

A Probabilistic Multispectral Pattern Recognition Method for Detection of
Overshooting Cloud Tops Using Passive Satellite Imager Observations

Author(s): Kristopher M. Bedka and Konstantin Khlopenkov

Source: *Journal of Applied Meteorology and Climatology*, Vol. 55, No. 9 (September 2016), pp. 1983-2005

Published by: American Meteorological Society

Stable URL: <https://www.jstor.org/stable/10.2307/26179814>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/10.2307/26179814?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



American Meteorological Society is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Applied Meteorology and Climatology*

JSTOR

A Probabilistic Multispectral Pattern Recognition Method for Detection of Overshooting Cloud Tops Using Passive Satellite Imager Observations[©]

KRISTOPHER M. BEDKA

NASA Langley Research Center, Hampton, Virginia

KONSTANTIN KHLOOPENKOV

Science Systems and Applications, Inc., Hampton, Virginia

(Manuscript received 15 September 2015, in final form 26 March 2016)

ABSTRACT

Deep convective updrafts often penetrate through the surrounding cirrus anvil and into the lower stratosphere. Cross-tropopause transport of ice, water vapor, and chemicals occurs within these “overshooting tops” (OTs) along with a variety of hazardous weather conditions. OTs are readily apparent in satellite imagery, and, given the importance of OTs for weather and climate, a number of automated satellite-based detection methods have been developed. Some of these methods have proven to be relatively reliable, and their products are used in diverse Earth science applications. Nevertheless, analysis of these methods and feedback from product users indicate that use of fixed infrared temperature–based detection criteria often induces biases that can limit their utility for weather and climate analysis. This paper describes a new multispectral OT detection approach that improves upon those previously developed by minimizing use of fixed criteria and incorporating pattern recognition analyses to arrive at an OT probability product. The product is developed and validated using OT and non-OT anvil regions identified by a human within MODIS imagery. The product offered high skill for discriminating between OTs and anvils and matched 69% of human OT identifications for a particular probability threshold with a false-detection rate of 18%, outperforming previously existing methods. The false-detection rate drops to 1% when OT-induced texture detected within visible imagery is used to constrain the IR-based OT probability product. The OT probability product is also shown to improve severe-storm detection over the United States by 20% relative to the best existing method.

1. Introduction and background

Overshooting cloud tops (OTs) indicate the presence of a deep convective updraft with sufficient strength to penetrate through the local level of neutral buoyancy (LNB) and into the lower stratosphere. OTs can transport tropospheric aerosols, chemical species, water vapor, and ice into the lower stratosphere, each of which has a significant impact on Earth’s climate system [see Kirk-Davidoff et al. (1999) and Shindell (2001) for examples]. Thunderstorms with OTs also frequently

produce hazardous weather such as heavy rainfall, lightning, aviation turbulence, damaging winds, large hail, and tornadoes, and these hazards are typically concentrated near OT regions (Reynolds 1980; Negri and Adler 1981; Brunner et al. 2007; Bedka et al. 2010; Dworak et al. 2012).

Satellite observations of convective storms show that OTs have a consistent appearance in images of the visible (VIS; $\sim 0.65 \mu\text{m}$) and longwave infrared window (IRW; $\sim 11 \mu\text{m}$) channels. OTs most often appear as isolated regions of cold IRW brightness temperatures (BTs) relative to the warmer surrounding anvil cloud. Anomalously cold BTs are caused by persistent moist-adiabatic ascent, allowing the BT to be much colder than any temperature found in a collocated sounding. Satellite observations and numerical-model simulations have suggested that some long-lived OTs can actually be warmer than the surrounding anvil, in part because mixing between the OT and the stratosphere affects

[©] Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JAMC-D-15-0249.s1>.

Corresponding author address: Kristopher Bedka, NASA Langley Research Center, Mail Stop 420, Hampton, VA 23681.
E-mail: kristopher.m.bedka@nasa.gov

DOI: 10.1175/JAMC-D-15-0249.1

© 2016 American Meteorological Society

IRW BT within shorter-lived OTs (Setvák et al. 2010). As a result of the turbulent motions within OT updrafts and the fact that they can penetrate more than 2 km above the surrounding cirrus anvil cloud (Homeyer and Kumjian 2015; Bedka et al. 2015), OTs exhibit a “cauliflower like” texture in VIS imagery and produce shadows upon the anvil when illuminated at off-nadir angles.

In recent years, there has been increased interest in the development of automated satellite-derived OT detection algorithms and the use of their output for weather and climate analyses. The GOES-R Aviation Algorithm Working Group was given the task of developing an objective OT and enhanced-V-signature detection algorithm for use with GOES-R Advanced Baseline Imager data [Bedka et al. 2010 (hereinafter B2010), 2011]. The B2010 algorithm focuses on detection of small and distinct IRW BT minima within anvil clouds that are colder than the tropopause that is determined by a numerical model. Several other satellite-based OT detection approaches have also been developed over time. The most commonly used BT difference (BTD) for OT detection is the $\sim 6.5\text{-}\mu\text{m}$ water vapor (WV) channel minus the IRW channel (Schmetz et al. 1997). The presence of stratospheric WV injection by OTs is thought to correlate with a positive WV–IRW BTD. Comparisons of the WV–IRW BTD product with CloudSat Cloud Profiling Radar observations indicate that the BTD is effective for detection of deep convection with a vertical depth often exceeding 12 km (Young et al. 2012). Many other studies have shown that the BTD product often cannot differentiate OT regions from convective anvils and does not offer accuracy that is comparable to the B2010 approach. (Bedka et al. 2010, 2012; Setvák et al. 2013). Another recent approach pairs the WV–IRW BTD with several other BTD combinations (Mikuš and Strelec Mahović 2013). It is important to note that none of these IR-based methods can detect the long-lived and anomalously warm OT regions that were noted by Setvák et al. (2010) because they rely upon the assumption that OTs are always colder than the surrounding cloud. To date, only one approach (Berendes et al. 2008) has attempted to use the VIS for OT detection, despite the fact that a human can identify an OT very easily in VIS imagery. Variable solar illumination throughout the day causes OT appearance to vary significantly, posing a challenge to automated detection algorithms.

The B2010 product is being used increasingly for weather analysis and forecasting applications. It is produced using current GOES 4-km IRW imagery and transmitted in real time to NOAA for evaluation within the GOES-R Proving Ground (Goodman et al.

2012; Gravelle et al. 2016). The product has provided the most value to operational settings 1) at night when VIS imagery is unavailable for indicating hazardous storm updrafts, 2) for recognition of hazardous updrafts in regions that are inadequately sampled by weather radars (e.g., mountainous terrain or oceanic waters), and 3) for recognition of storm intensification and decay prior to indicators in radar data (W. Line and M. Folmer, GOES-R Proving Ground satellite liaisons, 2015, personal communication). OTs were detected near $\sim 45\%$ of severe storms over the eastern United States and Europe from 2004 to 2009 using the method of B2010 (Bedka 2011; Dworak et al. 2012). Analysis of OT detections generated using GOES 1-min super rapid scan operation (SRSO) data indicates that OTs can precede severe weather events by 30+ min and are typically collocated with maxima in total lightning activity, providing forecasters with valuable situational awareness and forecast lead time (Dworak et al. 2012; Bedka et al. 2015). Imagery with temporal resolution ranging from 30 s to 2.5 min will be routinely collected by the next generation of geostationary (GEO) imagers such as the GOES-R Advanced Baseline Imager (Schmit et al. 2005) and the Japanese Advanced Himawari Imager (Bessho et al. 2016), enabling the operational forecasting community to obtain maximum benefit from OT detection and other satellite-based products for detection of convective weather (Goodman et al. 2012).

Long-term regional GEO-based OT detection databases have been developed using B2010, providing new applications and opportunities for understanding the spatial and diurnal distribution of hazardous storm activity (B2010; Bedka 2011; Proud 2015; Thiery et al. 2016). Using Meteosat Second Generation imagery, the B2010 product identified a distinct nighttime OT maximum over each of the African Great Lakes, including Lake Victoria, which has been characterized as the “world’s most dangerous lake” (CNN 2013; see also Semazzi et al. 2011). Over 5000 people, commonly subsistence fisherman, die every year on Lake Victoria—often as a result of intense thunderstorms. A 9-year European OT detection database has been combined with ground-based reports of hail and data from numerical weather analysis to develop a European hail-risk model for use within the reinsurance industry (Punge et al. 2014). This model is designed to determine the risk of severe hail at any location over Europe, allowing decisions on insurance pricing to be based on strong scientific evidence rather than on studies done on a per-country basis with inconsistent methods. Model development would not have been possible without the OT detection information because no other long-term

pan-European data product exists that can identify hazardous storms at the individual storm-cell scale.

A key question that has yet to be answered is: Has hazardous storm activity changed in response to observed climate change during the weather-satellite era? Polar-orbiting imagers such as the Moderate Resolution Imaging Spectroradiometer (MODIS) observe at fixed times during the day that are not aligned with the timing of peak storm frequency over land, which typically occurs near 1600–1800 local time (Yamamoto et al. 2008; Bedka 2011). A climatological description that is based on long-term GEO-based OT detection extending back to the mid-1990s when GEO imagers began collecting IRW data at high spatial resolution (≤ 5 km per pixel at nadir) would help to answer this question better, but current OT detection algorithms such as B2010 have a key limitation in that they rely upon fixed detection criteria that can cause seasonal and regional biases and thereby limit their utility for analysis of climate trends (Dworak et al. 2012). For example, warm OTs common to midlatitude regions during the cold season or errant tropopause analyses inhibit detection, skewing trends and potentially masking climate-change-induced impacts on hazardous-storm distribution and frequency.

This paper describes a new multispectral (VIS and IRW) OT detection approach that seeks to minimize use of fixed detection criteria and emulates the process used by a human to identify deep-convective anvils and embedded OT regions. The net result is an OT probability at the individual storm-cell scale, enabling an improved capability to identify and forecast hazardous storms in real time and to derive long-term trends in hazardous-storm activity. The probabilistic nature of the algorithm and the comprehensive validation presented in this paper allow users to optimize the product so as best to meet their needs with an understanding of detection accuracy for a particular probability threshold—a capability that has not been provided by any other algorithm to date. A simple use case for the algorithm—detection of severe storms—is also presented.

2. Data

a. MODIS and GOES imagery

Aqua MODIS 0.65- μm VIS reflectance and 11- μm IRW BT satellite observations are the primary datasets used for development and validation of the OT detection algorithm. A set of one hundred 5-min MODIS granules was manually analyzed by an independent human expert to determine the locations of OT signatures using the method described below in section 2b. MODIS VIS and IRW images were respectively kept at

the 0.25- and 1-km resolutions to provide the best possible data for human OT identification. These granules were selected to capture a diversity of OT-producing storms in terms of storm location, intensity, and size from a combination of 1) scenes with *CloudSat*-observed OTs (Bedka et al. 2012), 2) MODIS-observed severe-storm events identified via Internet searches, and 3) randomly selected scenes identified within NASA MODIS “quick look” imagery (<http://ladsweb.nascom.nasa.gov>). It is assumed that including storms across diverse regions and seasons encompasses the spectrum of storm size and intensity that could be present across the globe at any time. Subsets of IRW and VIS imagery from four granules with the same color enhancements are shown in Figs. 1 and 2. These examples illustrate the variability in IRW BT and overall storm characteristics that can be very challenging to an automated OT detection algorithm.

MODIS 1-km IRW BT is resampled to a 4-km resolution for input into the OT detection algorithm to approximate the nadir spatial resolution of many current and historical GEO imagers. This resampling allows the algorithm to be demonstrated in current forecast operations and to be used to generate climate-data records of hazardous-storm events using historical satellite data. Prior to resampling, MODIS swath data are remapped to an equal-area projection (Khlopenkov and Trishchenko 2008) to remove the so-called bow-tie effect that is present along the limb of the MODIS swath (<http://mcst.gsfc.nasa.gov/forums/images-seem-distorted-edges-scans-why-and-what-can-be-done>). The resampling process typically preserves the BT signal associated with OTs given that most exhibit a clear BT minimum that extends across several 4-km pixels. Resampling also dampens small-scale BT variability in convective anvils that could confuse the pattern-recognition scheme and lead to false detection. MODIS 1-km-resolution VIS imagery is also input to the detection algorithm, which also matches the resolution of current and historical GEO VIS data.

All resampling of imagery described in this paper is implemented by using a resampling function over a window of 6×6 pixels. At the image boundaries, the missing content of the 6×6 window is padded by replicating values of the edge pixels. The resampling function is based on Lanczos filtering (Duchon 1979) with the parameter $a = 3$ extended to the two-dimensional case. This interpolation method is based on the sinc filter, which is known to be an optimal reconstruction filter for band-limited signals such as digital satellite imagery.

GOES-14 4-km IRW BT and 1-km VIS data were used to generate OT detections for 25 May 2015 over a 10-h period to demonstrate algorithm performance

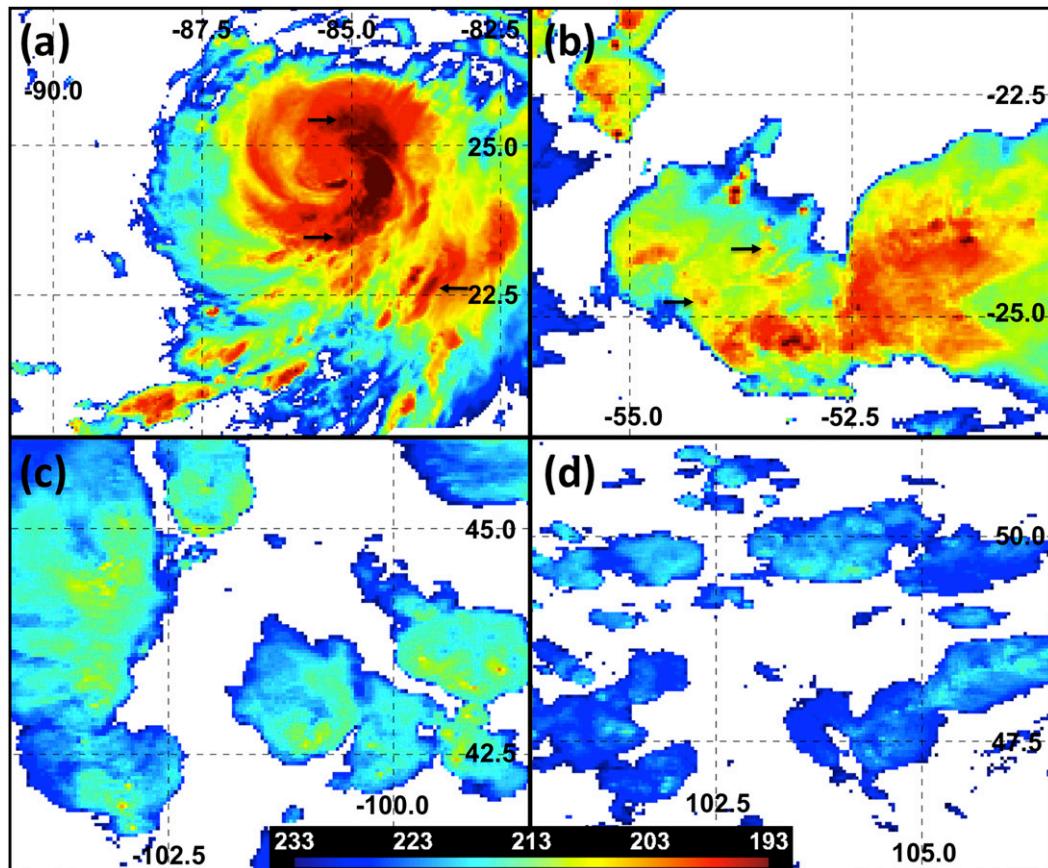


FIG. 1. MODIS 1-km IRW BT imagery (K) of (a) Hurricane Katrina over the southeastern Gulf of Mexico at 1925 UTC 27 Aug 2005, (b) a large mesoscale convective system over southern Brazil at 1735 UTC 8 Mar 2009, (c) tornado- and hail-producing storms over the U.S. central plains at 2005 UTC 24 May 2010, and (d) small storms over Mongolia at 0605 UTC 8 Jul 2012. Pixels with IRW BT warmer than 233 K are set to white. The black arrows in (a) and (b) show IRW BT minima that appear similar to other OT regions but that do not have any signal of an OT in VIS imagery.

across varying conditions of solar illumination that cannot be demonstrated with sun-synchronous imagers like MODIS. GOES-14 collected data in 1-min SRSO mode for this case, which featured a large complex of storms that produced flooding and severe weather across Texas and Oklahoma.

b. Database for training and validation of the OT detection algorithm

Algorithm development and validation required compilation of a database of regions corresponding to OTs and also regions within anvils that were not overshooting (i.e., “non-OT regions”). The algorithm uses differences in the satellite-observed characteristics of OT and non-OT regions and the relationships between IRW BT and the NASA Modern Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011) data to assign an OT probability

using the logistic regression described below in section 3a(3). MODIS 0.25-km VIS imagery was manually analyzed by a human expert to identify all OTs present in each of the 100 granules, 67 of which were used for algorithm training and 33 of which were used for validation. OTs were identified by their characteristic texture and shadowing cast upon the surrounding anvil cloud; see Setvák et al. (2013) for detailed examples of OT signatures in MODIS imagery. For questionable OT features, 1-km IRW BT imagery was examined to ensure that they had a BT that was consistent with nearby anvils. A map of human OT identifications within the 100 granules is shown in Fig. 3. Images of OTs over the United States are abundant because MODIS-observed thunderstorm events are frequently publicized on Internet resources. Of the 617 OTs identified over the United States, 166 were collocated with a report of severe weather ± 15 min from the time of the MODIS

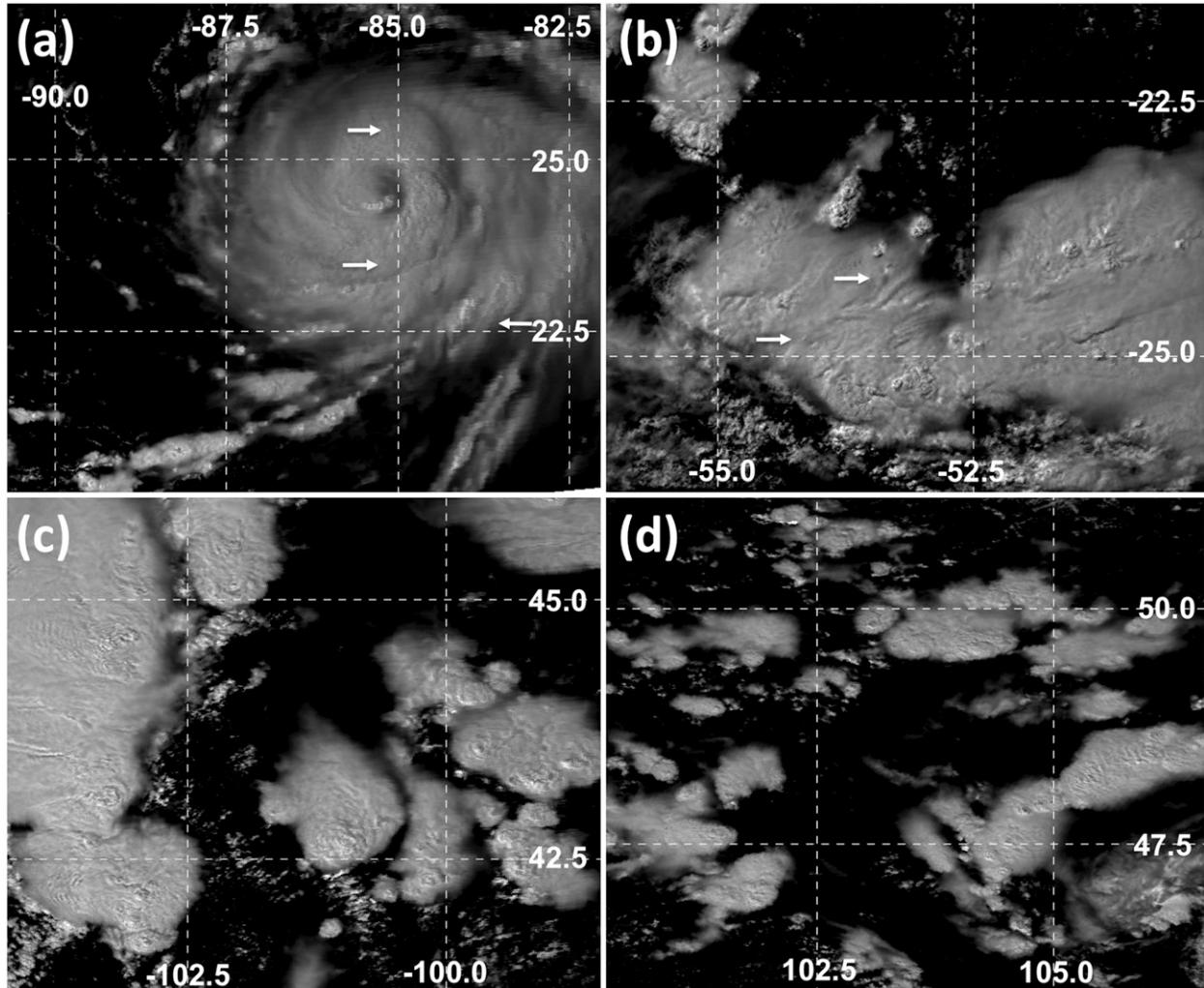


FIG. 2. MODIS VIS reflectance for the same four scenes as in Fig. 1. The white arrows in (a) and (b) show IRW BT minima that appear similar to other OT regions but that do not have any signal of an OT in VIS imagery.

image, allowing for an assessment of severe-storm detection capability (see section 3d). A combination of IRW and VIS imagery was used for non-OT database development to capture a variety of BT and reflectance patterns that are representative of anvils.

c. MERRA

MERRA reanalysis fields were used in the logistic-regression component of the algorithm. MERRA has a $0.5^\circ \times 0.66^\circ$ grid spacing and provides hourly surface analyses and 6-hourly vertical profiles at 42 vertical levels. MERRA data were interpolated to the time of the MODIS granules to represent the thermodynamic environment present near the OT and non-OT regions. Parameters were selected to assess their value for discriminating between OT and non-OT regions within a

logistic-regression framework. These parameters include the minimum IRW BT minus the LNB temperature for 1) surface-based and 2) “most unstable” parcels, 3) minimum IRW BT minus the tropopause temperature, and 4)–6) vertical wind shear in the layers at 0–1, 0–3, and 0–6 km above ground level (AGL), respectively. The minimum LNB and tropopause temperatures and maximum wind shear within 2.5° of an OT region or non-OT region were used to account for possible MERRA spatial and temporal analysis errors.

The term most unstable refers to the parcel within the layer at 0–3 km AGL that yields the greatest convective available potential energy (CAPE). The most unstable parcel is considered here because of 1) challenges in accurately capturing 3D temperature and moisture distribution, especially over data-poor regions, and 2) a

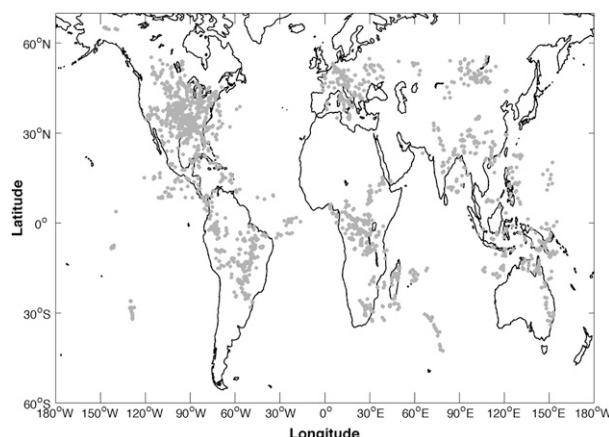


FIG. 3. A map of OT locations (gray circles) identified within 100 granules of *Aqua* MODIS imagery.

lack of a priori knowledge of exactly which height the rising parcels originated from, especially when convection is not surface based. The LNB temperature is set to 170 K for situations in which the surface-based or most unstable CAPE is 0, which serves to diminish the OT probability because it is unlikely that a detection is accurate in an environment that is not supportive of deep convection.

d. Severe-weather-report database

Severe-weather reports from the NOAA Storm Prediction Center (SPC) were used to identify MODIS-observed OT-producing storms that were severe. The severe-weather reports contain the date, time, and coordinates of tornado, large hail (≥ 1 in.; 1 in. = 2.54 cm), or severe-wind reports (≥ 50 kt; 1 kt = 0.51 m s^{-1}) in addition to the hail diameter and estimated and/or observed wind speed.

3. Method

The following section will describe the details of the IRW- and VIS-based OT pattern-recognition scheme, the logistic-regression framework used to generate OT detection probability, and the validation approach. A flowchart summarizing the primary components of the IRW and VIS pattern recognition is provided by Fig. 4 and will serve as a guide for this discussion.

a. Description of probabilistic OT detection algorithm

The goal of the algorithm is to emulate the analysis used by a human for identifying deep convection and embedded OT signatures. Deep convection typically has a spatially coherent area of cold BT with a distinct

anvil edge, marking the transition from cloud to the warmer Earth surface. Once deep-convective clouds are identified, a human would analyze their anvils to identify OT signatures.

The pattern-recognition components of the algorithm attempt to quantify how much the region around a particular pixel “looks like” deep convection and an OT on the basis of the spatial distributions of IRW BT and VIS reflectance. A primary goal is to enable the algorithm to discriminate between deep convection and cold but “ordinary” cirrus, because these cirrus clouds can trigger false detection within the B2010 approach (Proud 2015). Through pattern recognition, OT candidates are identified and are then assigned an OT probability. The OT probability is based entirely on IRW-based data, but areas of VIS texture detected through statistical, spatial, and frequency analyses are used during the daytime to increase confidence that a given IRW-based detection is accurate.

1) OT PATTERN RECOGNITION USING IRW BT IMAGERY

The first step in the OT detection process is to identify deep-convective anvils that may contain one or more OT regions. Previous studies such as B2010 used a 215-K BT threshold to define an OT candidate and 225 K to define an anvil. The 215-K threshold is typically effective for warm-season convection over the United States but could cause 1) OT under-detection during the cold season or in environments with warm cloud tops resulting from a low tropopause and 2) over-detection in tropical environments with a very high tropopause.

We attempt to improve upon previous efforts by transforming IRW imagery from BT into a normalized “score” parameter (hereinafter referred to as S) to create a dynamic approach for identification of deep convection. The S parameter has two primary terms that were derived through empirical analysis of many MODIS scenes:

$$S = [(BT_{\text{winavg}} - BT)^{0.7}(255 - BT)^{1.3}/16] + [2\sigma(BT_{\text{win}})]. \quad (1)$$

The first term of S has two components. The first component determines the difference between a pixel IRW BT (denoted as BT) and the mean BT within a large window surrounding the pixel (BT_{winavg}). This is done to normalize the pixel BT relative to the observed scene and to determine an initial likelihood that the pixel is within deep convection. The window size is 400×400 IRW pixels, which is typically large enough

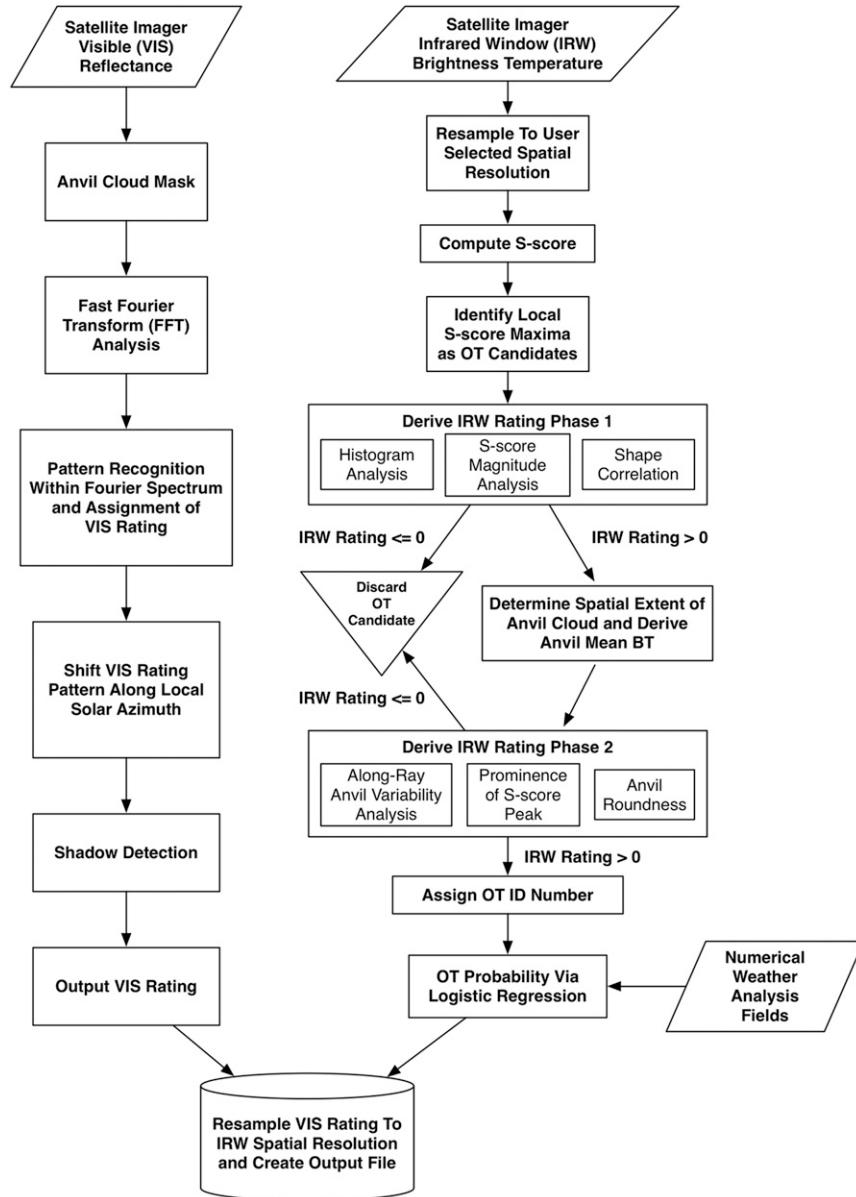


FIG. 4. A flowchart describing the IRW and VIS components of the probabilistic OT detection algorithm.

to encompass a large mesoscale convective system or tropical cyclone as well as some of the surrounding clear-sky area. The second component quantifies how cold a pixel is on the basis of the IRW BT difference from 255 K, because OTs and anvil clouds are typically much colder than this threshold. We weighted each component differently with exponentials because the pixel BT was found to be more important than the difference from the window mean. The two components are multiplied, which forces both to have high values to achieve a high S . The second term

incorporates the standard deviation σ of the window BT, or BT_{win} . The $\sigma(BT_{win})$ would be large and would further enhance S in windows in which a combination of deep convection and clear sky is present. An $S \geq 90$, a threshold found to work well from empirical testing, is required for a pixel to advance for further processing.

The obtained S field is used to detect local S maxima within the image. The process is similar to the automated detection of ground control points from image variance that is described by Khlopenkov and Trishchenko (2010). The routine ensures that detected

maxima have appropriate spacing, which is larger for small S and smaller for high S . The variation in spacing as a function of S attempts to preserve the more significant and colder regions but enables spatial filtering to account for the fact that OTs are typically less than 15 km in diameter (B2010). These maxima become initial OT candidates that are subsequently processed within pattern-recognition analyses.

The pattern-recognition analyses each produce an arbitrary rating, which are accumulated over all analyses to form a cumulative IRW rating. This IRW rating indicates the degree of confidence that a particular pixel is within deep convection and has characteristics similar to an OT. Such a cumulative approach reduces the sensitivity to failure of any single analysis, which helps the overall stability of the algorithm (as compared with a decision-branching tree) but also allows a user to customize the end product on the basis of their confidence and accuracy requirements. The series of analyses is implemented in two phases and includes histogram analysis, an S -magnitude test, an OT-shape test, an analysis of BT variability within the anvil cloud, the prominence of the OT relative to the anvil (i.e., the degree to which the OT is colder than the anvil), and an anvil-roundness analysis (see Fig. 4). These tests/analyses will be explained in as much detail as possible, but the reader should understand that the algorithms and their mathematical formulation cannot be explicitly presented in this paper.

The fundamental goals of and challenges for the following analyses are to automatically estimate the spatial extent of the anvil cloud without a priori knowledge of its average BT and then to find prominent cold regions within the anvil that are characteristic of an OT. Once the boundary of the anvil is determined, the process of estimating its BT variability and finding prominent embedded cold regions is relatively straightforward.

The first step in defining the anvil extent, which is the histogram analysis, is a simple evaluation of the S variance $\sigma(S)$ around each candidate pixel found above. The histogram is built within an octagon-shaped window of 7×7 pixels centered on a candidate pixel. The octagon shape was selected to approximate the typical shape of a convective cloud. There can be multiple peaks in the histogram, but the peak at the highest S would correspond to anvil pixels, and thus the $\sigma(S)$ is derived only for pixels that belong to that peak. A low $\sigma(S)$ would correspond to a spatially uniform anvil cloud that is characteristic of an OT-producing storm and would increase the IRW rating. A high $\sigma(S)$ would serve to decrease the rating. The peak in the S histogram becomes the first estimate of the anvil mean S .

The next analysis determines that the candidate and neighboring pixels resemble an idealized protrusion through the anvil that is characteristic of an OT region. This resemblance is quantified through the correlation between the S region and an idealized 5×5 pixel OT region that has a distinct S maximum in the middle that tapers off toward its edge. The OT regions could be circular or somewhat elliptical, and therefore two idealized regions with these shapes were created to compute correlations with the S field. The elliptical shape is tested four times with different orientations at 45° increments. The highest among the five correlation estimates is used as the basis for adjusting the IRW rating. The final analysis in phase 1 (see Fig. 4) simply examines the magnitude of the S value. A higher S indicates a greater likelihood that a pixel is an OT, and so the adjustment to the IRW rating is proportional to the S value.

Before proceeding to phase 2, the algorithm determines the spatial extent of the anvil cloud surrounding the OT candidate (Fig. 5). To accomplish this, 16 rays are directed outward from the OT candidate with equal angular spacing, each ray having a maximum length of 30 pixels. The S profile along each ray is analyzed to fulfill the following requirements: a rapid initial drop from the maximum (label A in Fig. 5), significant positive inflection (B), low deviation from the anvil mean (C), and then a steep drop (D) from the anvil mean. The distance from the OT candidate to the point of the steep drop (D) defines the extent of the anvil region in each of the 16 directions. This process is implemented in two passes: the first uses the anvil mean obtained from the histogram analysis above, and the second uses a more accurate anvil mean calculated within the 16-sided polygon found in the first pass. The second pass yields the final shape of the anvil region and allows for a better estimation of the anvil mean S score and anvil mean BT. The characteristics described in A–D above are quantified within the phase-2 along-ray anvil-variability analysis (Fig. 4) to reduce the IRW rating for OT candidates within regions that do not resemble typical anvils.

Knowledge of the anvil extent serves as the basis for the peak-prominence test in phase 2, which computes the S difference between the OT candidate and the anvil mean. A large difference implies that the OT candidate has a cold BT surrounded by a warmer anvil. A small difference could correspond to a weak OT feature, a region of cold anvil cloud from a recently decayed OT, or simply spatial BT variability that is common within anvils. The IRW rating adjustment is proportional to the calculated difference. Because this prominence parameter 1) is a key discriminator between OT and non-OT

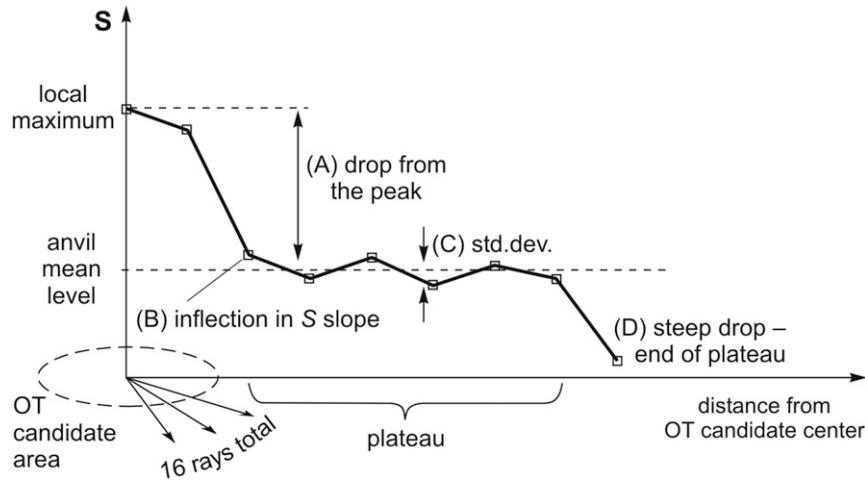


FIG. 5. A schematic diagram showing analysis of IRW-based anvil-cloud characteristics along rays extending from OT candidate regions.

regions and 2) can be used to accurately assign a cloud-top height to the OT (Griffin et al. 2016), the magnitude of the difference between the OT candidate and the anvil S is transformed back into BT space for use within the logistic-regression analysis. The final test in phase 2 estimates the “roundness” of the 16-sided anvil polygon, which is calculated as the ratio of its area to its circumference squared. A round anvil is typical of deep convection, and therefore a higher roundness value increases the IRW rating.

At this point, all IRW-based analyses are complete and an OT probability can be derived for OT candidates with a positive IRW rating. The IRW-rating field for all OT candidates for the Brazil scene (Figs. 1b and 2b) is shown in Fig. 6a. A map of all anvil-cloud detections for the OT candidates from the Brazil scene is shown in Fig. 6b.

2) VISIBLE-CHANNEL TEXTURE DETECTION

Detection of the OT signature within VIS imagery is done using a combination of statistical, spatial, and frequency analyses as summarized in Fig. 4. Locations of highly reflective pixels associated with anvils are first identified to limit the frequency analysis to regions where OTs are likely. An anvil is typically uniformly reflective, but use of a fixed reflectance threshold for anvil detection was found to be ineffective, and so a more sophisticated approach was developed. For this VIS-based anvil-cloud mask, the VIS image resolution is first degraded by a factor of 2 to improve efficiency and to suppress “noise” associated with singular bright pixels that are not likely to be part of an anvil; such an image is hereinafter referred to as I2X. One might assume that this anvil-cloud mask duplicates the anvils

identified in the IRW-based OT detection (Fig. 6b), but we seek to make the VIS and IRW components entirely independent so that they are unbiased.

A VIS reflectance histogram is calculated within an octagon-shaped window with a maximum diameter of ~ 12 km (Fig. 7). The octagon center is placed at every other pixel and line of the image for performance reasons, resulting in an image that is 4 times as coarse as the original image (hereinafter referred to as I4X). The pixel values of I4X are calculated from the histogram maximum prominence, which is defined as the ratio of the maximum bin count to the average count from the bins $i_{\text{center}-2}$ and $i_{\text{center}+3}$. The term “center” refers to the histogram bin corresponding to the nominal anvil reflectance as a function of the local solar zenith angle (SZA). The nominal anvil reflectance was derived from an empirical analysis of GOES-observed anvil clouds for a range of SZAs, which yielded the following function, with θ being the back-scattering angle:

$$C = \cos(\theta) + 1, \quad F = \frac{C^4 + 8C + 170}{202}, \quad \text{and}$$

$$\text{nominal anvil reflectance} = 0.42F(3.5 + \cos\text{SZA})^{1/2}. \quad (2)$$

Use of the range from $i_{\text{center}-2}$ to $i_{\text{center}+3}$ allows the histogram maximum to be broader on the brighter side of the distribution, which is often the case if an anvil or OT is present. If a region were to be within an optically thick anvil, one would expect to see a large count of bright pixels in one bin and minimal counts in the other bins, resulting in a higher prominence of the bright-bin maximum.

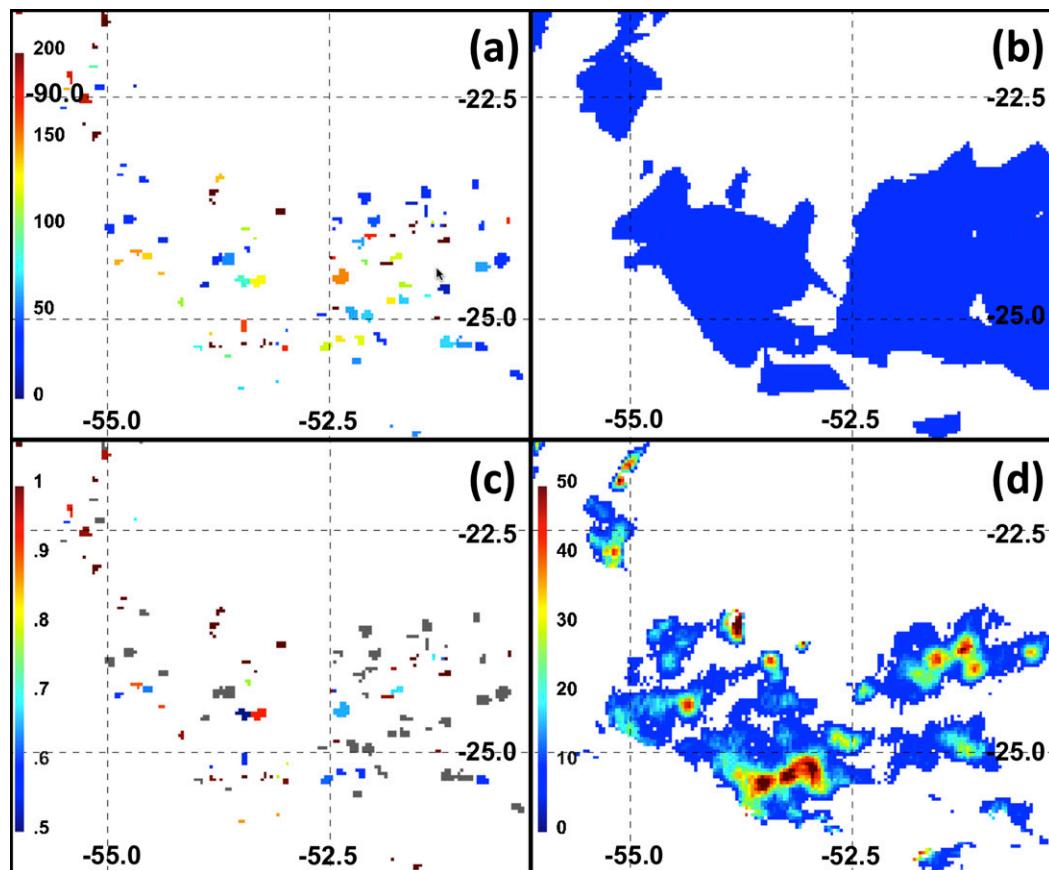


FIG. 6. (a) OT candidates colored by their IRW rating, (b) IRW-based anvil-cloud detection, (c) OT probability (values < 0.5 are colored gray), and (d) VIS rating for the Brazil scene shown in Figs. 1b and 2b.

The anvil-cloud mask is derived from I4X through a scoring system that uses an accumulation of histogram-derived information from an ensemble of nearby pixels. If the histogram maximum prominence is higher than a threshold multiplied by the cosine of the local solar zenith angle [$\cos(\text{SZA})$], then the value of 16 is stored in the image I4X. If the prominence is higher than $7 \times \cos(\text{SZA})$, then a value of 8 is stored, indicating a lower likelihood of anvil cloud. The value 8 or 16 is stored not only at the current pixel but also at 12 pixels spaced at equal angles surrounding the octagon center at a distance of ~ 16 km. As an octagon is moved throughout the image, a given pixel will be located at one of these 12 radii many times, and therefore the stored values for a given pixel are accumulated to best represent the regional scene characteristics. This allows the mask to include not only the inner areas of anvil clouds (where the octagon shape would fit completely) but also to include cloud boundaries regardless of cloud shape. After completion of the anvil-cloud mask, the image I4X is resampled back to the original 1-km VIS resolution.

Pixels with a mask < 50 are omitted from the following frequency analysis.

The frequency analysis is based on a two-dimensional spatial discrete fast Fourier transform (FFT) of the anvil reflectance. A power spectrum is calculated within a window of 32×32 pixels. The windowing is accomplished with a Gaussian function so as to gradually reduce the pixel values to zero at the window boundary. Such a windowing provides a smooth junction when the windowed data are duplicated by the FFT algorithm to form an infinite function and avoids artificial frequencies in the high spectral range.

The obtained spectrum is a 32×32 image with each pixel being the power represented in a particular spatial frequency. Ideal OT signatures and concentric gravity waves were found to produce the strongest signal in a ringlike pattern with a wavelength of $\sim 4\text{--}8$ km, and so we attempt to recognize rings in the spectrum at this wavelength interval. A copy of the FFT spectrum image is rotated relative to the origin (zero frequency) by 20° , and then each pixel of the rotated image is multiplied by the corresponding pixel of the original spectral image.

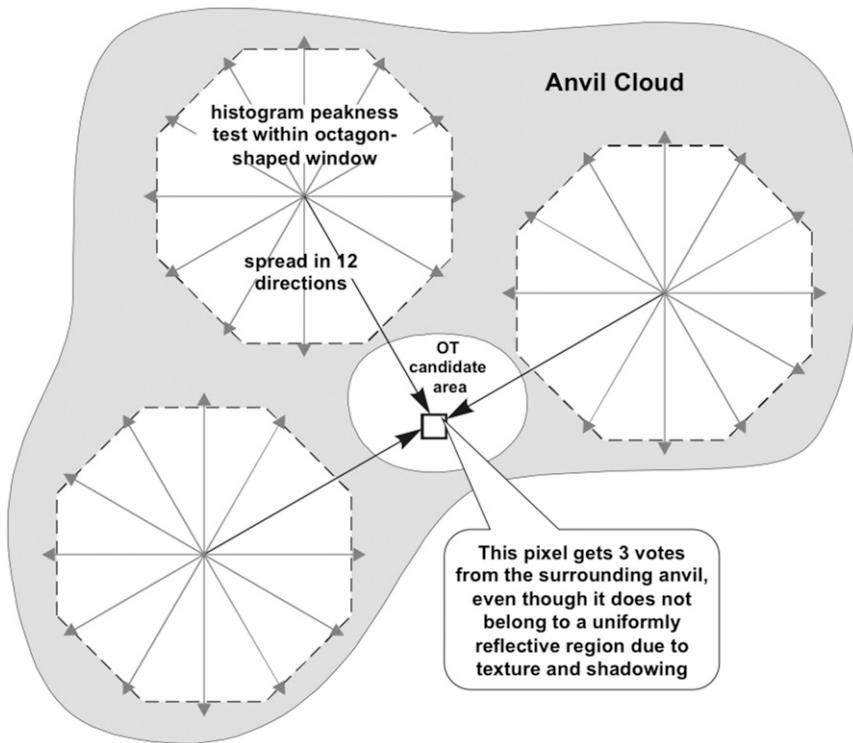


FIG. 7. A schematic diagram showing development of a VIS-based anvil-cloud mask using analysis of VIS reflectance within octagon-shaped windows.

This process suppresses any random noise but keeps all features with a ring shape that extend for at least 20° . Randomly fluctuating patterns that are typical of ordinary cirrus or other bright clouds will be suppressed by such a rotation filtering.

The spectral image is divided into four sectors of 90° each, and the total signal from the ring-detection analysis above is calculated for each sector independently. Then, the sum of the two smallest sectors is used to derive the VIS rating. The VIS rating will be high if the spectral image has sufficient circular symmetry, while allowing one of the four sectors (one term in the sum) to have weak spectral power. Again, this is designed to limit the detection to circular features like OTs and to suppress detection of other cloud types.

The FFT window may include dark cloud-free pixels if an OT area happens to be at the edge of an anvil. Because of the large difference in brightness, these pixels can cause the power spectrum to increase significantly. To correct for this effect, a sum of dark pixels is calculated within each 32×32 pixel subset of I_{2X}. Their input is weighted by an inverted Gaussian function, which makes the dark pixels at the boundary contribute the most. The VIS rating is reduced proportional to the effective sum of the dark pixels. The VIS rating is also

scaled by the $\cos(\text{SZA})^{1.5}$, which helps to equalize the rating for different illumination conditions. The resulting VIS rating is resampled back to the resolution of the original input VIS image.

As the SZA increases above $\sim 40^\circ$, the side of an OT region can appear anomalously bright, which causes the FFT analysis to produce a higher rating. This causes the enhanced VIS rating to be displaced relative to the actual OT centers. This is corrected by a morphing routine, which shifts the bright pixels in the image along the local solar azimuth. The amount of the displacement is proportional to pixel reflectance and the $\cos(\text{SZA})$.

For SZAs exceeding $\sim 60^\circ$, OT regions remain anomalously bright along their sides and top, but they also begin to produce prominent shadows on the surrounding anvil, resulting in a sharp reflectance gradient over a short distance. This not only causes the VIS rating to decrease because of the dark pixel correction described above but also can exclude an OT from FFT analysis because the shadowed area may not pass the anvil-cloud-mask test. For regions where shadows are likely, a shadow-detection component is implemented. A likely shadow region is defined where 1) the anvil-cloud-mask image exceeds 40, 2) the VIS reflectance exceeds an SZA-dependent level and that of its four

nearest neighbors, and 3) the SZA exceeds 60° . Starting from a bright pixel, the routine searches for a dark pixel in the direction opposite to the local solar azimuth, with a search distance that is proportional to the SZA. The search direction is allowed to be 10% wider than the local solar-azimuth-based prediction. If a pixel reflectance within the search region is below the original minus 0.6, it is assumed to be an OT-induced shadow. An area extending from the bright pixel to the corresponding dark pixel within the shadow is marked by a solid circle with a VIS rating that is proportional to the reflectance difference between the bright pixel and the dark pixel.

The VIS rating can be used as a mask to reduce the false-detection rate of the IRW-based OT probability algorithm. An example of the VIS rating for the Brazil scene is shown in Fig. 6d. As would be expected from the description above, regions with texture that is characteristic of an OT are assigned a high VIS rating that can exceed a value of 50 and smooth anvils are assigned a low or null rating. A VIS rating of 7 was found to be effective for discriminating OTs and gravity waves from anvils, and pixels that meet this threshold are outlined with a gray contour in Figs. 10 and 14, described in more detail below. A VIS rating of 15 highlights significant texture, typically immediately surrounding a prominent OT (see magenta contours in Figs. 10 and 14, below). Pixels with evidence of an OT via the OT probability and VIS-rating products are less likely to be errant than those without a VIS rating, a conclusion exemplified in Figs. 6c and 6d and reinforced by the validation that is described in section 4.

3) LOGISTIC REGRESSION

Logistic regression is a commonly used method for determining the probability of occurrence of an event, such as the existence of an OT, on the basis of a set of explanatory parameters. A description of logistic regression and an analogous application (hailstorm identification using a set of satellite-based parameters) is provided by Merino et al. (2014) and references therein. The logistic-regression model was constructed using a stepwise method from the MATLAB software package that evaluates a set of parameters incrementally to identify those that are statistically significant for discriminating OT regions from non-OT regions at a 95% confidence level. The parameters tested by the stepwise method should be physically related to the type of event being predicted, which in our case would be metrics of 1) convective-updraft intensity, 2) spatial BT gradients within the cloud top, 3) cloud-top height with respect to the potential maximum height dictated by the local thermodynamic

environment, and 4) vertical wind shear supportive of storm organization and increased severity.

Once a set of parameters is identified as being statistically significant on the basis of the training dataset (described in section 3b), coefficients that optimally weight each parameter and an intercept are derived. The equation takes the following form, where W_0 is the intercept, W_i is the weighting coefficient, and A_i is a satellite-based parameter:

$$\text{OT Probability} = \left\{ 1 + \exp \left[- \left(W_0 + \sum_{i=1}^n W_i A_i \right) \right] \right\}^{-1}. \quad (3)$$

When the coefficients and intercept are applied to the validation dataset, the net result is an OT probability that ranges from 0 (region is not an OT) to 1 (region is an OT). A probability threshold of 0.5 can be used to separate the OT and non-OT classes. An example of the OT probability field for the Brazil scene is shown in Fig. 6c, illustrating that the probability product effectively differentiates between anvil and OT-like regions within the OT candidate population.

b. Training and validation of the probabilistic OT detection algorithm

A number of parameters are derived for each OT region and non-OT region to be tested for statistical significance within the logistic regression. These parameters include MERRA-based fields listed in section 2d and a set of satellite-based parameters. The satellite-based parameters include 1) the minimum IRW BT within 5 km of the OT and non-OT coordinate and 2) the difference between the minimum IRW BT and the mean BT of the surrounding anvil cloud.

We derived the anvil mean BT differently than the sophisticated method described in section 3a(1) because we need an anvil mean BT computed for each OT region and non-OT region. Not all OT regions are necessarily recognized by the pattern recognition, and very few non-OT regions are recognized, and therefore much of the database would have a null anvil mean BT value. To alleviate this, we use a less sophisticated method than that described above in which we define the anvil as the set of pixels with an IRW BT ≤ 233 K located 8–15 km from the OT or non-OT coordinate. We exclude pixels in the 0–8-km range so that cold BTs within the OT region do not bias the anvil mean BT. An OT region or non-OT region is excluded from further analysis if the surrounding anvil is not large enough to compute a mean BT.

The database was separated into two groups: 67 MODIS granules (2581 total OT and non-OT regions) were randomly selected to train the logistic regression

TABLE 1. A list of satellite- and numerical-model-based parameters determined to be significant at the 99+ % level for discriminating between OT and non-OT regions using a stepwise-regression approach. The intercept and weighting coefficients used in Eq. (2) are also provided.

Statistically significant OT-region discrimination parameters and regression coef	Logistic-regression weighting coef (intercept = -2.2062)
OT-region min minus mean anvil IRW BT	-0.6946
OT-region min IRW BT minus most unstable LNB temperature	-0.0165
OT-region min IRW BT minus tropopause temperature	-0.1371

and the remaining 33 granules (1424 total OT and non-OT regions) were used for validation. Once a set of statistically significant OT predictors and regression coefficients was determined, the coefficients were used to derive OT probability for the validation database. An accuracy metric called the OT discrimination skill is computed using Eq. (4), below. An OT region can be considered to be “undetected” for two reasons: it does not have an IRW rating > 0 or its OT probability is < 0.5. The OT regions that were assigned a positive IRW rating and had a probability ≥ 0.5 are considered to be correct detections (category *A*). The OT regions that either did not have a positive IRW rating or had a probability < 0.5 are missed detections (*B*). Non-OT regions that either had a probability < 0.5 or were not assigned a positive IRW rating are correct null events (*C*). The remaining non-OT regions with a positive IRW rating and probability ≥ 0.5 are false detections (*D*). The accuracy metric is then

$$\text{OT Discrimination Skill} = \frac{A + C}{A + B + C + D}. \quad (4)$$

OT detection accuracy can also be quantified in terms of probability of OT detection (POD) and false-detection rate (FAR) derived as a function of OT probability. These statistics are presented under the assumption that the human analyst did not misidentify or fail to recognize OT regions. The number of misidentifications is probably small given that an experienced human analyst analyzed 0.25-km VIS imagery in which OTs are evident. In some situations, an OT can exhibit a prominent signal in the IRW but appear ambiguous in the VIS. The OT-probability algorithm would likely detect this feature, but it would be considered a false detection because the human analyst only recorded OTs for which she or he had high confidence on the basis of VIS analysis. Another complication is that cold regions without evidence of a VIS OT can appear to be very similar to cold regions with a VIS OT (see panels *a* and *b* of Figs. 1 and 2 for examples). If a region “looks like” an OT in IRW imagery, then the resulting OT probability will be high and the VIS rating would be required to constrain the IRW-based detection. During the night, VIS data are

unavailable, and so this high probability and false detection would be preserved.

c. Validation of the B2010 OT detection algorithm

As described in section 1, the B2010 OT detection product has been used for a variety of applications, and therefore the community needs to understand how the method described in this paper compares with B2010. The B2010 algorithm is analyzed in two different ways. The first analysis evaluates detection performance for the algorithm as described in the B2010 text. The B2010 algorithm was applied to the 33 MODIS granules in the validation database. This is referred to as “the B2010 candidates and detection criteria” in Table 3, described in more detail below. The second analysis uses the OT candidates derived from the pattern recognition described in section 3a(1) with the B2010 detection criteria. The anvil mean BT is computed using a more sophisticated approach and OT candidates are rigorously filtered here, which, in theory, should help to improve detection accuracy. This approach is referred to as the “improved candidates and B2010 detection criteria” in Tables 3 and 4, also described in more detail below.

d. Comparison of OT with the severe-weather database

Human-identified MODIS OT locations are compared with the SPC severe-weather-event database to determine the fraction of 1) confirmed OTs that were associated with severe weather and 2) severe storms that were detected by the OT probability, VIS-rating, and B2010-based methods. The OT locations over the United States were corrected for parallax to derive their true Earth-relative position. (Wang et al. 2011) An OT is associated with a severe storm if it is within ± 15 min and 25 km of a report. These match criteria are similar to those described by Bedka (2011) that take into account storm movement and severe-weather-event reporting error.

4. Results

Logistic regression determined that the three parameters in Table 1 are statistically significant at the 99+

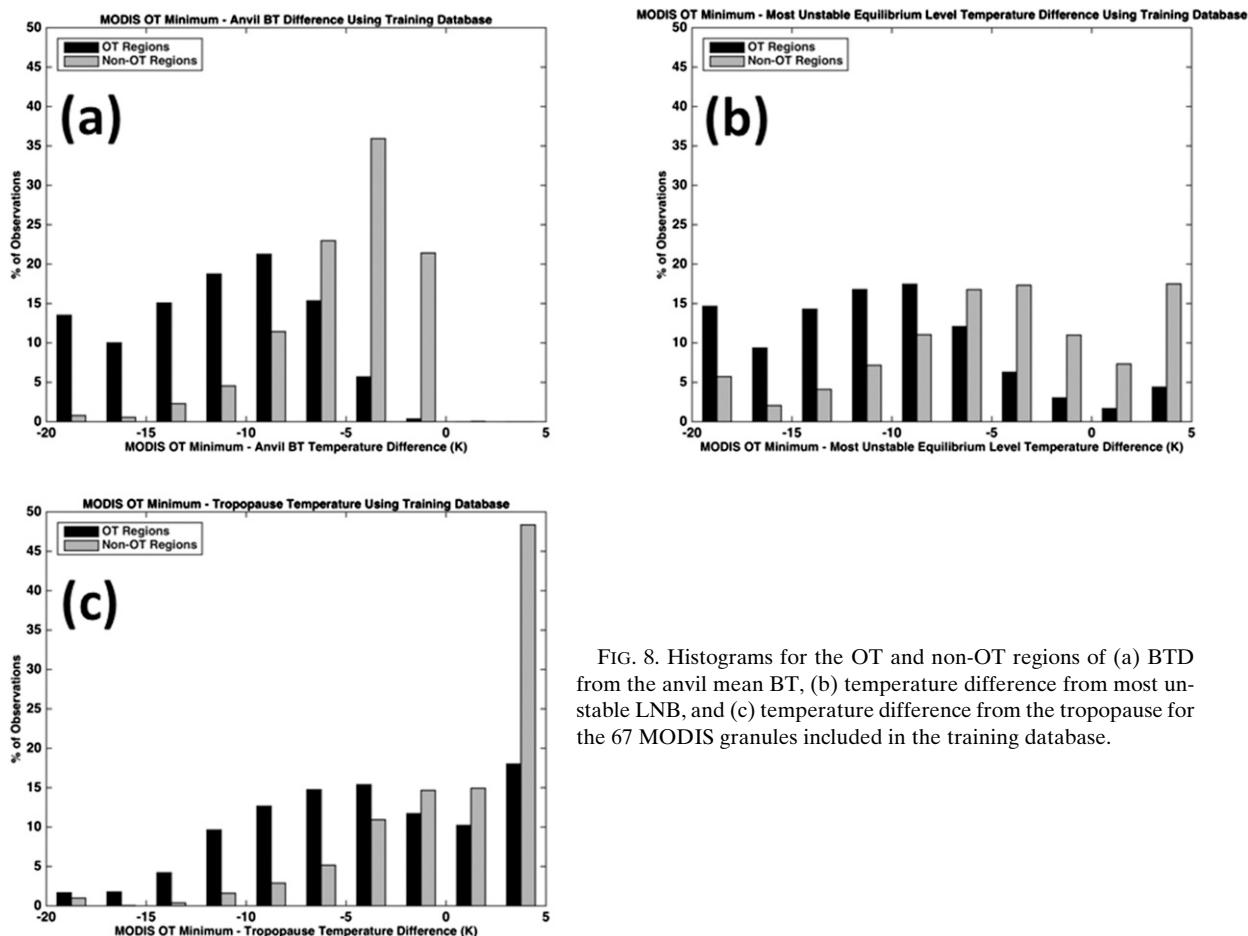


FIG. 8. Histograms for the OT and non-OT regions of (a) BTD from the anvil mean BT, (b) temperature difference from most unstable LNB, and (c) temperature difference from the tropopause for the 67 MODIS granules included in the training database.

level for discriminating between OT and non-OT regions. These parameters are indicators of updraft intensity and cloud-top height. One might assume that this parameter set is providing redundant information, but this is not necessarily the case here. The OT–anvil BT difference is based on pattern recognition within the satellite observations and is the biggest contributor to the probability. Although the satellite observations may depict an OT-like feature, the probability should also be anchored to a storm-environment analysis provided by a numerical model, which can decrease the probability for errant detections and increase the likelihood for correct detections. If there is no instability near the OT detection, it is likely errant and the most unstable LNB is set to 170 K to decrease probability. If the OT BT were significantly warmer than the tropopause, this would be further confirmation of an errant detection and the probability would decrease.

It is also useful to consider why some parameters were not useful for OT discrimination. The surface-based LNB temperature was not significant because

many (29%) of the OT-producing storms were in an environment with minimal or no surface-based instability. Although the presence of wind shear is often used to diagnose environments favorable for hazardous storms (Brooks 2013), the wind shear parameters were not statistically significant because OT and non-OT anvil pixels are often in close proximity to each other. These parameters would be the same or very similar for the two pixel populations and thus would not be useful.

Histograms of the three significant parameters for both OT and non-OT anvils are shown in Fig. 8. There is a clear difference between the OT and non-OT populations, especially for the OT–anvil BT difference (Fig. 8a). Nearly all OT regions have IRW BT colder than the most unstable LNB (Fig. 8b), but 28% of OT regions were not colder than the tropopause (Fig. 8c), indicating that not all OT features evident in VIS imagery penetrate the tropopause. These would be missed by detection approaches that require this criterion, such as B2010 and Rossow and Pearl (2007).

TABLE 2. A contingency table demonstrating the ability of the IRW-based pattern recognition and OT probability to discriminate between OT and non-OT anvil regions using the validation database that has 33 MODIS granules. The percentage of the total number of OT and non-OT anvil regions is shown in parentheses. The OT discrimination skill (percentage of OT regions plus non-OT regions correctly recognized) is 80.7%.

No. of OT regions: 809	No. of non-OT regions: 615
No. of OT regions with OT probability ≥ 0.5 : 593 (41.6%)	No. of non-OT regions with OT probability ≥ 0.5 : 58 (4.1%)
No. of OT regions with OT probability < 0.5 or lack of OT detection: 216 (15.2%)	No. of non-OT regions with OT probability < 0.5 or lack of OT detection: 557 (39.1%)

The OT discrimination skill is 81% using a 0.5 OT probability threshold with the Table 1 coefficients within Eq. (3) (Table 2). The greatest detriment to the skill was undetected OT regions, which is to be expected considering that OT regions were identified by a human using 0.25-km VIS imagery but were detected with an algorithm that uses 1-km VIS and 4-km IRW data. Only 4.1% of the non-OT regions were assigned an OT probability ≥ 0.5 , indicating that the combination of these three parameters is very effective for discriminating between OT and non-OT.

The POD and FAR as a function of OT probability are shown in Fig. 9. The POD for an OT probability > 0 is 81%, indicating that 19% of human-identified OTs were missed by pattern recognition. The OT POD drops by only $\sim 11\%$, but the FAR decreases by $\sim 50\%$ at the 0.7 probability threshold, providing the greatest separation between POD and FAR of any probability. A steep decline in POD occurs beyond probability ≥ 0.7 , reaching 0.46 for probability ≥ 0.95 . FAR also continued to decline to a value of 0.05.

The POD and FAR drop significantly when the VIS rating is used to constrain the IRW-based detections. The FAR is 5% for probability > 0 , indicating that the VIS texture detection can differentiate the correct detections from the much larger population of incorrect detections. Detections with an OT probability ≥ 0.95 overlaid with VIS rating are rarely incorrect (< 0.01 FAR). The combined VIS and IRW POD is significantly lower than IRW alone because of the fact that, even though an OT is apparent to a human in 0.25-km VIS data, it may not be quite as apparent in 1-km data nor may it be detectable by an automated algorithm. Analysis indicates that OTs in storms with small diameter are missed more often than those in larger storms (not shown). The frequency analysis requires a sufficient number of anvil-cloud pixels within the 32×32 pixel window that may not be present for a small storm.

Detection performance is compared with B2010 in Table 3. When OTs detected using the B2010 candidates and detection criteria are compared with the validation database, the POD (FAR) is 34% below (6.5% above)

that from the OT probability product. Bedka et al. (2012) showed a POD and FAR of 55% and 17%, respectively, using the B2010 algorithm applied to ~ 4 -km GEO imagery and based on CloudSat radar echo tops ≥ 0.5 km above the surrounding anvil as OT truth. We hypothesize that many human-observed OTs identified within 0.25-km VIS imagery in our study may not be ≥ 0.5 km above the anvil and could therefore be harder to detect. The FAR may be greater in this study relative to the Bedka et al. (2012) paper because only 265 OT detection pixels along the CloudSat overpass were analyzed and the OTs were typically over the tropics. In our study, the full domains of 33 granules were analyzed and the granules were better distributed across the globe, which may have incorporated more complex cloud scenes that triggered additional false detection. The FAR drops significantly when the B2010 detection criteria are applied to OT candidates identified by pattern recognition. This indicates that the simple spatial filtering of OT candidates by B2010 is a source of error within their algorithm. POD also decreases, which is an indication that the B2010 and Monette et al. (2012) detection criteria are too strict. In contrast, when the VIS rating constrains OT probability, FAR is

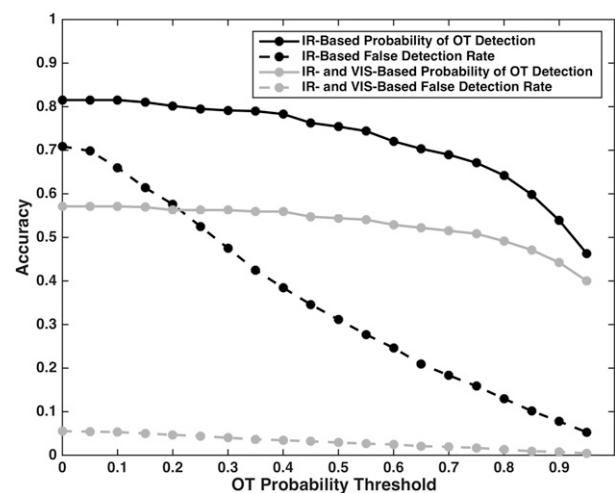


FIG. 9. OT POD and FAR derived as a function of OT probability for the IRW-only and IRW-VIS combined detections.

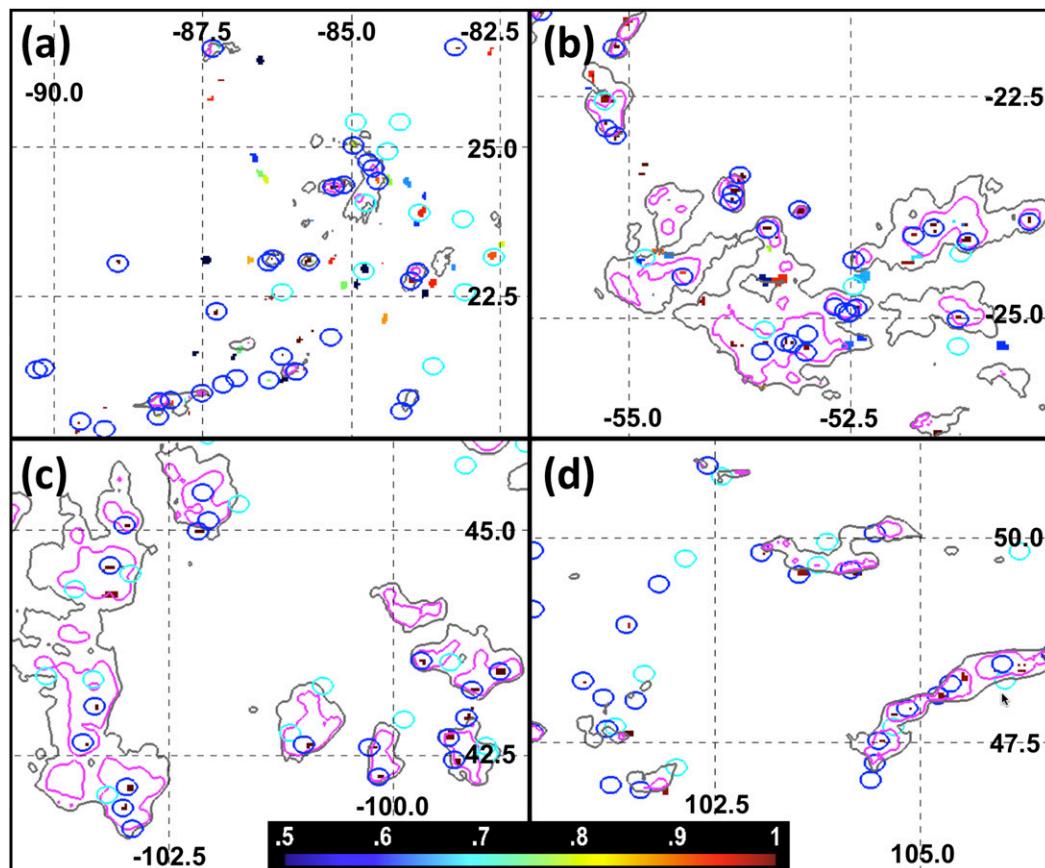


FIG. 10. The OT probability (color scale at bottom) along with VIS-rating values of 7 (gray contours) and 15 (magenta contours) atop human identifications of OT (blue circles) and non-OT regions (cyan circles) for the four MODIS scenes shown in Figs. 1 and 2.

comparable to the improved candidates and B2010 detection criteria but POD increases by 24%.

Detection results for four MODIS scenes show the characteristics of our algorithm relative to B2010. Figures 1 and 2 illustrate that deep convection and OTs have a similar appearance in VIS imagery regardless of storm location but vary significantly in the IRW, causing algorithms that use fixed IRW-based detection criteria to perform differently across cases. The OT probability and VIS-rating detection products overlaid with human-identified OT and non-OT anvil regions (Fig. 10) show that, when high OT probability (≥ 0.9) is collocated with VIS rating, the detections are rarely incorrect. Lower OT probability (< 0.7) detections represent the majority of those located outside human OT identifications, consistent with the results in Fig. 9. Comparable performance is evident for the Hurricane Katrina, central-U.S., and Brazil cases. A greater number of human OTs are missed for the Mongolia case. This is to be expected given how weak and warm the OT regions appear to be. Our experiences suggest that the Mongolia case will be a

challenge for any automated VIS- or IR-based algorithm. Very few non-OT regions are detected by the OT-probability product, reinforcing the results from Table 3.

The B2010-based detection results (Fig. 11) indicate that when the B2010 detection criteria are applied to the improved OT candidates (red), the detections are very accurate. The original B2010 framework (cyan) performs reasonably well, but some false detections are evident. B2010 performed well for the Brazil and central-U.S. cases but poorly for the Mongolia and Katrina cases. Most OTs in the Mongolia case were not colder than 215 K nor were they prominent enough relative to the anvil. Most of the OTs in the Katrina case were not colder than the tropopause or did not show much prominence in the IRW data.

Although we sought to emulate GEO imagery with MODIS by resampling to 4-km IRW resolution, one aspect of GEO that cannot be emulated is the effect of varying solar illumination on OT appearance and detection. OTs are more apparent to humans in VIS imagery at high solar zenith angle than when they are

TABLE 3. OT detection validation using the B2010 candidates and detection criteria, improved candidates and B2010 detection criteria, and the OT probability threshold and VIS-rating mask that are described in this paper.

	POD (%)	FAR (%)
B2010 candidates and detection criteria	35.1	24.9
Improved candidates and B2010 detection criteria	27.3	1.1
OT probability ≥ 0.7	69.2	18.4
OT probability ≥ 0.7 and VIS-rating detection	51.4	1.6

illuminated at solar noon, which can have an impact on detection with an automated algorithm. A time series of imagery of strong thunderstorms over Texas from GOES-14 VIS and IRW, OT detection, and radar reflectivity is shown in Figs. 12–15) to demonstrate how the detection algorithm performs with actual GEO data.

Despite the complexity of these storms, the OT-probability and VIS-rating products performed well. VIS texture was detected within all but three human-identified OTs, often with large (>15) VIS-rating values. Consequently, when OT probability was collocated with VIS rating, the detection was rarely incorrect. There are many regions in IRW imagery with no evidence of VIS

OT that appear very similar to IRW within VIS OTs in other parts of the images. These situations would be interpreted as false detection by a user but they are almost impossible to eliminate, especially at night. Correctly identified OTs, and the VIS-rating product in general, typically occur in regions with >20 -dBZ reflectivity at 10 km. Reflectivity cores at 10 km indicate strong updrafts with cloud tops that are likely penetrating through the anvil level, which was ~ 12 km for this case as based on comparison between IRW BT and a nearby sounding. The vast majority of these cores are outlined by VIS rating, and a smaller but still significant fraction of cores are in close proximity to high

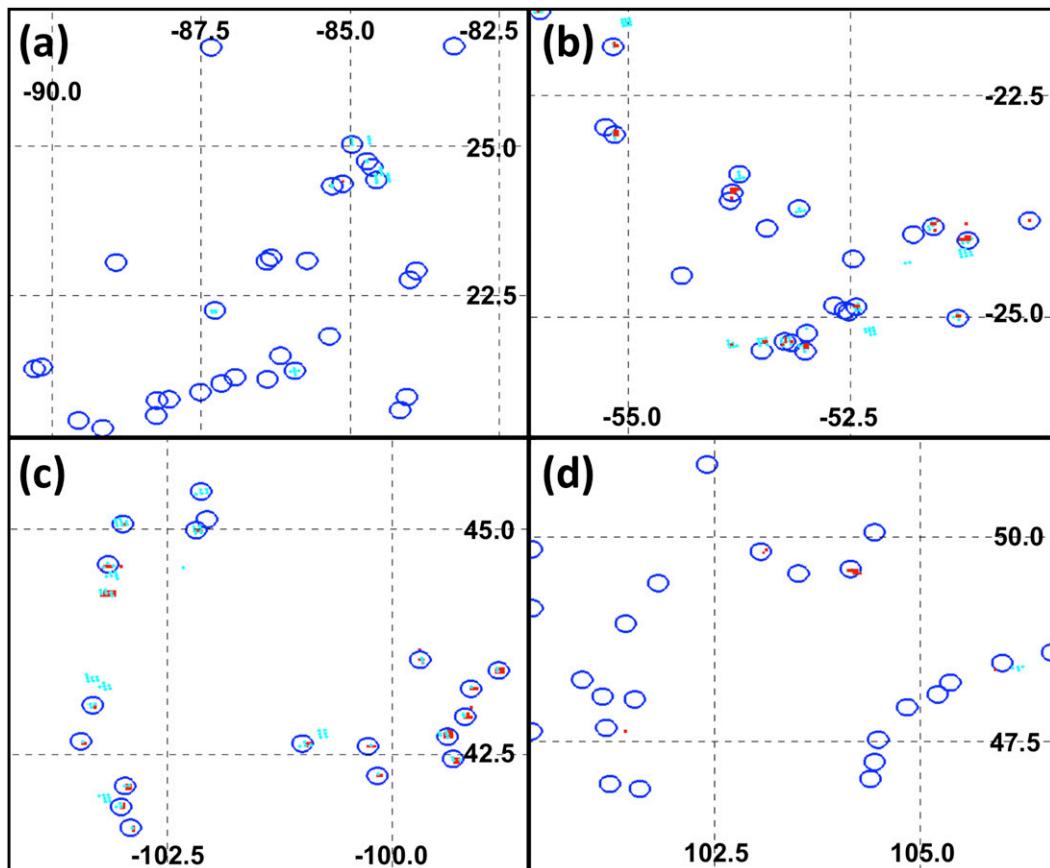


FIG. 11. B2010 detections using the original (cyan) and improved (red) OT candidates atop human OT identifications (blue circles) for the four MODIS scenes shown in Figs. 1 and 2.

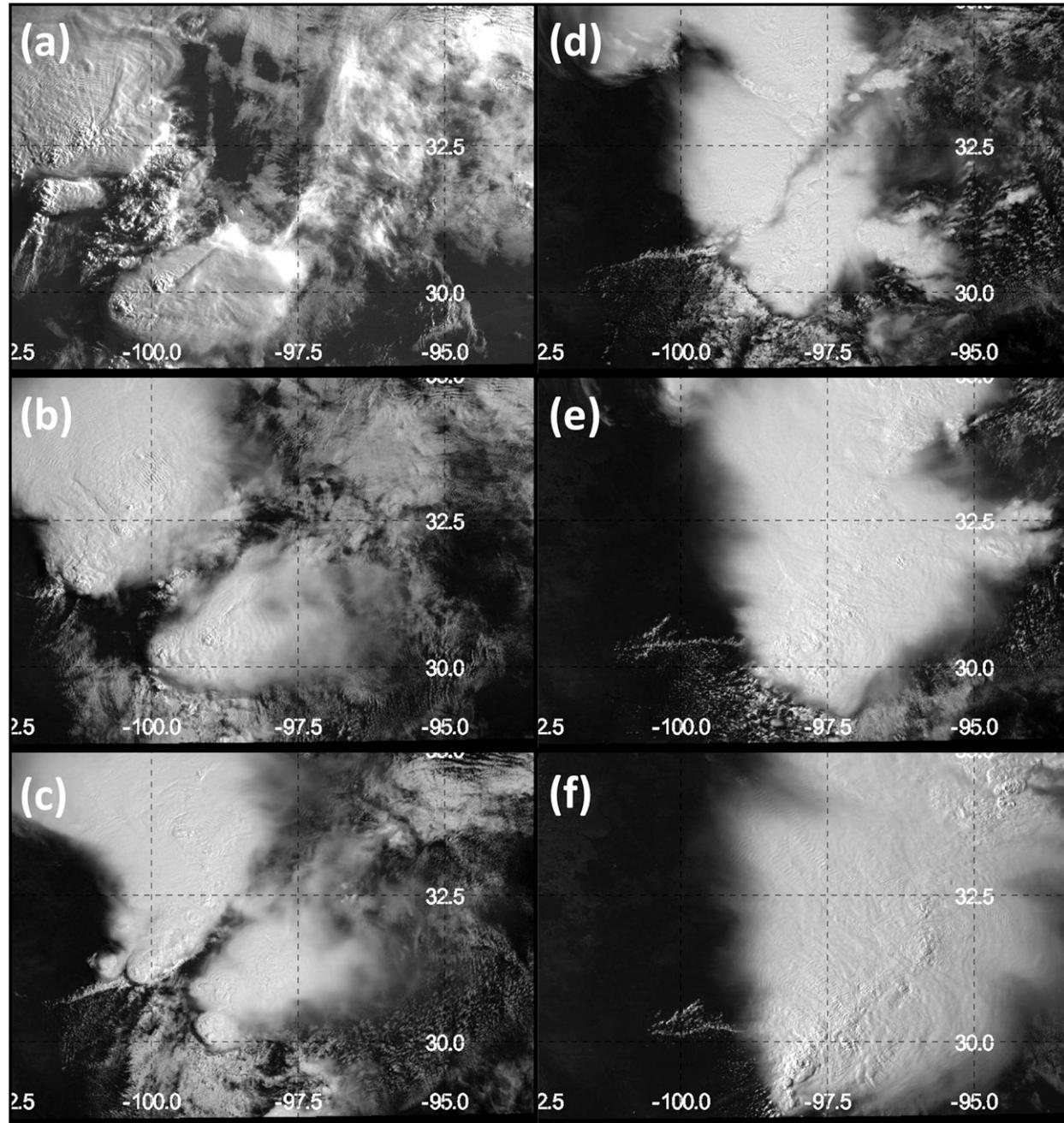


FIG. 12. *GOES-14* VIS reflectance imagery at (a) 1245, (b) 1445, (c) 1645, (d) 1845, (e) 2045, and (f) 2245 UTC 25 May 2015.

OT probability. Animations of OT detections using 1-min imagery from *GOES-14* for this case and two others (11 May 2014 and 16 August 2015) are provided as online supplemental material to demonstrate algorithm performance for a variety of storm types and intensities and solar-illumination conditions. From these results, it is clear that the OT detection algorithms can operate very effectively on both GEO and low-Earth-orbit

(LEO) imagery, providing a capability to monitor hazardous storms anytime and anywhere.

Of the 617 MODIS OTs identified over the United States, 166 (~27%) were found to be associated with severe weather. This relatively low fraction is not surprising because OTs are common but severe-weather reports are relatively rare—a point discussed by Bedka et al. (2015). Using the B2010 approach, Dworak et al.

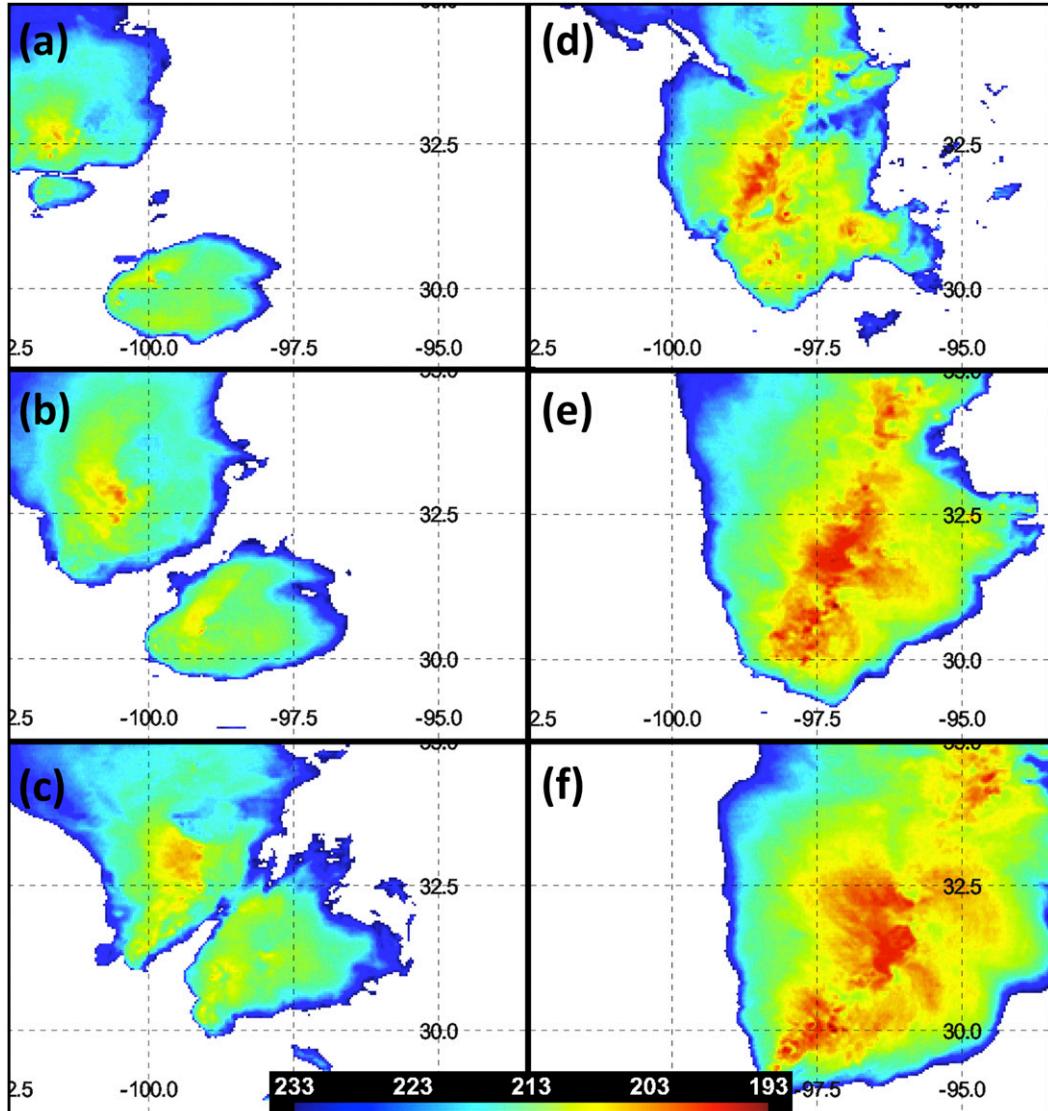


FIG. 13. GOES-14 IRW BT imagery for the same scenes shown in Fig. 12.

(2012) found that 42% of severe storms during the 2004–09 warm seasons were collocated with an OT detection. Approximately 45% of the severe storms in our sample were detected using B2010 (Table 4). Because the OT POD and FAR showed maximum separation for OT probability ≥ 0.7 , this threshold is used to investigate the severe-storm detection capability of our algorithm relative to B2010. The probability product offers an $\sim 20\%$ improvement over B2010. The VIS rating identified $\sim 8.5\%$ fewer severe OTs than did the OT probability, which is not surprising given the results from Fig. 9. Severe OTs were detected $\sim 7\%$ more often than non-severe U.S. storms using VIS data, which suggests that 1) severe-storm updrafts are stronger and produce greater texture and/or 2) severe weather preferentially occurs in

larger storms, which aids automated detection. The combined OT probability and VIS-rating detection identified 52% of the severe OTs, which still exceeds the detection rate of B2010 by $\sim 7\%$ but decreases the OT FAR to $\sim 1\%$ according to Fig. 9.

5. Summary and future work

A new satellite-based probabilistic OT detection algorithm has been developed to address the limitations of and improve upon the accuracy of previous methods. The algorithms attempt to emulate the analysis process used by a human by first identifying anvils and then performing pattern recognition to identify embedded OTs. The temperature differences between the OT and

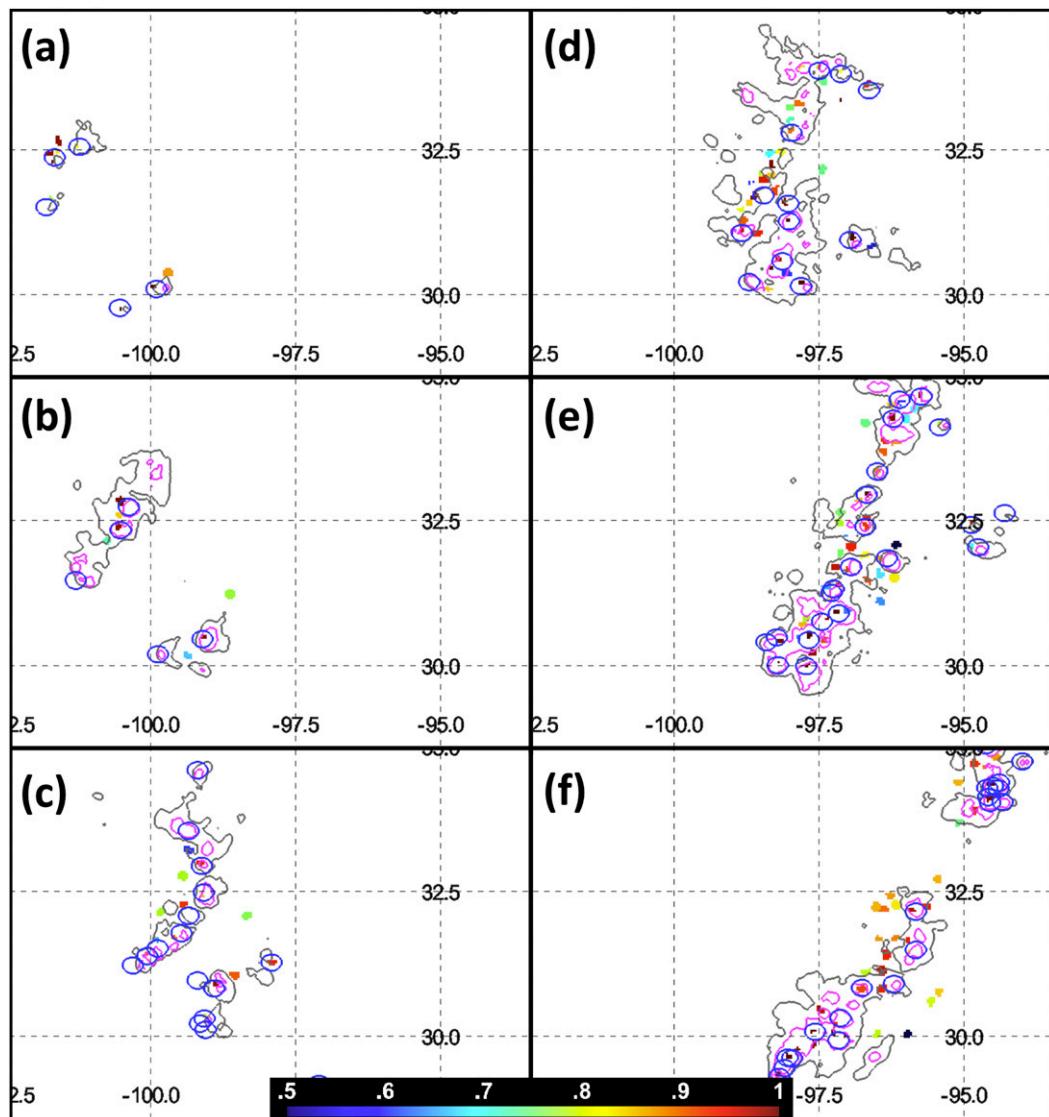


FIG. 14. The OT probability (color scale at bottom) along with VIS-rating values of 7 (gray contours) and 15 (magenta contours) atop human OT identifications (blue circles) for the same scenes shown in Fig. 12.

anvil, tropopause, and most unstable LNB temperatures were found to be statistically significant parameters for differentiating between OT and non-OT anvil regions. These three parameters are used to produce an OT probability for each OT detection. Spatial, statistical, and frequency analyses of VIS reflectance were used to develop a VIS rating that indicates the presence of texture within and near OTs. The VIS rating can be used to constrain likely errant IRW-based detections during daytime.

Validation using a database of human OT identifications indicates that, for a comparable FAR, the approach offers a significant improvement in OT detection rate relative to B2010. Four MODIS scenes and six

GOES-14 scenes show that the OT probability and VIS rating produced relatively consistent results for all of the cases. B2010 was unable to identify OTs in two MODIS scenes because of issues associated with strict B2010 detection criteria. The GOES-14 scenes showed that correctly identified OTs and VIS texture detections typically occurred in regions with high radar reflectivity values at a 10-km altitude, indicative of strong updrafts. The detections were rarely incorrect when the OT probability is constrained by texture detection from the VIS-rating product.

The relationships between OT signatures and severe-weather reports over the United States were examined. The OT probability product offered a 20% improvement

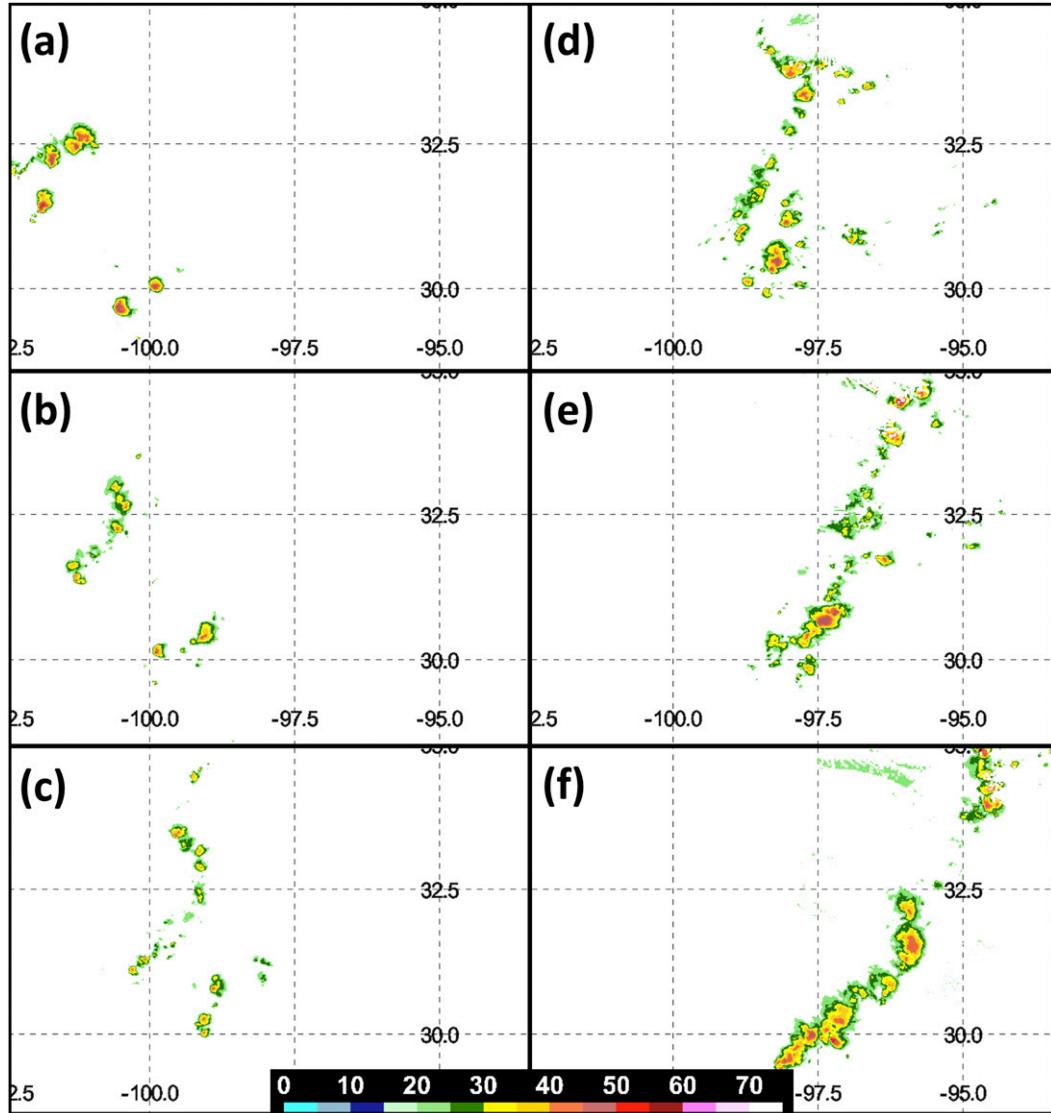


FIG. 15. WSR-88D reflectivity at 10-km altitude using the methods outlined in Homeyer (2014) and updated in Homeyer and Kumjian (2015) for the same scenes shown in Fig. 12.

in severe storm detection rate relative to what would be provided by B2010. Over one-half of the OTs within severe storms were detected by both the IRW and VIS algorithms.

This paper shows that the new probabilistic OT detection approach is a significant improvement over existing methods. The accuracy of the algorithm and applicability to historical, current, and future LEO and

TABLE 4. The detection rate for OTs in nonsevere storms and severe storms over the United States using the B2010 criteria, OT probability, VIS rating, and a combination of the OT probability and VIS-rating approaches.

	OT detection rate for U.S. nonsevere storms (%)	OT detection rate for U.S. severe storms (%)
Improved candidates and B2010 detection criteria	33.0	45.8
OT probability ≥ 0.7	66.5	65.1
VIS-rating detection	49.0	56.6
OT probability ≥ 0.7 and VIS-rating detection	45.2	52.4

GEO imagery make it suitable for weather analysis and forecasting applications as well as the development of long-term climate-data records of hazardous-storm events. We will work to intercalibrate IRW BTs throughout the historical GEO data record using long-lived and reliable LEO sensors such as TRMM VIRS, MODIS, AVHRR, and/or HIRS. This will help to ensure that the trends in OT activity are caused by climate variability and not by sensor artifacts. We will continue to quantify algorithm accuracy and relationships with other remotely sensed datasets such as WSR-88D (Homeyer and Kumjian 2015), TRMM (Liu and Zipser 2005), and GPM radar OT indicators (Liu and Zipser 2015). We will also continue to examine relationships between OTs and severe weather, especially with the advent of the 30-s imagery that will be provided by the GOES-R ABI.

Acknowledgments. This research has been supported by the GOES-R Risk Reduction Research (R3) program. In particular, we thank Dr. Steven Goodman, senior (chief) scientist, GOES-R System Program, for his guidance and support throughout this effort. We thank Patrick Minnis and Christopher Velden for their advice and collaboration throughout the algorithm-development process. We thank Cameron Homeyer for providing the WSR-88D data shown in this paper. We also thank Jake Smith for manually identifying OT locations in *GOES-14* satellite imagery.

REFERENCES

- Bedka, K. M., 2011: Overshooting cloud top detections using MSG SEVIRI infrared brightness temperatures and their relationship to severe weather over Europe. *Atmos. Res.*, **99**, 175–189, doi:10.1016/j.atmosres.2010.10.001.
- , J. Brunner, R. Dworak, W. Feltz, J. Otkin, and T. Greenwald, 2010: Objective satellite-based overshooting top detection using infrared window channel brightness temperature gradients. *J. Appl. Meteor. Climatol.*, **49**, 181–202, doi:10.1175/2009JAMC2286.1.
- , —, and W. Feltz, 2011: Overshooting top and enhanced-V signature detection. NOAA/NESDIS/Center for Satellite Applications and Research Algorithm Theoretical Basis Doc., 53 pp. [Available online at http://www.goes-r.gov/products/ATBDs/option2/Aviation_OvershootingTop_v1_no_color.pdf.]
- , R. Dworak, J. Brunner, and W. Feltz, 2012: Validation of satellite-based objective overshooting cloud-top detection methods using CloudSat Cloud Profiling Radar observations. *J. Appl. Meteor. Climatol.*, **51**, 1811–1822, doi:10.1175/JAMC-D-11-0131.1.
- , C. Wang, R. Rogers, L. Carey, W. Feltz, and J. Kanak, 2015: Examining deep convective cloud evolution using total lightning, WSR-88D, and *GOES-14* super rapid scan datasets. *Wea. Forecasting*, **30**, 571–590, doi:10.1175/WAF-D-14-00062.1.
- Berendes, T. A., J. R. Mecikalski, W. M. MacKenzie Jr., K. M. Bedka, and U. S. Nair, 2008: Convective cloud identification and classification in daytime satellite imagery using standard deviation limited adaptive clustering. *J. Geophys. Res.*, **113**, D20207, doi:10.1029/2008JD010287.
- Bessho, K., and Coauthors, 2016: An introduction to *Himawari-8/9*—Japan's new-generation geostationary meteorological satellites. *J. Meteor. Soc. Japan*, **94**, 151–183, doi:10.2151/jmsj.2016-009.
- Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmos. Res.*, **123**, 129–138, doi:10.1016/j.atmosres.2012.04.002.
- Brunner, J. C., S. A. Ackerman, A. S. Bachmeier, and R. M. Rabin, 2007: A quantitative analysis of the enhanced-V feature in relation to severe weather. *Wea. Forecasting*, **22**, 853–872, doi:10.1175/WAF1022.1.
- CNN, 2013: Lethal weather on world's most dangerous lake. [Available online at <http://www.cnn.com/2013/01/17/world/africa/lake-victoria-weather-deaths/>.]
- Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.*, **18**, 1016–1022, doi:10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2.
- Dworak, R., K. M. Bedka, J. Brunner, and W. Feltz, 2012: Comparison between *GOES-12* overshooting top detections, WSR-88D radar reflectivity, and severe storm reports. *Wea. Forecasting*, **27**, 684–699, doi:10.1175/WAF-D-11-00070.1.
- Goodman, S. J., and Coauthors, 2012: The GOES-R Proving Ground: Accelerating user readiness for the next-generation geostationary environmental satellite system. *Bull. Amer. Meteor. Soc.*, **93**, 1029–1040, doi:10.1175/BAMS-D-11-00175.1.
- Gravelle, C. M., J. R. Mecikalski, W. E. Line, K. M. Bedka, R. A. Petersen, J. M. Sieglaff, G. T. Stano, and S. J. Goodman, 2016: Demonstration of a GOES-R satellite convective toolkit to “bridge the gap” between severe weather watches and warnings: An example from the 20 May 2013 Moore, Oklahoma, tornado outbreak. *Bull. Amer. Meteor. Soc.*, **97**, 69–84, doi:10.1175/BAMS-D-14-00054.1.
- Griffin, S. M., K. M. Bedka, and C. S. Velden, 2016: A method for calculating the height of overshooting convective cloud tops using satellite-based IR imager and CloudSat cloud profiling radar observations. *J. Appl. Meteor. Climatol.*, **55**, 479–491, doi:10.1175/JAMC-D-15-0170.1.
- Homeyer, C. R., 2014: Formation of the enhanced-V infrared cloud-top feature from high-resolution three-dimensional radar observations. *J. Atmos. Sci.*, **71**, 332–348, doi:10.1175/JAS-D-13-079.1.
- , and M. R. Kumjian, 2015: Microphysical characteristics of overshooting convection from polarimetric radar observations. *J. Atmos. Sci.*, **72**, 870–891, doi:10.1175/JAS-D-13-0388.1.
- Khlopenkov, K. V., and A. P. Trishchenko, 2008: Implementation and evaluation of concurrent gradient search method for re-projection of MODIS level 1B imagery. *IEEE Trans. Geosci. Remote Sens.*, **46**, 2016–2027, doi:10.1109/TGRS.2008.916633.
- , and —, 2010: Achieving subpixel georeferencing accuracy in the Canadian AVHRR Processing System. *IEEE Trans. Geosci. Remote Sens.*, **48**, 2150–2161, doi:10.1109/TGRS.2009.2034974.
- Kirk-Davidoff, D. B., E. J. Hints, J. G. Anderson, and D. W. Keith, 1999: The effect of climate change on ozone depletion through changes in stratospheric water vapour. *Nature*, **402**, 399–401, doi:10.1038/46521.
- Liu, C., and E. J. Zipser, 2005: Global distribution of convection penetrating the tropical tropopause. *J. Geophys. Res.*, **110**, D23104, doi:10.1029/2005JD006063.
- , and —, 2015: The global distribution of largest, deepest, and most intense precipitation systems. *Geophys. Res. Lett.*, **42**, 3591–3595, doi:10.1002/2015GL063776.

- Merino, A., L. Lopez, J. L. Sanchez, E. Garcia-Ortega, E. Cattani, and V. Levizzani, 2014: Daytime identification of summer hailstorm cells from MSG data. *Nat. Hazards Earth Syst. Sci.*, **14**, 1017–1033, doi:10.5194/nhess-14-1017-2014.
- Mikuš, P., and N. Strelec Mahović, 2013: Satellite-based overshooting top detection methods and the analysis of correlated weather conditions. *Atmos. Res.*, **123**, 268–280, doi:10.1016/j.atmosres.2012.09.001.
- Monette, S. A., C. S. Velden, K. S. Griffin, and C. M. Rozoff, 2012: Examining trends in satellite-detected tropical overshooting tops as a potential predictor of tropical cyclone rapid intensification. *J. Appl. Meteor. Climatol.*, **51**, 1917–1930, doi:10.1175/JAMC-D-11-0230.1.
- Negri, A. J., and R. F. Adler, 1981: Relation of satellite-based thunderstorm intensity to radar-estimated rainfall. *J. Appl. Meteor.*, **20**, 288–300, doi:10.1175/1520-0450(1981)020<0288:ROSBTI>2.0.CO;2.
- Proud, S. R., 2015: Analysis of overshooting top detections by Meteosat Second Generation: A 5-year dataset. *Quart. J. Roy. Meteor. Soc.*, **141**, 909–915, doi:10.1002/qj.2410.
- Punge, H. J., A. Werner, K. M. Bedka, M. Kunz, and M. Puskeiler, 2014: A new physically based stochastic event catalog for hail in Europe. *Nat. Hazards Earth Syst. Sci.*, **73**, 1625–1645, doi:10.1007/s11069-014-1161-0.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.
- Reynolds, D. W., 1980: Observations of damaging hailstorms from geosynchronous satellite digital data. *Mon. Wea. Rev.*, **108**, 337–348, doi:10.1175/1520-0493(1980)108<0337:OODHFG>2.0.CO;2.
- Rossow, W. B., and C. Pearl, 2007: 22-year survey of tropical convection penetrating into the lower stratosphere. *Geophys. Res. Lett.*, **34**, L04803, doi:10.1029/2006GL028635.
- Schmetz, J., S. A. Tjemkes, M. Gube, and L. van de Berg, 1997: Monitoring deep convection and convective overshooting with Meteosat. *Adv. Space Res.*, **19**, 433–441, doi:10.1016/S0273-1177(97)00051-3.
- Schmit, T. J., M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, 2005: Introducing the next-generation Advanced Baseline Imager on GOES-R. *Bull. Amer. Meteor. Soc.*, **86**, 1079–1096, doi:10.1175/BAMS-86-8-1079.
- Semazzi, F., and Coauthors, 2011: Enhancing safety of navigation and efficient exploitation of natural resources over Lake Victoria and its basin by strengthening meteorological services on the lake. North Carolina State University Climate Modeling Laboratory Final Rep., 104 pp. [Available online at http://climlab02.meas.ncsu.edu/HYVIC/Final_Report_LVBC.pdf.]
- Setvák, M., and Coauthors, 2010: Cold-ring-shaped cloud top features atop convective storms. *Atmos. Res.*, **97**, 80–96, doi:10.1016/j.atmosres.2010.03.009.
- , K. Bedka, D. T. Lindsey, A. Sokol, Z. Charvát, J. Št'áská, and P. K. Wang, 2013: A-Train observations of deep convective storm tops. *Atmos. Res.*, **123**, 229–248, doi:10.1016/j.atmosres.2012.06.020.
- Shindell, D. T., 2001: Climate and ozone response to increased stratospheric water vapour. *Geophys. Res. Lett.*, **28**, 1551–1554, doi:10.1029/1999GL011197.
- Thiery, W., E. L. Davin, S. I. Seneviratne, K. M. Bedka, S. Lhermitte, and N. van Leipzig, 2016: Climate change intensifies hazardous storms over Lake Victoria. *Nature Commun.*, in press.
- Wang, C., Z. J. Luo, and X. Huang, 2011: Parallax correction in collocating *CloudSat* and Moderate Resolution Imaging Spectroradiometer (MODIS) observations: Method and application to convection study. *J. Geophys. Res.*, **116**, D17201, doi:10.1029/2011JD016097.
- Yamamoto, M. K., F. A. Furuzawa, A. Higuchi, and K. Nakamura, 2008: Comparison of diurnal variations in precipitation systems observed by TRMM PR, TMI, and VIRS. *J. Climate*, **21**, 4011–4028, doi:10.1175/2007JCLI2079.1.
- Young, A. H., J. J. Bates, and J. A. Curry, 2012: Complementary use of passive and active remote sensing for detection of penetrating convection from *CloudSat*, *CALIPSO*, and *Aqua* MODIS. *J. Geophys. Res.*, **117**, D13205, doi:10.1029/2011JD016749.