r > −2 km possible-OTN) locations. All circles, regardless of color, represent Geostationary Operational Environmental Satellites (GOES)-16 OTG detection locations with Overshooting Top (OT) probability ≥0.5. Image produced using gridrad_viewer software, created by Dr. Cameron Homeyer.

Figure 5. Number of instances in the selected analyses that overshooting convection is identified for each GridRad pixel. The number in the parentheses of each map is the total number of overshooting events identified by the respective methods. (a) 10 dBZ OTN above MERRA-2 tropopause. (b) 20 dBZ OTN z_r > −2 km. (c) Geostationary Operational Environmental Satellites (GOES)-16 OTG. Panels (a), (b), and (c) include only pixels where the number of radar azimuth scans observing a grid column ≥40. (d) GOES-16 OTG with no radar azimuth scan threshold applied minus Panel (c).
OTG. Even in the Western US, where OTN identifications were rare on the dates being studied and the radar coverage is poor, there are few OTG, which indicates that there are, in fact, few OT occurrences there during our study days.

3.3. Storm Lifecycle OT Detection Analyses

The examples shown throughout this study highlight both the successes and challenges associated with IR-based OT detection. While it is impossible to achieve 100% success with such detections from a statistical standpoint, GOES OT detections can be used to map storm tracks for analyzing severe- or aviation-weather hazards, as well as identify concentrated regions of tropopause-penetrating updraft activity in support of climate research goals like those of the NASA Dynamics and Chemistry of the Summer Stratosphere (DCOTSS) mission. The 1 min sampling of GOES-16 provides a unique opportunity to compare detections with the coarser 5 min NEXRAD echo tops throughout the lifetime of convection events. Figure 6 (S2) shows examples from April 5–6, 2017 and May 16 to 17 derived from 5 min Next-Generation Radar (NEXRAD) data with Geostationary Operational Environmental Satellites (GOES)-16 Overshooting Top (OT) probability derived from 1 min imagery (panels c–d) from the same events.

Figure 6. A comparison of 20 dBZ \( z_r \) from (a) April 5 to 6, 2017 and (b) May 16 to 17 derived from 5 min Next-Generation Radar (NEXRAD) data with Geostationary Operational Environmental Satellites (GOES)-16 Overshooting Top (OT) probability derived from 1 min imagery (panels c–d) from the same events.

OTG. Even in the Western US, where OTN identifications were rare on the dates being studied and the radar coverage is poor, there are few OTG, which indicates that there are, in fact, few OT occurrences there during our study days.

3.3. Storm Lifecycle OT Detection Analyses

The examples shown throughout this study highlight both the successes and challenges associated with IR-based OT detection. While it is impossible to achieve 100% success with such detections from a statistical standpoint, GOES OT detections can be used to map storm tracks for analyzing severe- or aviation-weather hazards, as well as identify concentrated regions of tropopause-penetrating updraft activity in support of climate research goals like those of the NASA Dynamics and Chemistry of the Summer Stratosphere (DCOTSS) mission. The 1 min sampling of GOES-16 provides a unique opportunity to compare detections with the coarser 5 min NEXRAD echo tops throughout the lifetime of convection events. Figure 6 (S2) shows examples from April 5–6 and May 16–17, 2017 (May 18–19, 2017 and June 28–29, 2017), where the maximum 20 dBZ \( z_r \) and highest OT probability in each NEXRAD and GOES GridRad grid box is recorded throughout the event. Tropopause-relative echo tops are not as high on April 5–6, 2017, which corresponds to slightly lower OT probabilities. For both cases though, GOES-16 detections align quite well with regions in which the \( z_r > -2 \) km possible-OTN criteria is met, however, brief drops to lower echo-top heights will impact the validation statistics.

Figures S3–S4 provide the OT storm tracks for the analysis dates observed by GOES-16 at 5 min intervals. One notable observation is that early in some storm lifetimes, OTs are detected by NEXRAD but not GOES. NEXRAD echo tops are high relative to the tropopause, but the GOES \( T_b \) are warm relative to the tropo-
This is likely due to storm tops not being optically thick when storms initially develop. During this time, particles at the tops of updrafts may not be very dense so it is possible radiation measured by the satellite is from deeper within the cloud where it is warmer (Sherwood et al., 2004). This reduces the magnitude of the temperature difference between the top and the tropopause and ultimately reduces the OT probability, causing the GOES IR-based OT detection method to miss the OT event.

3.4. Matching Statistics Overview

GOES-13/16 OT detection accuracy is quantified in terms of POD and FAR, defined in Equations 1 and 2 in Section 2.5, which are combined to form a receiver operating characteristic (ROC) curve, shown in Figure 7 for GOES-13/16 OT detections with varying OT probability thresholds. GOES-13 (solid blue line) at OT probability ≥0.5 (OTG) corresponds to ~0.64 POD and ~0.49 FAR, relative to OTN. GOES-16 (solid red line) at OT probability ≥0.5 corresponds to ~0.76 POD and ~0.49 FAR with OTN. The GOES-16 POD is better than GOES-13 for all OT probability thresholds, with comparable FARs between the two satellites. It is important to note that these statistics are presented under the assumption that the NEXRAD algorithm does not misidentify or fail to recognize OT regions. It has been shown with Figure 3 and Section 3.1 that

![Figure 7](image_url)

**Figure 7.** Probability a Geostationary Operational Environmental Satellites (GOES) Overshooting Top (OT) detection is within 10 km of a Next-Generation Radar (NEXRAD) detection as a function of GOES OT detection false-alarm ratio for nine different probability thresholds. The red line corresponds to GOES-16 OT detections while the blue line corresponds to GOES-13 OT detections. The solid, dashed, and dotted lines correspond to OTN, possible-OTN 20 dBZ $z_r > -1$ km, and possible-OTN 20 dBZ $z_r > -2$ km, respectively. The numbers next to the points indicate the OT probability threshold.
echo tops beneath the tropopause can generate OT-like texture in visible imagery and prominent cold regions detected by the GOES algorithm. Therefore, the statistics based on identification of echo tops above the tropopause are not wholly representative of OTG performance.

To account for the tropopause-relative echo-top height and width (i.e., number of overshooting pixels) criteria failing to recognize some OT regions, the dashed and dotted lines on the figure correspond to POD and FAR for possible-OTN $z_r > -1$ km and $> -2$ km, respectively. Figure 7 shows that possible-OTN are less frequently detected by OTG, especially $z_r > -2$ km. Thus, not all possible-OTN generate a reliably detectable signal in GOES IR imagery and may not actually have cloud top above the tropopause. For OT probabilities $\geq 0.5$, POD progressively decreased to $\sim 0.61$ ($\sim 0.45$) for $z_r > -1$ km ($-2$ km), but FAR decreased by a greater amount to $\sim 0.22$ ($\sim 0.12$), relative to the statistics for the strict OTN criteria. Because FAR decreased by a greater amount than POD, this result reinforces our belief that including possible-OTN comparisons with OTG provides a more complete assessment of GOES detection performance than OTG comparisons with OTN alone, even though not all possible-OTN have cloud top above the tropopause.

The 0.45–0.76 POD range identified here for GOES-16 across the possible-OTN to OTN categories falls within the 0.51–0.95 POD range identified by Khlopenkov et al. (2021), using a database of human-identified OTs, from GOES visible and IR imagery, where inclusion of lower confidence OTs (i.e., their “liberal mask”) corresponded to lower POD. As noted above, GOES-13 POD decreased relative to GOES-16 by $\sim 15\%$ for the same OT probability, similar to Khlopenkov et al. (2021). FAR from Khlopenkov et al. (2021) for the liberal mask was $\leq 0.15$ for both satellites, with GOES-16 being more accurate than GOES-13, which again is comparable to the results from this study. Though the two validation datasets differ, comparable results give us confidence that the NEXRAD-based method accurately represents OT detection performance.

Figure 8. Box plot histogram showing the maximum composite reflectivity measured within 10 km of Geostationary Operational Environmental Satellites (GOES)-13 and GOES-16 Overshooting Tops (OTs) for varying OT probability identifications. OT probability bin sizes are 0.05. The horizontal line within each box corresponds to the median reflectivity value. The blue and red boxes indicate GOES-13 and GOES-16 identifications, respectively. The corresponding colored numbers near the bottom x-axis provide the total number of GOES OT detections at 0.1 level increments. The magenta shading illustrates the overlapping reflectivities between GOES-13 and GOES-16 OT detections for each bin.
3.5. GOES OT Detection Statistical Relationship With Composite Reflectivity and Echo Top

Larger NEXRAD reflectivity measurements often correspond with stronger updrafts, and stronger updrafts are more likely to overshoot the local equilibrium level and thus be OTs. However, as shown in Figure 3d, large column maximum reflectivity does not guarantee high echo tops near to or above the tropopause. Therefore, to further quantify detection performance and consistency between GOES-13 and -16, Figures 8 and 9 present box plot histograms of the inter-quartile range (IQR) of maximum NEXRAD composite reflectivity and 20 dBZ $z_r$, respectively, measured within 10 km of GOES OT algorithm detections for varying OT probability bins. The extrema are not shown because they occupy almost the entire range, often due to temporal/spatial matching issues between GOES and NEXRAD data as well as sometimes errant GOES OT detections, so we just focus on IQR in order to highlight notable relationships in the data. The premise behind this analysis is that if GOES-13 and -16 $T_b$ and OT probability are adequately normalized, then the IQR should be similar between satellites for a particular OT probability bin, regardless of differences in POD discussed above.

As shown in Figure 8, higher OT probabilities, indicative of colder, more prominent anvil-relative tops, correspond to higher reflectivity measurements by NEXRAD. The IQR also narrows considerably for OT probability $\geq 0.5$, indicating that such detections are most reliable. Figure 9 shows that there is a steady increase in median echo top relative to the tropopause with increasing OT probability, further reinforcing that, in general, colder cloud tops correspond to higher echo tops. Unlike composite reflectivity, the overlap in data from the two satellites is much closer, with GOES-13 being biased lower by 0.5–1.0 km, most notably for detections with OT probability <0.5. For OT probability >0.5, the tropopause-relative echo-top height distributions are above $z_r$ –1 km and –2 km, the two possible-OTN criterion. Data beneath the bottom end of the IQR for $z_r$ –2 km is what contributes to the FAR for the dotted lines shown in Figure 7. This indicates that even extremely cold, prominent cloud tops can be dislocated from true updraft regions. Selected examples are presented in the following sections that provide context for these results (Figures 7–9) and highlight challenges associated with IR-based OT detection.
3.6. GOES-13 and GOES-16 OT Detection Comparisons

Khlopenkov et al. (2021) shows how OT regions IR $T_b$ were measured progressively warmer as pixel spacing increased from 375 m in VIIRS to 2 km in GOES-16 and 4 km in GOES-13. Larger pixels smooth (average) temperature values across the pixel while smaller pixels and improved sensor noise characteristics with more modern sensors like GOES-16 and VIIRS are able to more precisely observe the extremely low temperatures within OTs (Hillger et al., 2013; Khlopenkov et al., 2021). In addition, smaller pixel sizes more precisely depict spatial temperature gradients, influencing the OT IR-anvil BTD computation and therefore the OT probability.

The Khlopenkov et al. (2021) OT detection algorithm attempts to correct for these resolution differences, but the examples provided in Figures 10 and 11 demonstrate the challenges associated with detection consistency across sensors. Since both the GOES-13 and GOES-16 algorithm use the same MERRA-2 tropopause data, differences in $\Delta T_{\text{MERRA2}}$ (Figures 10a and 10b) are due to $T_b$ measurement differences. In the easternmost cell in Figure 10a (black arrow, 45.2°N, 96°W), the GOES-13 $\Delta T_{\text{MERRA2}}$ is over 5 K warmer than…
that measured by GOES-16 (0 K vs. −5 K, respectively). Note there is both an OTN (blue contours in Figure 10c) and GOES-16 OTG detection (black contours in Figure 10d) in this cell. Such $T_b$ differences also impact the IR-anvil BTD (Figures 10c and 10d), where GOES-13 (GOES-16) BTD is −3 to −4 K (−8 to −9 K). The combination of warmer GOES-13 $\Delta T_{\text{MERRA}2}$ and weaker IR-anvil BTD reduce OT probability below 0.5, as shown in the OT probability maps provided by Figures 11a and 11b.

The possible-OTN regions in Figures 10d and 11b are better correlated with areas of visible texture (Figure 11d) that human analysts associate with OT regions. While the highest OT probability areas (≥0.9, magenta shading in Figures 11a and 11b) are co-located with texture and 20 dBZ echo top just beneath the tropopause, other relatively high probability regions with $\Delta T_{\text{MERRA}2} \lesssim −5$ K (yellow and brown arrows) are dislocated from updrafts in stratiform precipitation regions (Figure 11c). Such detections would be considered false alarms, penalizing the statistics depicted by Figures 7–9.

The case shown in Figure 10 and Figure 11 was chosen to visually illustrate how satellite image resolution can impact OT detections from IR data. In order to correct for resolution, we need to understand the magnitude and range of the $T_b$ biases. Khlopenkov et al. (2021) shows a cumulative frequency diagram...
of ΔT_{MERRA} and IR-anvil BTD for OTN, possible-OTN, and time periods during storm lifetimes when severe weather was reported near GOES OT regions. The results clearly show differences in distributions between GOES-13 and -16, where increasingly high echo tops were, on average, colder and slightly more prominent relative to the anvil. From a qualitative standpoint, Figures 10a and 10b show comparable OT probability values between GOES-13 and -16 for the same OT regions, with differences typically less than 0.1. A magnified region within Figures 10 and 11 is provided in Figure 12 and shows T_b gradients are not as sharp within GOES-13 as GOES-16. This reduces the IR-anvil BTD magnitude by as much as 5.5 K and contributes to these small OT probability differences.

Figure 13 shows histograms of the difference between the coldest T_b measured by Geostationary Operational Environmental Satellites (GOES)-16 subtracted by the minimum T_b measured by GOES-13 within 10 km of OTN (blue). Histogram of the minimum IR Overshooting Top (OT)-anvil BTD calculated from GOES-16 IR data subtracted by the minimum IR OT-anvil BTD from GOES-13 OT data within 10 km of an OTN (red). The average difference and standard deviation are plotted near the bottom x-axis in corresponding colors for each histogram.

Figure 12. Comparison of Geostationary Operational Environmental Satellites (GOES)-13 and GOES-16 ΔT_{MERRA} for zoomed in region of Figures 10a and 10b valid June 13, 2017 at 2300 UTC. (a) GOES-13 ΔT_{MERRA}. The black contours represent GOES-13 Overshooting Tops (OTs) with OT probability ≥0.5 (i.e., OTG). (b) Same as Figure 12a but for GOES-16.
3.7. Additional Challenges Associated With IR-Based OT Detection

Figure 8 shows that the majority, but not all, of GOES OT detections are within high reflectivity regions. One noted source of GOES OT detection error from Figures 10 and 11 is identification of cold anvil cloud as an OT. Detections in anvil regions may be triggered by cold outflow from a previous OT that decayed a few minutes prior to a GOES image, simply random IR perturbations in cold outflow regions not necessarily directly associated with a recent updraft, or temperature gradients associated with an above anvil cirrus plume (Bedka et al., 2018) near to an OT.

A decayed OT can inject small cloud particles above the tropopause that appear as prominent cold cores from a satellite perspective but are dislocated from any NEXRAD OT at the time of comparison. For the case shown in Figure 11c, black contours in green shaded regions ($Z_H < 25 \text{ dBZ}$) indicate OTG identifications co-located with NEXRAD identified anvil. Analysis of the history of storm cells near the brown arrows in Figures 10c and 10d show these echoes above the tropopause were present prior to this image. Therefore, we believe these identifications are residual from a previous OT and not a current OT. This is observed in other cases as well, but the exact contribution of this to the FAR statistics is difficult to quantify.

Yellow arrows on Figures 10c and 10d point to OTG identified in an anvil region due to a random IR perturbation from nearby OT outflow and gravity waves. These OTG identifications are at least 5 K colder than...
the tropopause and 3–4 K colder than the surrounding anvil. Khlopenkov et al. (2021) show that ~40%–65% of NEXRAD-detected OTs (percentage range dependent on \( z_r \) magnitude) had tropopause-relative \( T_b \) warmer than this 5 K threshold. Detections of this nature are observed in other cases as well (not shown). The goal of an automated detection algorithm is to not only detect large, prominent overshoots but also be sensitive enough to detect smaller, weaker overshoots. If IR perturbations occur in very cold regions, relative to the tropopause, then the OT detection algorithm will have challenges distinguishing a random perturbation from a true OT region, especially for comparable IR-anvil BTD. This is an inherent limitation associated with IR-based OT detection.

IR satellite imagery often depicts a distinct cold “U” or “V” shaped feature, bounding a region of much warmer \( T_b \) associated with an above anvil cirrus plume, that is generated by tropopause-penetrating updrafts and often severe thunderstorms (Adler et al., 1981; Fujita, 1982; Homeyer, 2014; Negri, 1982). The U/V shaped region, also referred to here and in the literature as an “enhanced-V”, appears cold because of lift associated with the intersection of jet stream flow with the OT region that can be 3+ km above the anvil that acts as an obstacle to flow. The OT is located at the apex of the enhanced-V, and cold cloud emanates for some distance downwind, forming the “arms” of the U/V. Clusters of cold pixels in the enhanced-V can have enhanced prominence relative to the anvil due the presence of the warm plume. Figure 14 shows OTG behavior in conjunction with several enhanced-Vs, denoted by yellow arrows in Figure 14c. These enhanced-Vs are evident in Figure 14a via \( \Delta T_{MERRA} \) < 0 (green, red, and magenta shading) bounding an area of \( \Delta T_{MERRA} > 0 \), associated with above anvil cirrus plumes (blue shading). Because enhanced-Vs are often generated by wide tropopause-penetrating updrafts (Bedka et al., 2018), most of the OTG detected in this case, including those along the V “arms,” are matched with an OTN, as shown by the intersection of black and blue contours in Figure 14c. OTG further away from the updraft regions and downwind of the arms (brown arrows) are associated with cold cloud that was once within the arms that has still retained cold \( T_b \). Such OTG detections along the arm regions are not truly associated with updraft cores though, and if the updrafts nearby were not as wide as those here, they would be considered false detections in our statistical analysis. Given the cold \( T_b \) and spatial gradients within enhanced-V producing storms, false detections are inevitable and represent another inherent limitation associated with IR-based OT detection.

Another source of IR-based OT detection error is detection of updrafts with cloud tops colder than the tropopause but echo top below it (in altitude). Figure 15 shows an example case where the yellow arrows in Figure 15c point to regions in which OTG are identified but no OTN are detected. These are heavy precipitation regions, as shown in Figure 15b. Figure 15d shows radar echo ≥20 dBZ measured within 2 km of \( Z_{MERRA} \), which meets the possible-OTN criteria.

Therefore, these are not necessarily false detections, but the echo may not be observed above the tropopause. Similar detections were observed most frequently in the Southeast US where the tropopause is often colder and higher than in the Central Plains, evident via the differences in Figures 5a and 5b.
While it is assumed that OTs are cold in IR satellite imagery, this is not always what we see in reality, as evidenced by the example shown in Figure 16. This is a case where an OTN is identified that is missed by the GOES algorithm because $\Delta T_{\text{MERRA}^2} > 0$. The yellow arrow in Figure 16a points to a narrow OTN ~1 km above the tropopause but ~2 K warmer than the tropopause. The narrow OT may not be adequately filling a GOES IR pixel, resulting in warm $T_b$. This OT also rapidly develops and decays. In the initial stages of development, ice concentration near storm top is not as dense as later in the storm lifetime, so it is possible radiation emanates from deeper within the cloud where it is warmer (Sherwood et al., 2004). Particles that are large, but not dense could produce a signal observable by NEXRAD but not GOES. As the OT develops, the cloud top $T_b$ gradually becomes colder as the ice density and optical depth increases, so that radiation emanates from higher in the cloud where it is colder (Sherwood et al., 2004), as was discussed above in relation to Figures S3 and S4. Additional analysis (not shown) indicates that narrow OT regions, comprised of <25 NEXRAD -2 km possible-OTN pixels, were only detected 33% (18%) of the time from GOES-16 (GOES-13). Narrow OTs may not sufficiently fill a GOES IR pixel, thereby inhibiting detection. Detection rate nearly linearly increases to 85% (73%) for wide OT regions that exceed 75 pixels. Another explanation for warm OT $T_b$ is subsidence, where an OT collapses to near anvil level as the updraft near cloud top briefly decays. It is difficult to determine the exact cause without additional datasets that provide more cloud top information. Such behavior was noted in other cases as well (not shown) which contributes to reducing the POD shown in Figure 7.

4. Summary and Conclusions

This study compares OT detections from GOES-13 and -16 with tropopause-relative precipitation echo-top heights derived from the NEXRAD GridRad data set across 15 days and a broad area of the contiguous US with widespread and long-lived convection during 2017. Satellite OT detection validation, such as that described by Khlopenkov et al. (2021), has typically focused on comparison with OT identifications by a human analyst which limits the sample size and diversity of storm types/intensity. Comparison of satellite detections with a stable reference like the NEXRAD GridRad data set enables a more comprehensive characterization of detection performance and provides new insights into the strengths and limitations of IR-based detection methods. Such characterization is important given increasing weather and climate community interest in OT detection methods from geostationary satellites.

NEXRAD-based OT detections are a reliable data source for validating a large sample of satellite-based detections but there are many challenges associated with comparing OTs identified from NEXRAD and GOES. NEXRAD radars have a limited sensitivity to small particles, so the echo-top altitude is generally below the cloud top altitude. In addition, reanalysis estimates of the tropopause altitude can have errors as large as the model vertical resolution. Errors in echo-top and tropopause altitude assignment can impact whether NEXRAD radars observe an OT or not. In addition, there is uncertainty in determining the altitude where IR radiation emanates from within a deep convective cloud top which adds to the challenges associated with comparing the two datasets. In order to provide a comprehensive overview of GOES satellite OT detections, this paper characterizes GOES performance using a set of NEXRAD tropopause-relative echo-top height metrics; 10 dBZ echo tops that penetrate the tropopause (OTN) and 20 dBZ echo tops 1–2 km below tropopause (possible-OTN). The two
possible-OTN criterion help illustrate the OTG detection performance, however, not every possible-OTN has a cloud top above the tropopause. Due to NEXRAD and GOES instrument limitations, quantifying exactly how many possible-OTN have cloud tops above the tropopause is impossible with the current set of existing sensors. Thus, the possible-OTN provides a “best-case scenario” that allows us to better estimate GOES OT detection performance.

It was found that GOES OT detections were often well correlated with high altitude echo tops, but that OT-like perturbations in the GOES IR $T_b$ can be generated by echo tops below the tropopause level. For example, if only tropopause-penetrating echoes are considered truth, ~49% of GOES detections with OT probability $\geq 0.5$ (OTG) would be false alarms. The false alarm ratio drops to 12% if 20 dBZ echoes 2 km beneath the tropopause are the target, however, these lower altitude echo tops are not all part of updrafts with cloud tops above the tropopause. Approximately 76% (45%) of tropopause-penetrating echoes ($z > -2$ km) were detected. The coarser GOES-13 pixel size resulted in a ~15% drop in detection rate, despite efforts to account for the ~2 K mean $T_b$ difference relative to GOES-16 within OT regions in the detection algorithm formulation. Despite this drop in detection rate, echo top and composite reflectivity distributions were quite comparable between the two satellites, suggesting that the normalization was as effective as possible.

Case study examples revealed that IR temperature perturbations within convective cloud tops are reliably detected provided that they are cold enough relative to the tropopause and distinct relative to the surrounding anvil. Such perturbations, however, are not always correlated with updraft cores near to or penetrating the tropopause observed by GridRad. Disagreement between GOES and NEXRAD detections was found to occur when (a) a cluster of cold pixels retains its temperature as it is advected away from the parent OT region, (b) cold cloud generated along the arms of an enhanced-V signature is incorrectly detected as an updraft, and (c) narrow/developing NEXRAD OTs do not generate a prominent IR $T_b$ or IR-anvil BTD signal. NEXRAD and GOES OT detections aggregated across entire events showed good agreement though, indicating that the GOES detection method, in general, performed well.

These results coupled with those from Khlopenkov et al. (2021) demonstrate that tropopause-relative IR $T_b$ and spatial gradients can be used to detect features that have similar patterns to OTs that are observed by humans and precipitation radars. OTs occur around the globe and are often associated with some of the most extreme weather on Earth. Therefore, in addition to extending the algorithm to past satellites, future work should focus on replicating these IR-based detection techniques to other geostationary satellites such as FY-4, GEO-KOMPSAT-2, and the Meteosat Third Generation. Detections could be validated with other national weather radar networks using approaches described in this paper, provided that the radars can observe convective cloud tops and resolve OT regions with ~1 km or better precision. Such IR-based proxies are imperfect, however, and false detections are unavoidable, especially with coarser-resolution satellites without including other datasets that confine detections to updraft regions. As was demonstrated with MODIS by Bedka and Khlopenkov (2016), future satellite OT detection algorithms could be improved by including visible channel texture in conjunction with the IR detections. Additional IR channels could perhaps be included as well, in addition to GOES-16/17 Geostationary Lightning Mapper flash extent density (Cintineo et al., 2020). Neither the visible texture nor lightning based updraft proxies are perfect, and their performance varies throughout the day and as a function of cloud optical depth (Fuchs et al., 2016). Thus, determining how each channel or flash detection can be properly weighted to detect OTs at the satellite pixel scale is still an open question. Future studies should consider utilizing segmentation-based neural network machine learning approaches to determine the importance of each variable in order to maximize OT detection capabilities at the satellite pixel scale (Everingham et al., 2015).

Data Availability Statement

Radar data were downloaded from the NOAA National Centers for Environmental Information (http://www.ncdc.noaa.gov). The NEXRAD data files were translated to netCDF format using the NOAA Weather and Climate Toolkit (https://www.ncdc.noaa.gov/wct/). ECMWF ERA-Interim reanalysis data were downloaded from the National Center for Atmospheric Research Data Archive (NCAR RDA; http://rda.ucar.edu/datasets/ds627.0/). GOES OT detection and human-identified OT region databases are available upon request from NASA Langley Research Center. MERRA-2 data was provided by the NASA Global Modeling and Assimilation
Acknowledgments

Funding for this work was provided by the National Science Foundation through grants AGS-1522906 and AGS-1522910 to Texas A&M University and the University of Oklahoma, respectively, the NASA ROSES award 18-DISASTER18-0008, and support from the NASA Earth Venture Suborbital DCOTSS mission. John Cooney’s research was supported by an appointment to the NASA Postdoctoral Program at NASA Langley Research Center, administered by Universities Space Research Association under contract with NASA. The authors would like to thank Dr. Cameron Homeyer for his guidance with interpreting GridRad data within deep convection.

References


