Ensemble Prediction of Oceanic Convective Hazards (EPOCH) Assessment: Part II

November 2019

Laura D. Melling
Arlene G. Laing
Matthew S. Wandishin
Joan E. Hart
Melissa A. Petty

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

A Technical Memorandum OAR GSD-65

Ensemble Prediction of Oceanic Convective Hazards (EPOCH) Assessment: Part II

Laura D. Melling[1]
Arlene G. Laing[2]
Matthew S. Wandishin[1]
Joan E. Hart[1]
Melissa A. Petty[2]

[1]Cooperative Institute for Research in Environmental Sciences (CIRES) and NOAA/ESRL/GSD
[2]Cooperative Institute for Research in the Atmosphere (CIRA) and NOAA/ESRL/GSD

UNITED STATES DEPARTMENT OF COMMERCE	NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION	Office of Oceanic and Atmospheric Research
Wilbur Ross Secretary	Dr. Neil Jacobs Acting NOAA Administrator	Craig N. McLean Assistant Administrator

Ensemble Prediction of Oceanic Convective Hazards (EPOCH) Assessment: Part II
February 6, 2019

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Quality Assessment Product Development Team (QA PDT) was tasked with assessing the Ensemble Prediction of Oceanic Convective Hazards (EPOCH). EPOCH was developed by the Convective Weather Product Development Team (CW PDT) at the National Center for Atmospheric Research (NCAR), under sponsorship from the FAA's Aviation Weather Research Program (AWRP). EPOCH provides synoptic 6-hourly global forecasts of the likelihood of thunderstorm occurrence and convective clouds exceeding 30, 35, and 40 kft, on a 1.0° grid.

The EPOCH products were assessed and compared with the baseline World Area Forecast System (WAFS) thunderstorm (Cb) forecasts through the application of climatological mapping, statistical metrics from contingency tables, and case studies. In Part I of the EPOCH assessment, forecasts were evaluated for the convective season in the Northern Hemisphere (1 June 2017 - 30 September 2017). The active season in the Southern Hemisphere (1 December 2017 - 31 March 2018) was the focus for this report, Part II of the assessment.

As a probabilistic forecast, the EPOCH fields are not directly analogous to the WAFS Cb fields. In particular, EPOCH gives the likelihood of thunderstorm occurrence, as well as the likelihood of convective clouds exceeding 30, 35, and 40 kft, while WAFS provides two deterministic fields: Cb horizontal extent and Cb tops. In order to compare the two products, forecast thresholds were identified by matching the cumulative distributions produced by EPOCH and WAFS. For the EPOCH likelihood of thunderstorm occurrence and WAFS Cb tops fields, this yielded forecast thresholds corresponding to "occasional" and "rare" event occurrence (~9.6% and 2.6% global frequencies, respectively).

Multiple observation datasets were used as proxies for thunderstorm occurrence and convective cloud top observations:

- US Multi-Radar/Multi-Sensor Echo Top height (MRMS-ET)
- European Maximum Composite Reflectivity (EU-CR)
- Ground-based Lightning Detection Networks (LTG)
- Integrated Multi-satellitE Retrievals for GPM (IMERG) Precipitation Rate
- Geostationary satellite Cloud Top Height (CTH)
- Geostationary satellite Infrared Anvil Detection (IRA)

Although no direct global measures of convection were available for verification of EPOCH, two ground-based radar datasets (MRMS-ET and EU-CR) served as primary truth fields over their respective domains. The remaining datasets, all of which have global coverage, were employed as upper or lower bounds on the forecasts. Two modified datasets were added for Part II of the assessment:

- Lightning within a radius of 0.25° (LTG25)
- Cloud Top Height filtered to areas with a minimum cloud depth of 15 kft (CTH15)

Both forecast products (EPOCH and WAFS Cb) have coarse spatial and temporal resolution relative to the observation datasets. Upscaling the finer-scale observed data to match the forecast grid spacing (1° and 6-hr) required several choices with respect to the scale of a convective weather event considered to be relevant. As in Part I of this assessment, independent thresholds for minimum intensity, spatial extent, and temporal duration were used to explore a range of possible event definitions.

The global performance of EPOCH was found to be similar to that of the existing WAFS Cb product. As illustrated in Figure A, both EPOCH (top) and WAFS (middle) captured the main climatological features. The relative skill of the two forecasts is summarized in Figure B. WAFS had a higher Probability of Detection (POD) against all datasets except IRA at the "occasional" forecast threshold (left), and IRA and IMERG at the "rare" threshold (right). By contrast, EPOCH generally had a higher Success Ratio (SR) across the different observation datasets for both forecast thresholds. As in Part I of the assessment, EPOCH was found to produce more consistent forecasts than WAFS, with a greater difference in consistency between the two products at the "rare" forecast threshold (see Figure 26).
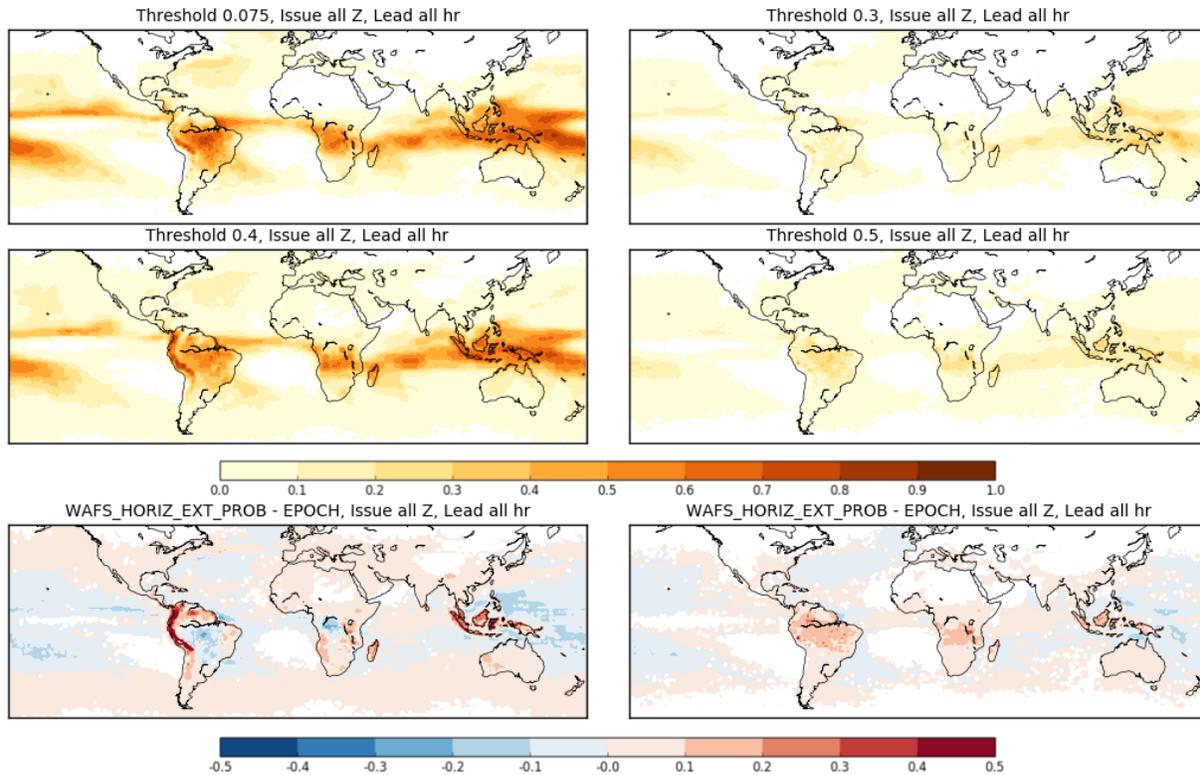
Figure A. Forecast climatological maps for EPOCH (top), WAFS (middle), and WAFS-EPOCH (bottom; red=more WAFS, blue=more EPOCH). The thresholds used for each forecast correspond to the "occasional" (left) and "rare" (right) event frequencies.
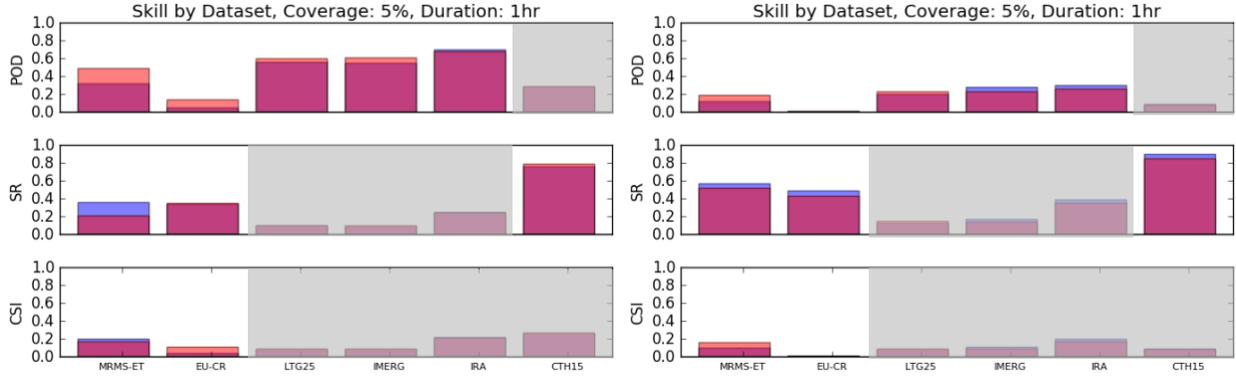
Figure B. Performance metrics by dataset using the equalized forecast thresholds corresponding to "occasional" (left) and "rare" (right) event occurrence. The scores for EPOCH (blue) and WAFS (red) are overlaid using 50% transparency. Gray shading indicates equivocal scores from bounding datasets. The observation definition for each dataset uses 5% spatial coverage, 1-hr temporal extent, and the primary intensity thresholds given in Table 5.
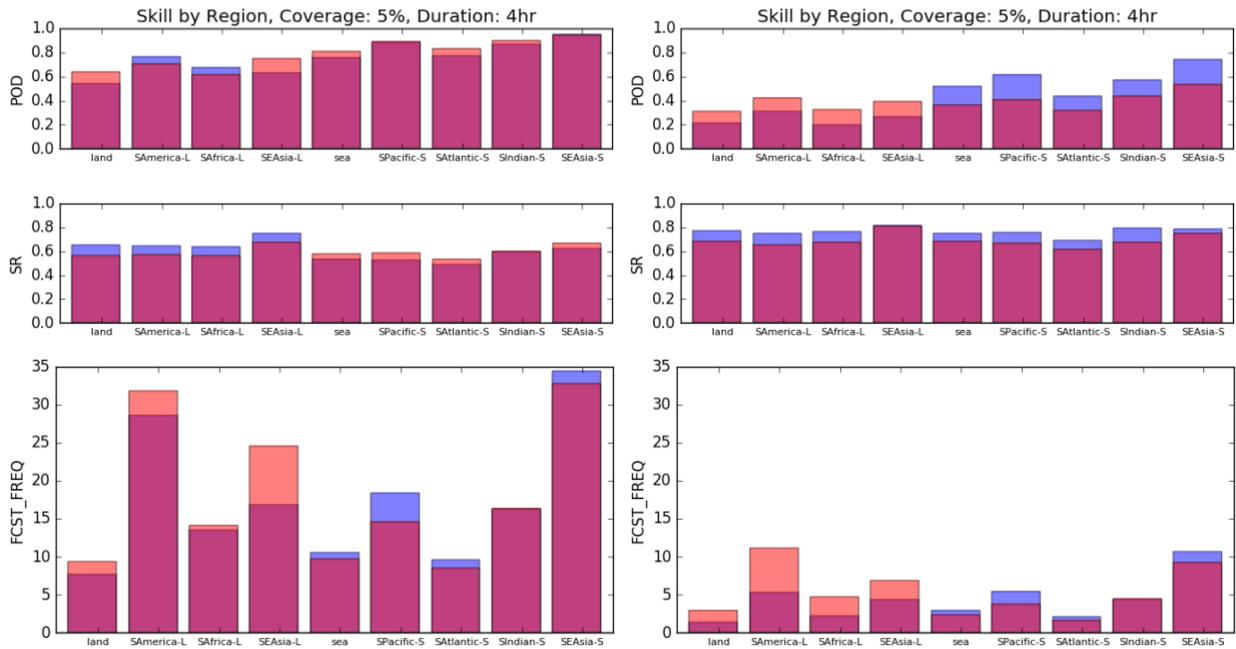
Figure C. Performance metrics for the land (L) and sea (S) regions using the equalized forecast thresholds corresponding to "occasional" (left) and "rare" (right) event occurrence. The POD (top) is obtained using IMERG ≥ 10mm/hr, while SR (middle) is obtained using CTH15 ≥ 30 kft. Forecast event frequencies (bottom) are given in percent. Scores using the 5% and 4-hr observed event thresholds are shown.

As seen in Part I, the frequency of events from EPOCH exceeded that from WAFS in the tropics, while WAFS surpassed the EPOCH rate at higher latitudes (see Figure 23). Results of forecast performance by land and sea region are summarized in Figure C. EPOCH produced more events in oceanic regions while WAFS produced more over land, at either forecast threshold. Despite a lower event frequency, EPOCH detected more events than WAFS in the South America and Southern Africa regions at the "occasional" threshold. Similarly, WAFS outperformed EPOCH in oceanic regions at the "occasional" threshold, having both better detection rate and lower false alarms. However, the reverse was true at the "rare" forecast threshold, where EPOCH outperformed WAFS for all oceanic regions.
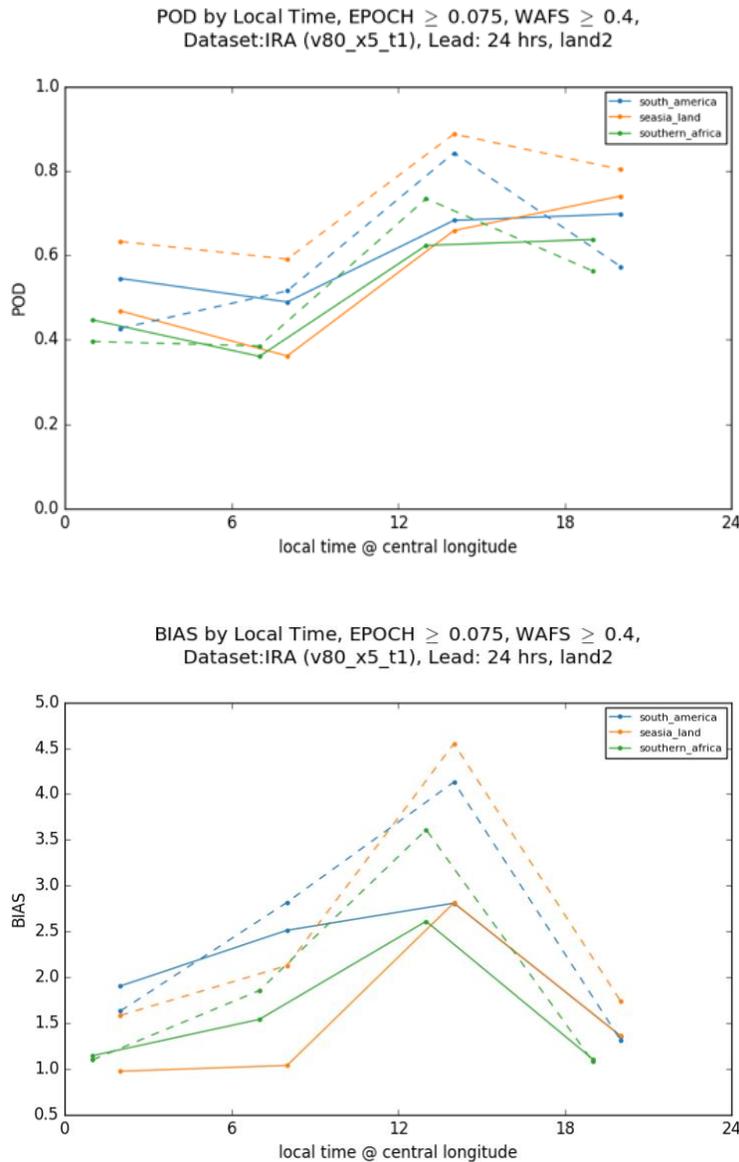
Figure D. POD (top) and forecast bias (bottom) versus local valid time for each land region for both EPOCH (solid) and WAFS (dashed) at the "occasional" forecast threshold using the IRA (v80_x5_t1) observation definition.

In Part I of this assessment, climatological studies over two land regions (Africa and South America) showed that WAFS tended to initiate convection earlier, while EPOCH maintained convection later into the evening hours. The diurnal pattern was investigated further in Part II by computing a "local valid time" using the UTC offset corresponding to the central longitude of each region. The POD and forecast bias against the IRA dataset for land regions are shown in Figure D. Both forecast products had higher detection rates for afternoon and evening convection, compared with overnight events (top). Both products also had a higher bias in the afternoon, but the difference was more pronounced in WAFS. Notably, EPOCH had a higher detection rate for evening convection, despite having a reduced bias (value approaching 1).

The performance of the forecast products in identifying convective cloud tops (CCT) at or above 30, 35, and 40 kft was evaluated against the filtered cloud top height observations, in addition to the MRMS-ET

used in Part I, in order to provide coverage in the Southern Hemisphere. Thresholds for each EPOCH CCT field were chosen to match the event occurrence rate found using the corresponding height threshold applied to the WAFS Cb tops field. Although WAFS generally detected more CCT events when using the MRMS-ET data, the skill for EPOCH and WAFS was found to be similar when using the filtered cloud top height dataset (see Figure 29).

Two case studies were conducted: a Southern Hemisphere case (South America) and a Northern Hemisphere case (Asia). Both EPOCH and WAFS were able to capture deep, organized convection over southern Brazil (Figure E, red rectangle) and near the Bolivia-Brazil border (solid yellow circle), while WAFS placed higher probabilities near the Andes (magenta oval). Over the Amazon basin (dashed yellow circle), EPOCH produced broad areas of moderate probabilities (exceeding the "occasional" threshold), while WAFS produced isolated pockets exceeding the "rare" threshold. In the Asia case, which was chosen to represent winter hemisphere convection that caused an aviation hazard, both products captured the core of the storm over the Persian Gulf, but WAFS covered a greater area that extended further inland and captured smaller storms in northern India that were missed by EPOCH (see Figure 35). The case study results were consistent with the global performance characteristics.
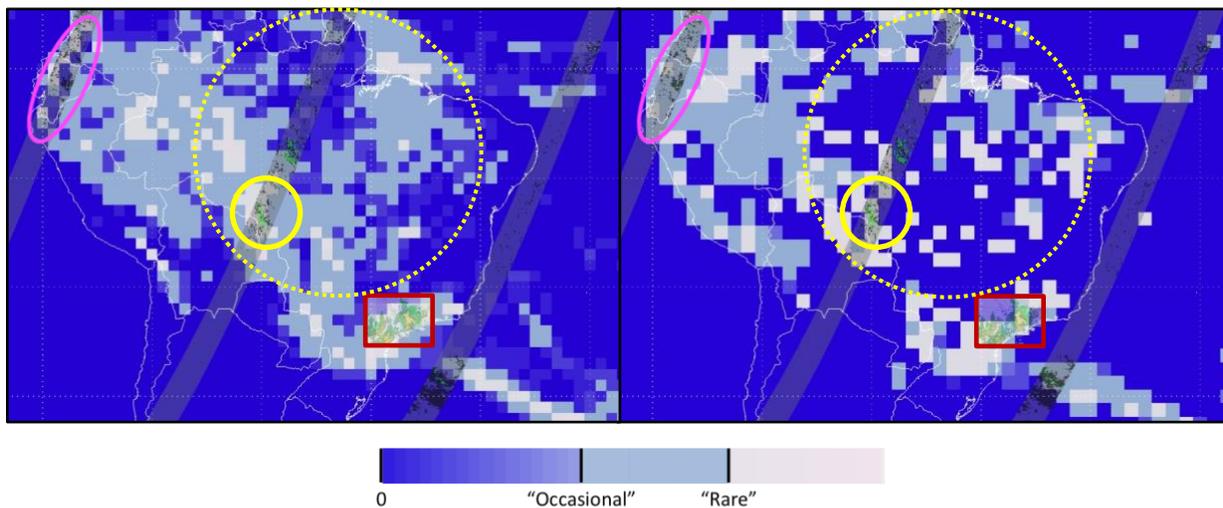
Figure E. The 24-h forecasts of EPOCH thunderstorm likelihood (left) and WAFS Cb horizontal extent (right) valid for 1800-2400 UTC, 20 March 2018. The main color scale differentiates forecast values exceeding the "occasional" (light blue) and "rare" (grey) thresholds. The red box contains an overlay of the 2200 UTC radar image from Figure 31. The swaths show the storm top height observed by the GPM Ku-radar using green to represent tops exceeding 30 kft.

Assessing global forecasts of convection poses challenges, both due to the lack of direct observations and the lack of a commonly held definition of the spatial and temporal extent of convective weather events that are operationally relevant. Methods were derived to address each of these issues for this assessment, although with the consequence that a single interpretation of the results is not always straightforward. Future availability of datasets that serve as robust proxies for global thunderstorm occurrence could help alleviate the difficulties of defining a truth field. Additionally, more detailed input from forecasters can provide direction to help focus subsequent assessments.

1 INTRODUCTION

This document is Part II of a report of the Quality Assessment Product Development Team's (QA PDT) assessment of the Ensemble Prediction of Oceanic Convective Hazards (EPOCH). EPOCH was developed by the Convective Weather Product Development Team (CW PDT) at the National Center for Atmospheric Research (NCAR). EPOCH aims to forecast convection-related hazards to aviation that may affect transoceanic flights, which often need lead times of 24 hours or more for planning. The QA PDT was tasked with an independent quality assessment of the EPOCH product to establish a baseline of performance and inform stakeholders of product skill and characteristics.

EPOCH consists of synoptic, 6-hourly global forecasts of the likelihood of thunderstorm occurrence and the likelihood of convective clouds exceeding 30, 35, and 40 kft, on a 1.0° grid. The EPOCH likelihoods are derived using the Global Ensemble Forecast System (GEFS) and the Canadian Meteorological Centre's Ensemble (CMCE) forecasts of accumulated precipitation, Convective Available Potential Energy (CAPE), and outgoing longwave radiation (OLR) (Stone et al. 2016). The performance and characteristics of EPOCH likelihoods were evaluated relative to the operational thunderstorm (Cb) products of the World Area Forecast System (WAFS), namely WAFS Cb horizontal extent and Cb Tops. The current WAFS Cb products are derived from the deterministic National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model and the UK Met Office Unified Model (UM). The WAFS forecasts are produced on a 1.25˚ grid.

The verification of EPOCH employed a diverse set of observations (Section 2.2), including: two regional radar data sets, which directly measure convective cloud hydrometeors; lightning, which defines thunderstorms; and global satellite cloud and precipitation products, which are linked, though less directly, to the identification of tall thunderstorms. The analysis comprised an evaluation of: (i) characteristics of the products' fields, (ii) statistical performance against truth data sets, (iii) consistency of forecasts between successive forecast cycles, and (iv) performance in specific case studies.

The document is organized as follows. Section 2 provides details of the datasets, including forecast products and observations. The methods and evaluations used in this assessment are described in Sections 3 and 4, respectively. Assessment results are presented in Section 5, while case studies appear in Section 6. Finally, findings are summarized in Section 7. The evaluation will serve to provide a baseline of performance for future versions of EPOCH.

2 DATA

This section describes the data used during the assessment. Stratifications are discussed in Section 3 (Methods). The period of evaluation for Part II of this assessment was 1 December 2017 through 31 March 2018, the convectively active season in the Southern Hemisphere. Results from Part I, which covered the active season in the Northern Hemisphere (1 June 2017 - 30 September 2017), will also be referenced when useful for synthesizing findings across assessment periods.

2.1 FORECAST PRODUCTS

2.1.1 ENSEMBLE PREDICTION OF OCEANIC CONVECTIVE HAZARDS (EPOCH)

The EPOCH likelihood of thunderstorm occurrence is based on the number of GEFS and CMCE ensemble members that exceed specific thresholds for the accumulated precipitation, CAPE, and OLR. Model-specific thresholds were chosen by comparison with the Climate Prediction Center morphing method global precipitation analysis (CMORPH, Joyce et al. 2004) and cloud top height (CTH) (Stone et al. 2016).

The likelihood of convective cloud tops (CCT) exceeding 30 kft (9.144 km) is based on the same fields as the thunderstorm occurrence. Likelihoods are also generated for 35 kft (10.668 km) and 40 kft (12.192 km). These probabilistic CCT products were compared to the WAFS Cb Top in a similar manner as the thunderstorm likelihood and WAFS Cb horizontal extent (Section 3.5.1).

EPOCH forecasts, which are on a 1° grid, are interpreted as representing the 6-hr period prior to the forecast timestamp (CW PDT, personal communication).

Table 1. Summary of forecast product properties.

Forecasts	EPOCH	WAFS Cb
Grid Spacing	1.0°	1.25°
Issuances	0Z, 6Z, 12Z, 18Z	0Z, 6Z, 12Z, 18Z
Lead Times	6, 12, 18, 24, 30, 36, 42, 48	6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36
Fields	• Likelihood of thunderstorm occurrence (0 to 1) • Likelihood of CCT > 30 kft (0 to 1) • Likelihood of CCT > 35 kft (0 to 1) • Likelihood of CCT > 40 kft (0 to 1)	• Cb horizontal extent – extent of Cb cloud in each grid box (0 to 1) • Cb Tops – calculated at flight levels then converted to meters above sea level
Model Input	GEFS & CMCE [Ensemble]	GFS & UM [Deterministic]

2.1.2 WORLD AREA FORECAST SYSTEM (WAFS)

The WAFS produces 1.25° gridded Cb Horizontal Extent and Cb Top forecasts (ICAO, 2012). The horizontal extent predicts the fraction of the sky covered by Cb cloud in a given grid cell. Where Cb clouds are forecasted, heights for the Cb top and Cb base are provided. The WAFS Cb forecast algorithm is based on cloud amount and convective precipitation intensity.

New WAFS Cb forecasts are issued every three hours and are available about five and a half hours after the model forecast issue. To match EPOCH, only the 0, 6, 12, 18 UTC issuances are used for this study. Grid point forecasts are provided for lead times of 6 to 36 hours at 3-hour intervals. In the harmonization of the forecasts between the two WAFCs, the higher value of the two forecasts is taken. These products served as baseline forecasts for comparison with EPOCH.

WAFS forecast grids are to be interpreted as representing ± 1.5 hours around the forecast valid time (AWC, personal communication). However, for this assessment, a window of ± 3 hours is used in order to match

the temporal resolution of EPOCH products; the temporal alignment of the two forecast products used in this assessment is detailed in Section 3.3.2.

2.2 OBSERVATIONS

The observation datasets used in this assessment are listed in Table 2. Despite the very limited oceanic coverage of continental radar networks, they are included in this evaluation because radar (whether ground- or aircraft-based) is the primary data source for the identification of hazardous convection.

Table 2. Observation datasets.

Dataset	Notation
Multi-Radar/Multi-Sensor Echo Top	MRMS-ET
European Max Composite Reflectivity	EU-CR
Ground-based Lightning Detection Networks	LTG
Integrated Multi-satellitE Retrievals for GPM	IMERG
Infrared Anvil Detection	IRA
NASA Cloud Top Height	CTH

2.2.1 MULTI-RADAR/MULTI-SENSOR (MRMS) ECHO TOP

The 18-dBZ echo top (ET), the maximum altitude of the 18-dBZ reflectivity in the vertical column, is derived from the three-dimensional merged reflectivity (Maddox et al. 1999). ETs are used in aviation to identify areas of potentially high turbulence in thunderstorm anvils (Smith et al. 2016). The high-resolution MRMS echo top field, developed for aviation applications over the US, was used as a primary truth set for verification (Figure 1).

Level III Echo Top–18dBZ @ Specific altitude above mean sea level
2017–07–01 1800 UTC

Data Min = –3.0, Max = 52.0
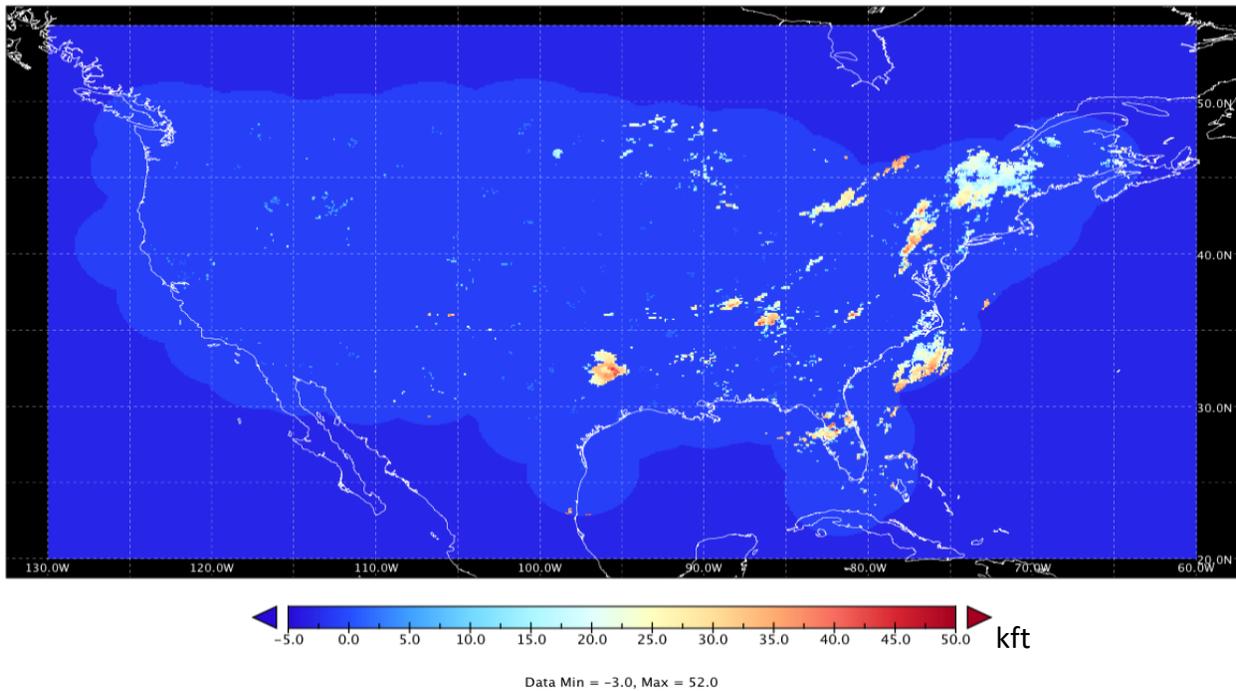
Figure 1. Example of the MRMS Echo Top field on a 0.1° grid over the CONUS.

2.2.2 EUROPEAN MAXIMUM COMPOSITE REFLECTIVITY

European radar reflectivity mosaicked data (EU-CR) were a primary truth set for the assessment. The maximum reflectivity in each column and a quality indicator are provided on 2-km x 2-km grid over Europe, including Iceland (e.g., Figure 2). The convective echoes of interest over Europe were identified by merged composite reflectivity of 35 dBZ or greater. Data was available every 15 minutes (00, 15, 30, 45 of each UTC hour) and were provided by MeteoFrance.

Figure 2. Example of the European radar mosaic of maximum composite reflectivity for 1 July 2017 at 1800 UTC.

4

2.2.3   GLOBAL PRECIPITATION MEASUREMENT (GPM) KU-RADAR

Global Precipitation Measurement Dual-frequency Precipitation Radar (GPM DPR) Ku-band storm top height was used in this assessment for evaluating EPOCH products in case studies. The GPM Ku-band has a swath width of 245 km, a spatial resolution of ~5.2 km at nadir and ~5.6 km at swath edge, and a vertical resolution of 250 m (Iguch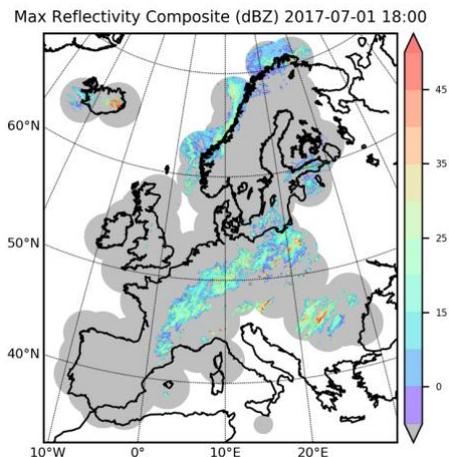i et al. 2010). The DPR Storm Top Height measures the altitude of the highest range bin that contains precipitating echoes, where radar reflectivity is required to exceed a given threshold for six range bins in succession, to remove noise.

The GPM Ku-radar coverage is global, as its host GPM Core Observatory satellite makes 16 orbits around the earth each day, with an orbit inclination of 65 degrees. However, while the GPM Ku-radar storm top height has a high probability of detecting tall thunderstorms, it misses events because of its narrow swath and the extremely short period of time it covers any given location. For this reason, the GPM storm top height data was used for case studies only.

2.2.4   BRAZILIAN RADAR IMAGES

Images of CAPPI (Constant Altitude Plan Position Indicator) reflectivity from two single-site radars in the vicinity of Sao Paulo, Brazil, were used for the South America case study presented in this assessment. Radar data was retrieved from the SOS CHUVA (**S**ystem of **O**bservation and Forecast of **S**evere Weather, **C**loud processes of t**H**e main precipitation systems in Brazil: A contrib**U**tion to cloud resol**V**ing modeling and to the Glob**A**l Precipitation Measurement) database. Due to the limited geographic extent of the radar coverage, statistics were not computed against this dataset.

2.2.5   LIGHTNING

The production of lightning (LTG) distinguishes thunderstorms from other clouds, making lightning strike counts (Figure 3) useful for verifying thunderstorm occurrence. Lightning data are the primary data used for WAFS Cb verification by WAFC-London.
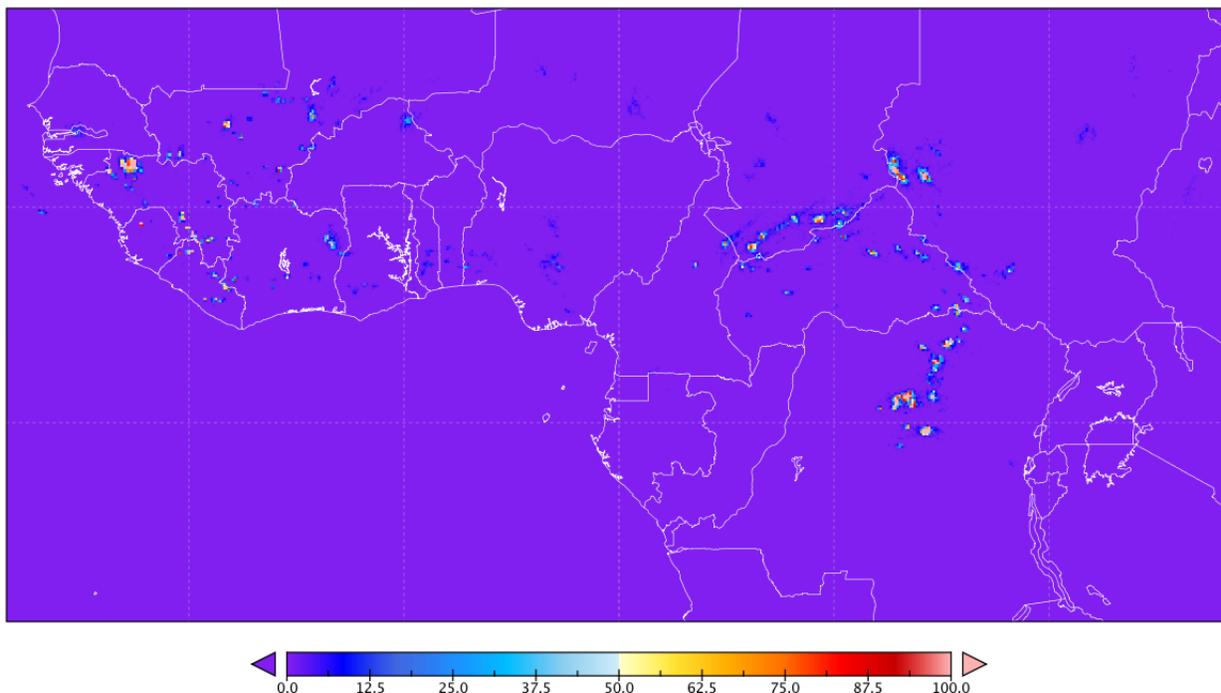
Figure 3. Lighting strike count for 1800 to 1829 UTC, 1 July 2017 over Tropical Africa.

The detection of lightning from distant places is based on the propagation characteristics of very low frequency (VLF, 3–30 kHz) waves generated by lightning discharges. Because VLF waves propagate through the Earth-ionosphere waveguide with relatively low attenuation, they can be detected at great distances from the lightning discharge (Said et al. 2013, Mallick et al. 2014).

A combination of data from the Vaïsala National Lightning Detection Network (NLDN; cloud-to-ground lightning in continental US and vicinity) and Earth Network (ENTLN; total lightning for the globe) were used as a truth set for thunderstorm occurrence. However, those data will be used primarily to determine a bound on the POD because of the low detection efficiency outside of North America and Europe. In addition, lightning within a neighborhood of 0.25° (LTG25) was used to provide a less conservative bound.

2.2.6 INTEGRATED MULTI-SATELLITE RETRIEVALS FOR GPM (IMERG)

The Integrated Multi-satellitE Retrievals for GPM (IMERG) precipitation rate is a combination of all satellite microwave precipitation estimates, infrared satellite precipitation estimates, and precipitation gauge analyses (Huffman et al. 2015). IMERG precipitation (Figure 4) was chosen for verification of EPOCH because it has global extent, relatively high spatial (0.1 deg) and temporal resolution (30-minute), and information from multiple sensors. IMERG "Late" run half-hourly data, calibrated with climatological gauge data with a latency of a few days, were used in this assessment.
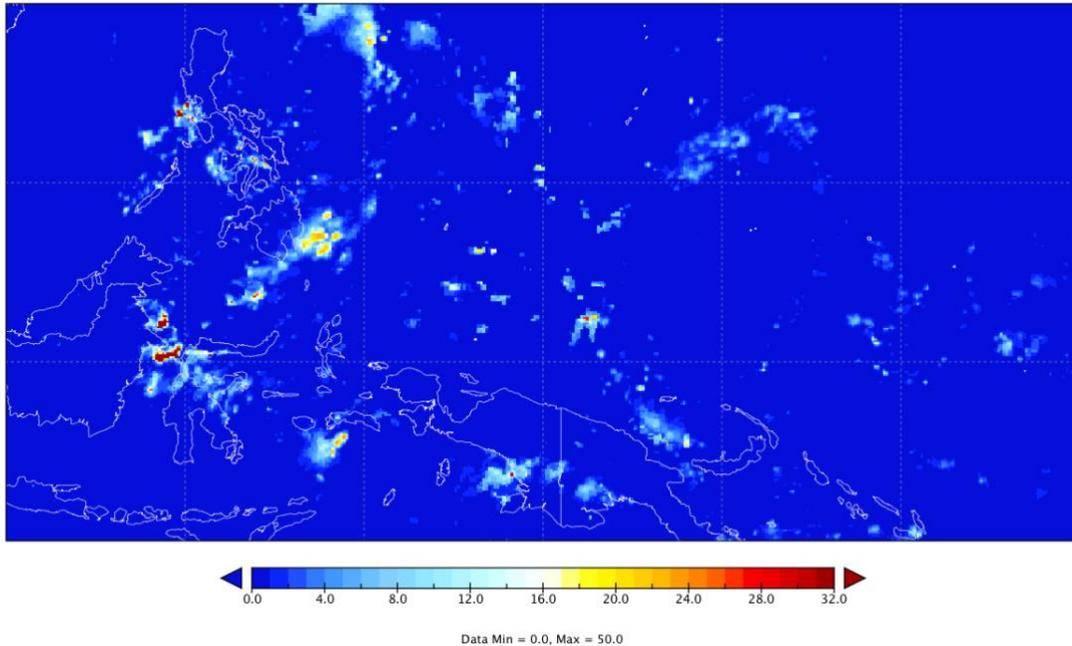
Figure 4. Map showing IMERG precipitation rate for 1700-1729 UTC, 1 July 2017.

From a global perspective, there is a weak correlation between the heaviest rainfall and the tallest storms (Hamada et al. 2015). However, convective systems typically have heavier rain rates than stratiform systems, therefore higher precipitation thresholds can be treated as useful proxies for identifying thunderstorms. Section 4.2.3 has more details of IMERG use in verification.

2.2.7   CLOUD TOP HEIGHT

Cloud top height (CTH) and cloud bottom height (CBH) data, received once per hour in near-real time at NOAA/ESRL/GSD, are derived from geostationary satellite images by NASA Langley Research Center (Minnis et al. 2008). The classification of pixels into cloud or clear fields is based on the surface temperatures, clear-sky radiances, surface types, and spectral atmospheric corrections in each tile. Various surface, aircraft, and other satellite datasets have been used to validate the accuracy of satellite-derived cloud top products (Mace et al. 1998, Min et al. 2004, Khaiyer et al. 2008).
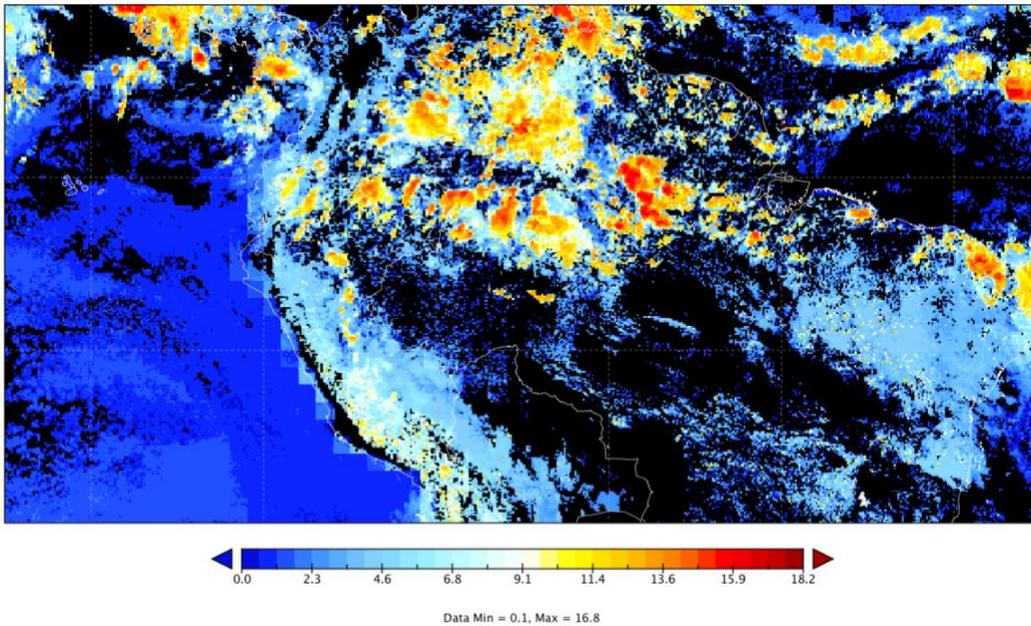
Figure 5. Cloud top height (km) derived from GOES-13 and assigned to the hour 1700-1759 UTC, 1 July 2017. CTH exceeding 30 kft are in warm colors.

Snapshots from multiple geostationary satellites were combined to create a global mosaic. The spatial resolution of the snapshots increased during the assessment period. For the cases where satellites came online during the assessment period, data from the new satellite were used as soon as available. Specific coverage periods used in this assessment are detailed in Table 13 (Appendix A).

Data were available on the hour, except for GOES-13 images, which were available at 45 minutes after the hour. To create an hourly global mosaic, a GOES-13 image was combined with images from the remaining satellites available at the next hour (a 15-minute difference between file times). Observations were assigned to the subsequent hour, e.g., values at 1700 UTC were assigned to 1700-1759 UTC (Figure 5).

The CTH field detects all convective cloud top events, resulting in high detection rates; however, it does not distinguish between convective and non-convective clouds. A filtered cloud top height field requiring a minimum "cloud depth" of 15 kft (CTH15) was used to provide a less conservative bound. "Cloud depth" was computed as the difference between "cloud top height" and "cloud bottom height".

2.2.8 INFRARED (IR) ANVIL DETECTION

Overshooting tops (OTs) are produced by convective updrafts that penetrate the surrounding cirrus anvil and rise through the tropopause. Therefore, there is high confidence in the ability of the OT data to detect thunderstorms, with few false alarms. The spatial and temporal distributions of OTs have been found to match well with those of lightning activity (e.g., Jurkovic et al. 2015).

Infrared (IR) anvil areas (color-shaded areas in Figure 6) are defined about OT cold spots by sending out rays in 16 directions. The brightness temperature gradient is monitored along each ray, and if the temperature either increases or decreases abruptly, then the ray is truncated at the spot (Bedka and Khlopenkov 2016). The "IR anvil detection" field (IRA), which identifies the surrounding thunderstorm

8

anvils, was used for storm detection rather than the "OT probability", which identifies only the strongest, and often short-lived, convective cores.
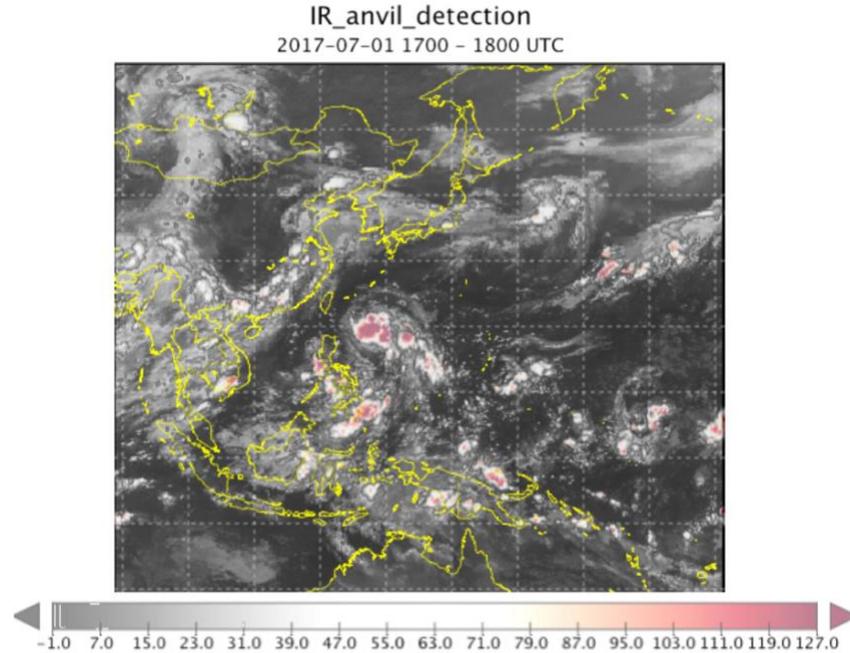

Figure 6. Examples of IR Anvil Detection for 1700-1800 UTC 1 July 2017, overlaid on grey-scale IR image, 1800 UTC 1 July 2017. Color shades are for confidence rating > 80.

3 METHODS

The EPOCH product was evaluated at different thresholds and over different geographic regions. Additionally, the quality of the EPOCH forecasts was compared with that of WAFS Cb products for the same forecast periods. The application of observation data to verifying EPOCH likelihoods and WAFS Cb forecasts is described below.

3.1 GRID DEFINITION

All comparisons between forecast products and observations were made with data placed on the grid used by EPOCH, which has a resolution of 1° x 1° x 6-hr. The valid time of EPOCH represents the end of the 6-hr period. The grid points lie at integer degrees and are considered to be at the center of a 1° x 1° grid cell. The WAFS Cb product and the observations were placed onto this grid using techniques described in subsequent sections.

3.2 FORECAST EVENT DEFINITION

To compute statistics, forecasts were converted to binary "Yes" or "No" events using a range of thresholds that span the relevant forecast values. For the EPOCH probabilistic fields, thresholds that gave the closest match to the event frequency from specific WAFS Cb thresholds were also included.

3.3 FORECAST ALIGNMENT TECHNIQUES

3.3.1 SPATIAL ALIGNMENT

The EPOCH product is available on a grid with 1.0° spacing, while WAFS Cb product is issued using a 1.25° grid spacing. To align the two products, the WAFS Cb forecast fields are re-gridded to 1.0° spacing using a nearest neighbor method. The choice of re-gridding direction is motivated by the resolution of the observation datasets used in this assessment, in particular the IMERG precipitation data, which upscales cleanly to 1.0° but not 1.25°.

3.3.2 TEMPORAL ALIGNMENT

The EPOCH product is issued with 6-hourly lead times, with the forecasts representing the 6 hours preceding the valid time. The WAFS Cb product is issued with 3-hourly lead times, with forecasts representing the 3-hr period centered on the valid time. Due to the mismatch in the valid period start and end times, there is no clean way to align the two products. In this assessment, EPOCH and WAFS lead times which yield the same center point of the valid interval were matched. For example, a WAFS lead of 9 hours was matched with an EPOCH lead of 12 hours, which corresponds to the period from 6-12 hours after issuance for EPOCH and 7.5 to 10.5 hours for WAFS, both of which are centered on 9 hours. The alignment of additional lead times is provided in Table 3 and illustrated in Figure 7. Since the WAFS forecast valid period only covers a 3-hr window, the number of forecast events should be lower than if the WAFS valid period covered the entire 6-hr window, which would lead to a reduction in correct detections and potentially also fewer false alarms.

Table 3. Alignment of EPOCH and WAFS forecast valid periods (hours).

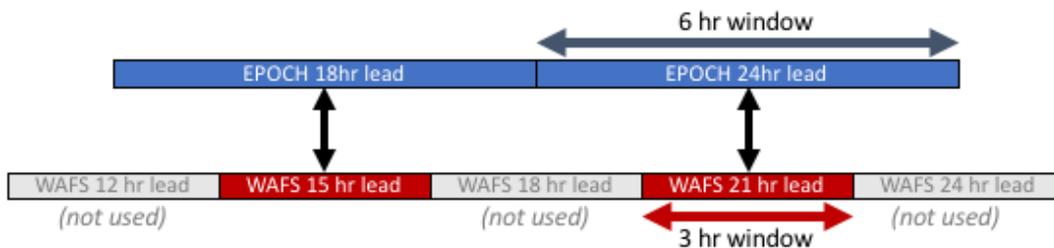| WAFS Lead time	EPOCH Lead time	EPOCH Valid Period	WAFS Valid Period
9	12	6 – 12	7.5 – 10.5
15	18	12 – 18	13.5 – 16.5
21	24	18 – 24	19.5 – 21.5
27	30	24 – 30	25.5 – 28.5
33	36	30 – 36	31.5 – 34.5

Figure 7. Temporal alignment of the EPOCH and WAFS forecast lead times.

3.4 FORECAST/OBSERVATION PAIRING TECHNIQUES

This section describes the pairing of model output to gridded observations. All radar and satellite data were re-gridded (up-scaled) to match the EPOCH grid. The result is a set of binary observations (i.e., yes/no) for each observation type.

The original spatial and temporal resolution of each observational dataset (Table 4) is much finer than the resolution of the forecast product (1° x 1° x 6-hr). The up-scaling required to pair observations with the forecast was accomplished in two steps: (1) placing each observation dataset on a common grid spacing, and (2) generating binary observation values using a range of criteria for defining an event, based on the intensity, spatial coverage, and temporal extent of weather present. The specific techniques employed for step (1) are described in this section while the approach for step (2) are described in section 3.5.1.

Table 4. Summary of observational data resolution.

Observation Dataset	Horizontal Spacing	Frequency
MRMS-ET	0.01°	10-minute
EU-CR	2 km	15-minute
LTG	Point data	Point data
IMERG	0.1°	30-minute
IRA	5 km	60-minute
CTH	8-10 km	60-minute

3.4.1 COMMON GRID FOR OBSERVATIONS

To allow for consistent upscaling to the EPOCH grid, each observation dataset was placed onto a common grid with 0.1° spacing. Datasets with high temporal resolution were also up-scaled to 30-minute data. The input data was required to be present for 100% of the spatial and temporal extent of the output grid cell; otherwise, a missing value was assigned to the output grid cell. The specific methodology used for each dataset is described below.

3.4.1.1 GPM IMERG

The common grid was chosen to match the native resolution of the IMERG product (0.1° every 30 minutes), therefore no processing was required for this dataset.

3.4.1.2 LIGHTNING

The lightning data (LTG) was available as point data. As in Part I of the assessment, counts of lightning strikes were aggregated over each 30-minute period within each 0.1° grid box. In addition, lightning strikes within a neighborhood of 0.25° (LTG25) were tabulated using the same common grid.

3.4.1.3 GROUND-BASED RADAR

The MRMS Echo Top (MRMS-ET) and European Max Composite Reflectivity (EU-CR) both have spatial and temporal resolution finer than the 0.1° and 30 minutes used for the common observations grid. These

datasets were up-scaled using the 90th percentile of all values contained within each 0.1° grid cell during each 30-minute interval.

3.4.1.4 SATELLITE CLOUD PRODUCTS

The NASA Cloud Top Height (CTH) and IR Anvil Detection (IRA) products were re-gridded to the common observations grid using a nearest-neighbor method. The original hourly temporal resolution was retained. The filtered cloud top height data (CTH15) was placed on the common grid using the same methodology.

Table 5. Intensity thresholds for observation datasets.

Dataset	Intensity Thresholds	Primary Threshold
MRMS-ET	30, 35, 40 kft	30 kft
EU-CR	35, 40, 45, dBZ	40 dBZ
IMERG	1, 10 mm/hr	10 mm/hr
IRA	20, 40, 80	80
LTG	1, 10 strikes	1 strike
LTG25	10 strikes	10 strikes
CTH	30, 35, 40 kft	30 kft
CTH15	30, 35, 40 kft	30 kft

3.4.2 OBSERVED EVENT DEFINITION

Data on the common observation grid were converted to binary "Yes" or "No" events using a range of combined thresholds for the intensity, spatial coverage, and temporal extent of the weather present. The thresholds were selected in consultation with the Aviation Weather Center (AWC).

An appropriate range of intensity thresholds was explored for each observation type (Table 5).

The spatial and temporal thresholds used were consistent for all observation datasets, specifically:

- Spatial Coverage: 5, 10, 25% of the 1° x 1° box
- Temporal Extent: 1, 2, 4 hours of the 6-hr interval

Observed data are included when there is full data availability on a grid-cell basis, specifically:

- All 0.1-deg sub-cells within the 1-deg cell have valid data
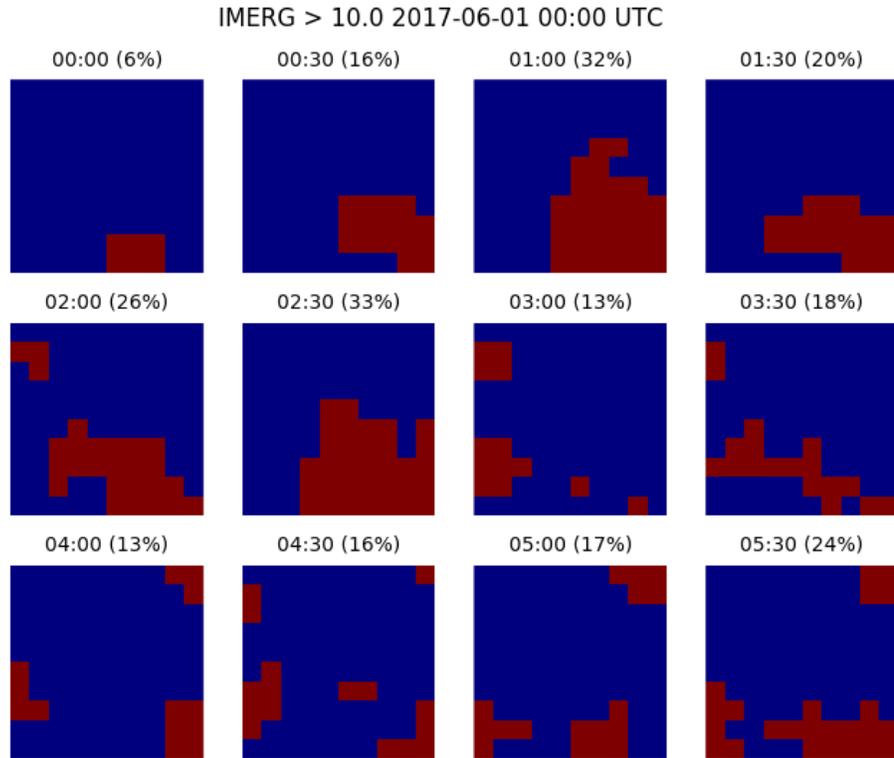- All sub-periods within the 6-hr interval have valid data

Figure 8. Example of observational data contained within a single 1° x 1° x 6-hr forecast grid cell, using IMERG data with a threshold of 10 mm/hr (red). Each panel represents a 30-minute period and contains 100 sub-cells (0.1°). The heading of each panel gives the observation timestamp as well as the spatial coverage as a percentage of the 1° grid cell.

An illustration of the amount of observational information contained within a single forecast grid cell is given in Figure 8. Each 1° x 1° cell contains 100 sub-cells (0.1°), and each 6-hour period contains 12 sub-intervals (30-minute). In this example, the spatial coverage of IMERG ≥ 10mm/hr varies from 6% to 33% during the 6-hour period. The assignment of a "Yes" or "No" event label to this case is dependent on the spatial and temporal thresholds chosen, as illustrated in Table 6.

Table 6. Example of converting observations to binary events using the IMERG data shown in Figure 8.

Spatial Coverage Threshold	Temporal Extent Threshold	Example considered "Yes" event?
5%	1-hr	Yes
5%	4-hr	Yes
25%	1-hr	Yes
25%	4-hr	No

Using this framework for defining observed events allows different spatial and temporal scales to be explored independently. Note that contiguity is not required to meet the spatial or temporal thresholds; for example, the 26% spatial coverage at 0200 UTC consists of two distinct objects.

3.5 STRATIFICATIONS

Results were stratified according to forecast thresholds, geographic region, issuances, and lead times.

3.5.1 FORECAST THRESHOLD STRATIFICATIONS

For each forecast variable, a range of thresholds was used to span the relevant values. Values from 0 to 1 were used to generate Receiver Operating Characteristic (ROC) curves for the EPOCH fields and WAFS Cb Horizontal extent, while thresholds of 30, 35, 40 kft were selected for the WAFS Cb Tops field. In the case of the EPOCH thunderstorm likelihood and WAFS Cb horizontal extent fields, additional thresholds were chosen based on frequency matching of the forecast distributions. Matching thresholds for each altitude (30, 35, and 40 kft) were also selected based on frequency matching between the EPOCH CCT fields and the distribution of the WAFS Cb tops.

3.5.2 GEOGRAPHIC STRATIFICATIONS

The assessment was composed of global statistics as well as regional stratifications based on World Meteorological Organization (WMO) and WAFC designated areas (Figure 9, top), as well as land and sea regions (Figure 9, bottom). Differences in the prevalence of intense convection by latitude (i.e., between 20°-40° and 40°-60°) were also investigated. The regions are listed in Table 7 and their geographic boundaries are defined in Table 14 (Appendix A).

Table 7. Regional focus areas used in this assessment (shown in Figure 9).

Part I Regions		
Global		
Tropics		
Northern Hemisphere Extratropics (NH ExTrop)		
Southern Hemisphere Extratropics (SH ExTrop)		
North Atlantic (NAtlantic)		
North America (NAmerica)		
North Pacific (NPacific)		
Asia		
Australia/New Zealand (AusNZ)		

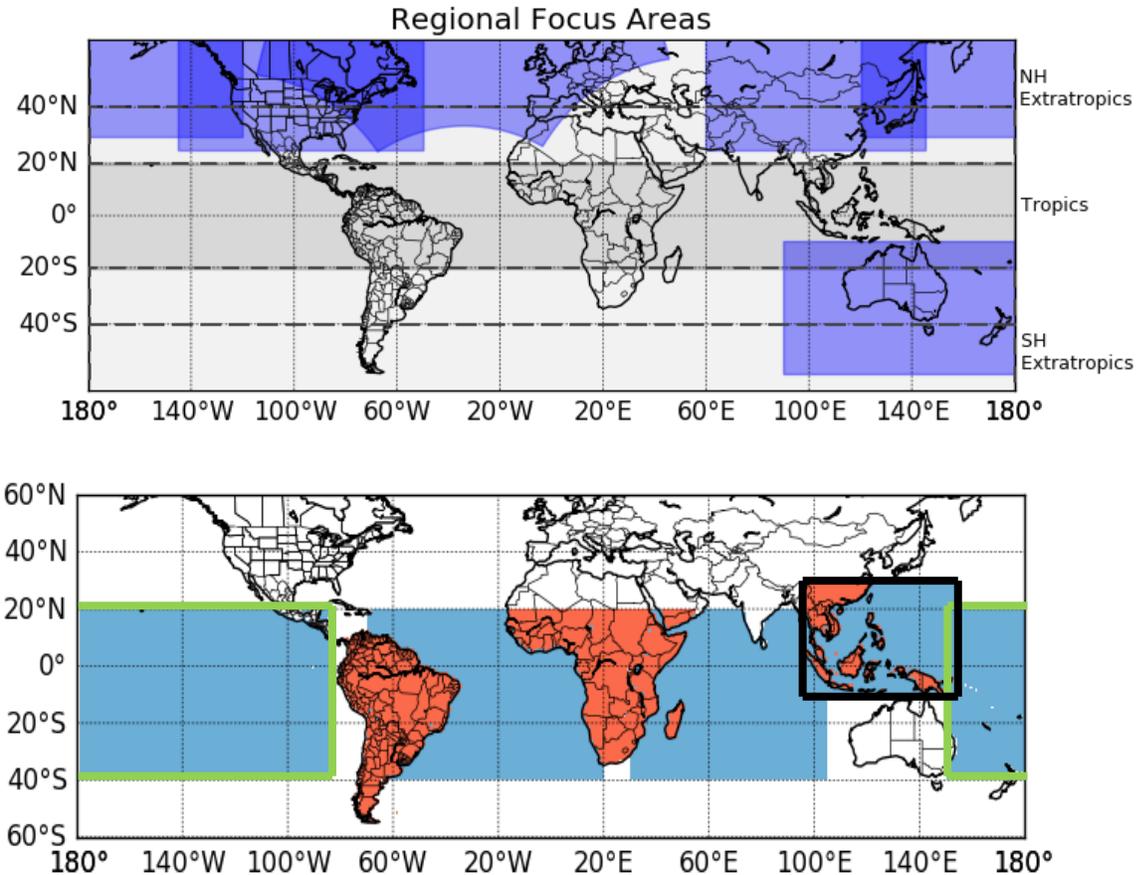| Part II Land Regions	Part II Sea Regions
Land	Sea
South America (SAmerica-L)	Southern Pacific (SPacific-S)
Sub-Saharan Africa (SAfrica-L)	Southern Atlantic (SAtlantic-S)
Indo-Australian Archipelago (SE Asia-L)	Indian Ocean (SIndian-S)
	Maritime Southeast Asia (SEAsia-S)

Figure 9. Regional focus areas defined in Part I (top, indigo) and Part II (bottom). The dashed lines (top) delineate latitude bands. The red and blue shading (bottom) indicate regions defined using land and sea masks, respectively. The boxes (bottom) highlight the boundaries of the southern Pacific (green) and southeast Asia (black) regions, which have a small area of overlap.

3.5.3   TEMPORAL STRATIFICATIONS

The skill of the forecast products was assessed over a range of operationally relevant lead times (EPOCH leads of 12, 18, 24, 30, 36 hrs; see Table 3 for aligned WAFS leads). Skill scores were also evaluated for each issuance time (0, 6, 12, 18 UTC).

4   EVALUATIONS

This assessment evaluated the EPOCH product, using the WAFS Cb product as a baseline. Both products were evaluated for field distributions, climatological maps, and statistical measures, where the specific metrics used varied according to observation set, as described in Section 4.2.

4.1   FIELD CHARACTERISTICS

4.1.1   FORECAST FIELD DISTRIBUTIONS

Distributions were generated for the EPOCH and WAFS Cb fields. The native bin spacing of each forecast field (Table 8) was retained, with no additional binning performed. The counts of forecast values by bin were normalized by the total number of forecasts to create frequencies. Cumulative distributions were also generated and matched with the goal of removing any effects of different biases in the two products. Given

the bin spacing of the forecast values, particular EPOCH and WAFS thresholds were chosen based on minimizing the difference in the resulting frequencies.

Table 8. Native bin spacing of forecast fields.

Forecast Field	Native Bin Spacing
EPOCH (all)	2.5%
WAFS Cb Extent	10%
WAFS Cb Tops	5 kft

4.1.2 NORMALIZED EVENT FREQUENCIES

Forecast and observed "Yes" event frequencies were computed for each forecast field as well as each observation type and definition (given by the combined intensity, spatial coverage, and temporal extent thresholds). Counts were converted to frequencies by normalizing by the total number of events ("Yes" or "No") at each grid point in the assessment period. For the forecasts, a particular valid time interval was only included if both EPOCH and WAFS products were available. For the observational datasets, individual grid points may have variable number of events, due to partial coverage outages or other data issues.

4.1.3 CLIMATOLOGICAL MAPS

Maps showing the geographic distributions of both forecasts and observations were generated using the normalized event frequencies. Climatological maps of EPOCH thunderstorm likelihood and CCT exceeding selected thresholds were compared with the current WAFS Cb horizontal extent and Cb tops, respectively. Difference maps were created by subtracting one forecast climatological map from another to highlight relative biases. Global and regional maps of observations serve as additional checks on the forecasts and illustrate the variation among observation types.

4.2 STATISTICS CALCULATED USING TRUTH DATA SOURCES

Terminology and score definitions referenced throughout the assessment are defined in Table 9. The use of particular scores was tailored by observation type, as summarized in Table 10. A brief discussion of the use of each observation type follows.

4.2.1 GROUND-BASED RADAR

Ground-based radar is the current standard for thunderstorm identification. In this assessment, observations from MRMS Echo Top and European Maximum Composite Reflectivity were therefore treated as primary truth datasets. The EPOCH likelihood of thunderstorm occurrence and WAFS Cb horizontal extent forecasts were compared to both radar datasets over their respective domains. MRMS echo tops exceeding 30, 35, and 40 kft were also compared with EPOCH likelihood of CCT and WAFS Cb top height at the corresponding altitudes.

16

Probability of Detection (POD)	Proportion of all observed events that are correctly forecast to occur, where in this case, an event is thunderstorm existence or cloud tops exceeding specific thresholds.
Probability of False Detection (POFD)	Proportion of all observed non-events that are mistakenly forecast to be events.
Success Ratio (SR)	Proportion of all forecasted events in which an observed event did occur.
Critical Success Index or Threat Score (CSI)	Proportion of all observed events and/or forecast events that were correctly predicted. It measures how accurate the forecast is when correct negatives have been removed from consideration.
Bias	Measures the ratio of the frequency of forecast events to the frequency of observed events.
Receiver Operating Characteristic (ROC)	Measures the ability of the forecast to discriminate between events and non-events. The POD is plotted relative to POFD (probability of false detection or false alarm rate) conditioned on observations.
Reliability	Measures the average agreement between the forecast values and the observed values. If all forecasts are considered together, then the *overall reliability* is the same as the *bias*. If the forecasts are stratified into different ranges or categories, then the reliability is the same as the *conditional bias*, i.e., it has a different value for each category.

4.2.2 LIGHTNING

As noted previously, lightning indicates the presence of a storm that should be captured by a forecast product, but the absence of lightning does not preclude the presence of convection. Therefore, lightning was used to evaluate skill of storm detection, and products are not penalized for forecasting convection where no lightning was detected. The EPOCH forecast likelihood of thunderstorm occurrence and WAFS Cb horizontal extent were therefore compared to the binary lightning field, with a focus on POD. The lightning within a neighborhood (LTG25) was used in a similar manner.

Table 10. Statistical measures by observation type.

Usage in Assessment	Statistical Measures	Observation Datasets
Primary Truth	POD, SR, CSI, Bias, ROC, Reliability	MRMS-ET EU-CR
Lower Bound	POD	LTG, LTG25 IMERG IRA
Upper Bound	SR	CTH, CTH15

4.2.3 IMERG

Due to the poor correlation between IMERG and the tallest thunderstorms, and the inclusion of non-convective precipitation, the accumulated precipitation data was used in a manner similar to the lightning for comparison with EPOCH and WAFS Cb forecast products. The CW PDT used an accumulated precipitation value of 4 mm per 6 hours in their calibration. In this assessment, a threshold of 10 mm/hr was used in an attempt to exclude non-convective events. Since many convective events may not reach this threshold, the IMERG data is used as a bound on forecast skill, with an emphasis on storm detection (POD).

4.2.4 IR ANVIL DETECTION

The IR Anvil Detection (IRA) product identifies strong updrafts that occur with deep convection that penetrates the tropopause and are surrounded by broad thunderstorm anvils. Thus, if an event is detected by the IRA algorithm, it should be present in the EPOCH thunderstorm forecast and in the corresponding WAFS Cb horizontal extent. The IRA field was used to verify the EPOCH and WAFS thunderstorm forecasts skill, with a focus on storm detection (POD).

4.2.5 CLOUD TOP HEIGHT

The CTH was used as an upper bound on the EPOCH and WAFS Cb forecasts. The CTH field is expected to capture both convective and non-convective elements. Therefore, a forecast of convection should not fall outside the area identified by CTH, but it is not possible to determine whether forecasts within the CTH field are correct. Consequently, CTH data was used to compute a bound on the Success Ratio (SR). The filtered cloud top height data (CTH15) was also treated as an upper bound.

4.3 CONSISTENCY

The forecast products were compared across successive forecasts to determine the consistency in the predicted values from one issuance to the next.

4.4 CASE STUDIES

Case studies for specific events were also part of the analysis. The events included in this report represent a Southern Hemisphere case and a Northern Hemisphere case that provides an example of an aircraft incident associated with thunderstorms.

5 RESULTS

The results of the comparison of the EPOCH and WAFS Cb products for the 4-month Southern Hemisphere summer assessment period (Dec 2017 – Mar 2018) are detailed in the subsequent sections.

5.1 FIELD CHARACTERISTICS

5.1.1 FORECAST FIELD DISTRIBUTIONS

The frequency distributions of EPOCH thunderstorm likelihood and WAFS Cb extent are given in Figure 10. (Note that all frequency plots in this report are presented using a log scale.) The shape of the distributions is notably different: the frequency of EPOCH likelihood was strictly decreasing, while the WAFS Cb extent had a strong peak in the 40% bin. Recall that the WAFS Cb and the EPOCH fields are defined differently. For instance, a WAFS Cb grid cell with a horizontal extent of 0.3 is a deterministic forecast that 30% of that grid cell will be covered by Cb clouds. An EPOCH thunderstorm likelihood of 0.3 in a grid cell means that 30% of the ensemble members reached the thunderstorm algorithm threshold for that grid cell (i.e., it contains no information about the coverage of thunderstorms within the cell). For the purposes of constructing ROC curves and reliability diagrams, the WAFS coverage thresholds were treated as probabilities.
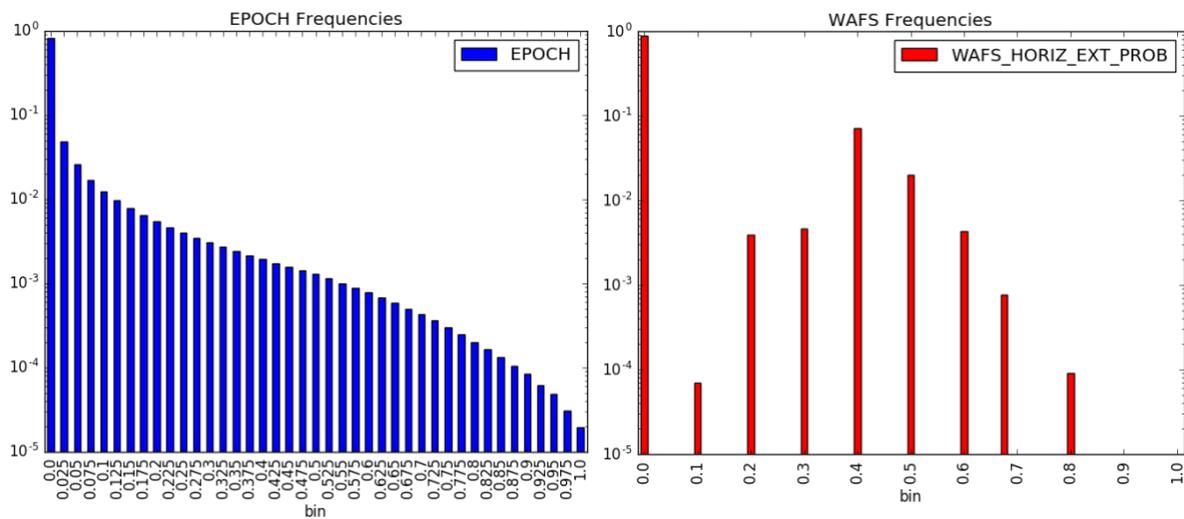
Figure 10. Frequency distribution of EPOCH thunderstorm likelihood values (left) and WAFS Cb horizontal extent (right) for the Southern Hemisphere summer period. Frequencies are presented using a log scale.

The cumulative distributions for both EPOCH and WAFS, showing the frequency of forecast values exceeding each threshold, are presented in Figure 11. Equalized thresholds, selected to yield the same event frequency in the two forecast products, are given in Table 11 and also illustrated in the figure. Note that the specific forecast thresholds and event rates that resulted from this bias-equalization process changed slightly compared with Part I (see Table 15 in Appendix A).
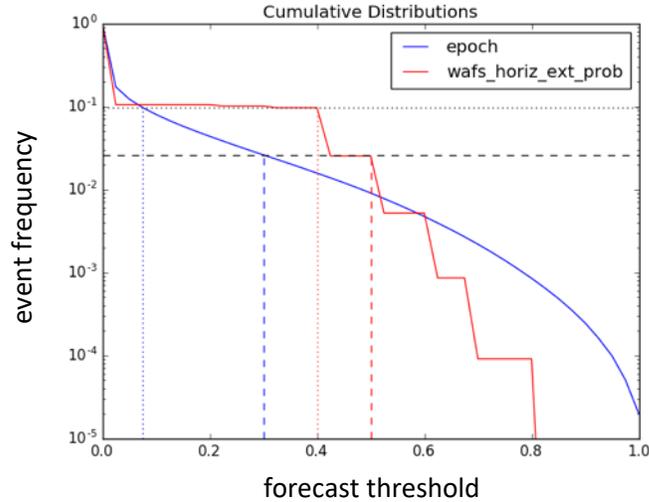
Figure 11. Cumulative distributions for EPOCH thunderstorm likelihood (blue) and Wafs Cb extent (red). Equalized forecast thresholds yielding "occasional" (10.8%, dotted) and "rare" (2.6%, dashed) event frequencies are also shown.

Table 11. Equalized forecast thresholds for the Part II assessment period (Dec 2017 – Mar 2018).

Terminology	EPOCH Threshold	WAFS Threshold	EPOCH Frequency	WAFS Frequency
"Occasional"	7.5%	40%	9.68%	9.61%
"Rare"	30%	50%	2.60%	2.53%

5.1.2 OBSERVED EVENT FREQUENCIES

Observed event frequencies depend on the definition of a relevant weather event. As described in Section 3.4.2, a range of intensity, spatial coverage, and temporal extent thresholds were used for each observation type. The following shorthand notation is used throughout the assessment to refer to a specific combination of these thresholds:

v(intensity threshold)_x(spatial coverage threshold)_t(temporal extent threshold)

As an example, an observation definition of ***v30_x5_t1*** for MRMS Echo Top data indicates that ***30 kft*** is exceeded for at least ***5%*** of the 1-deg grid box for at least ***1 hour*** of the 6-hr interval.

The frequency of "Yes" events for each dataset and observation definition are displayed in Figure 12 for the same datasets as used in Part I, and Figure 13 for the newly introduced LTG25 and CTH15 datasets. As expected, the intensity thresholds chosen for the upper (lower) bound datasets resulted in frequencies below (above) the "occasional" (dotted line) and "rare" (dashed line) event frequencies. For example, the 10 mm/hr IMERG threshold (which occurs less than 2% of the time for all space-time combinations; Figure 12 middle right panel) was used a lower bound, while the CTH thresholds (Figure 12 lower right panel) had an event frequency of 11% or more, with the lone exception of v40_x25_t4, consistent with usage as upper bounds.
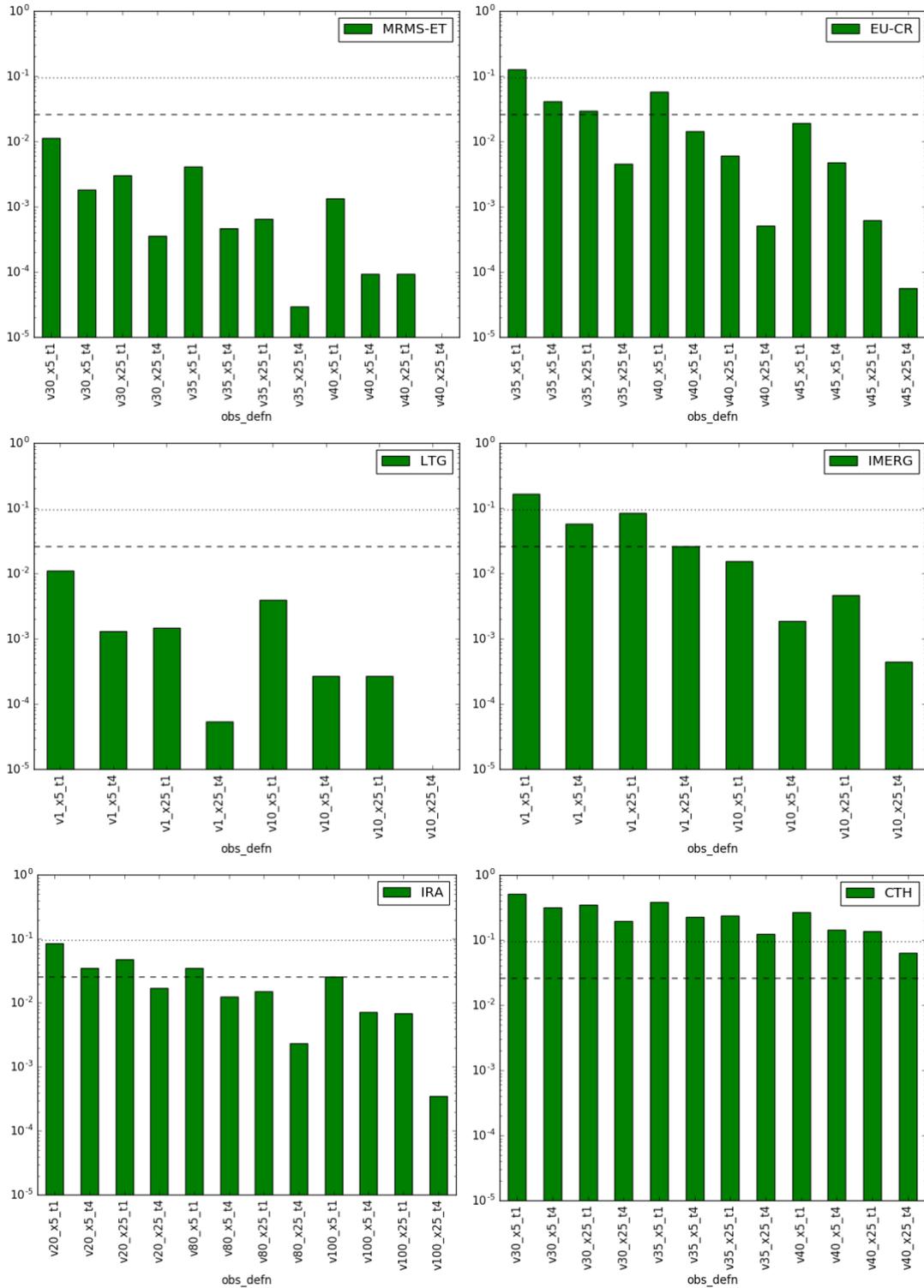
Figure 12. Observed event frequencies for each observation dataset using a range of observation definitions. For reference, the "occasional" (dotted) and "rare" (dashed) frequencies corresponding to the equalized forecast thresholds are also shown.
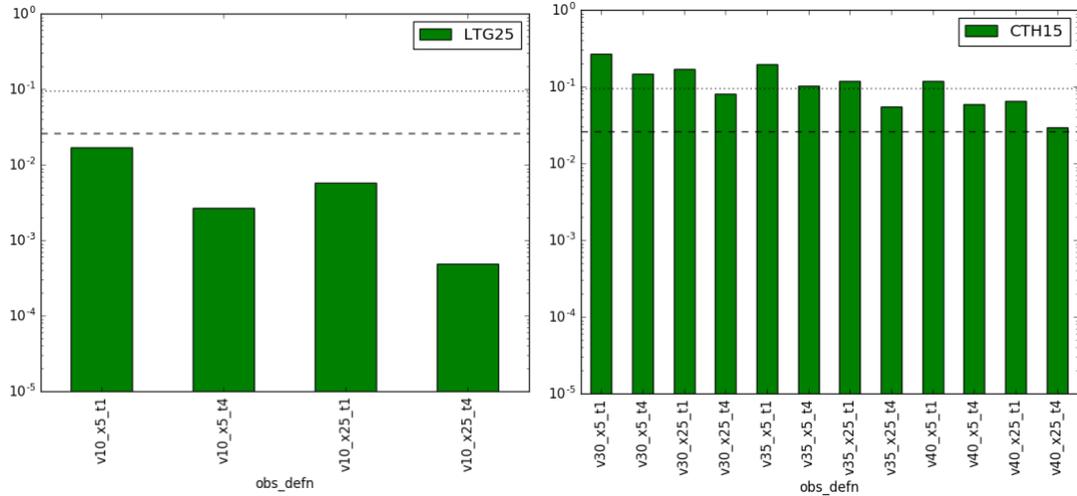
Figure 13. Same as Figure 12, for the neighborhood lightning (LTG25) and filtered cloud top height (CTH15) datasets.

5.1.3 CLIMATOLOGICAL MAPS

Forecast climatological maps using the two event-equalized thresholds are shown in Figure 14. EPOCH and WAFS both captured the large-scale climatological features such as the ITCZ and sub-tropical highs. Compared with the Part I results, a southward displacement of the ITCZ was observed, consistent with the change in season. Since the forecast thresholds were chosen such that global event frequencies were equalized across EPOCH and WAFS, the two products had the same global average bias. However, there were notable regional differences. As seen in Part I, EPOCH forecast more events over the tropical oceans, while WAFS forecast more in the higher latitudes and over most land regions. At the "occasional" forecast threshold (Figure 14, left), EPOCH placed more events over central Africa and the Amazon region, while WAFS forecast more events over some mountainous areas, particularly the Andes and Indonesia. The tendency of WAFS to place more events over regions of higher terrain is consistent with the findings for Part I, although the regions where this signal was observed varied by season. Specifically, a stronger signal was seen over the Andes and Indonesia during the SH Summer, while the signal over the Himalayas was reduced compared with the NH Summer assessment period, consistent with the general southward shift of peak convective activity associated with the change in season.
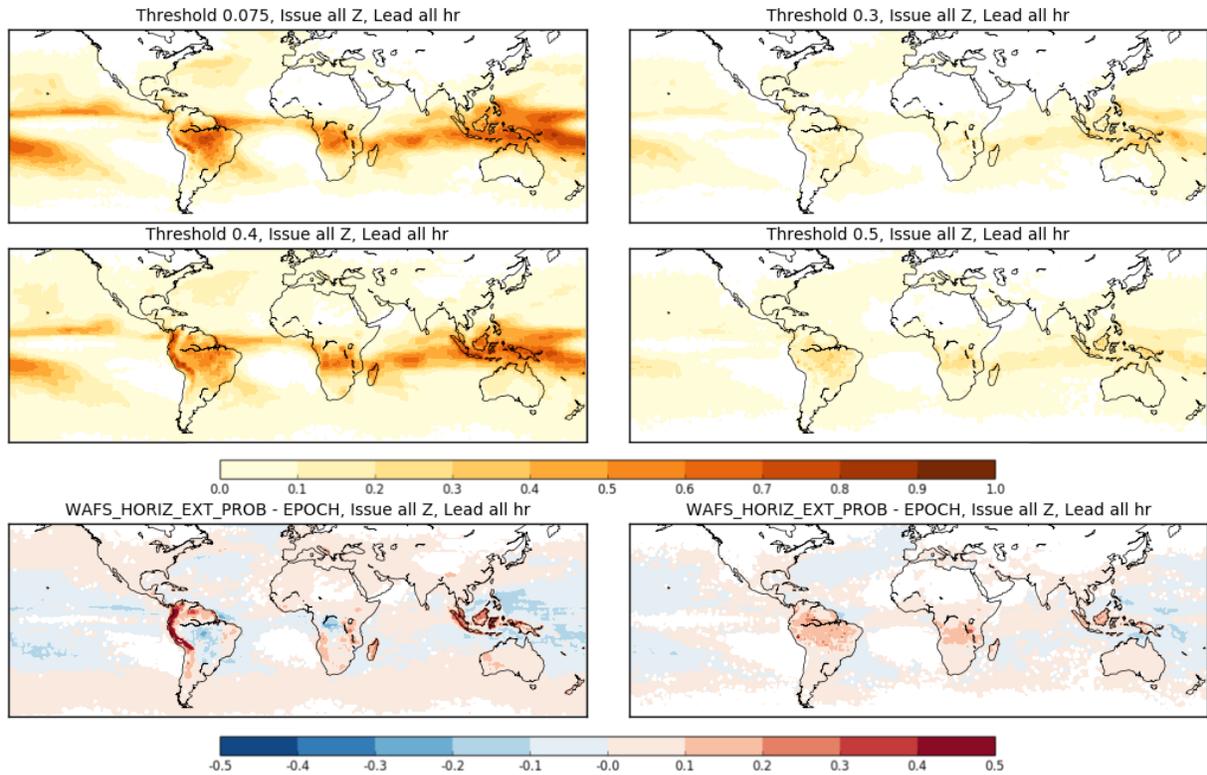
22

Figure 14. Forecast climatological maps for EPOCH (top), WAFS (middle), and WAFS-EPOCH (bottom; red=more WAFS, blue=more EPOCH). The thresholds used for each forecast correspond to the "occasional" (left) and "rare" (right) event frequencies.

Climatological maps of observed event frequencies were created for comparison between datasets and both forecast products. The results for global datasets using select observation definitions are presented in Figure 15. In agreement with the observed event frequencies presented in Figure 12, these maps illustrate the bounding nature of each dataset on a climatological basis, with the lower bound fields (IMERG, IRA, LTG, and LTG25) presenting lower event frequencies than the forecasts and the upper bounds (CTH and CTH15) showing higher event frequencies. Climatological maps for all observation definitions used in this assessment are included in Section 10.2.1 (Appendix B).
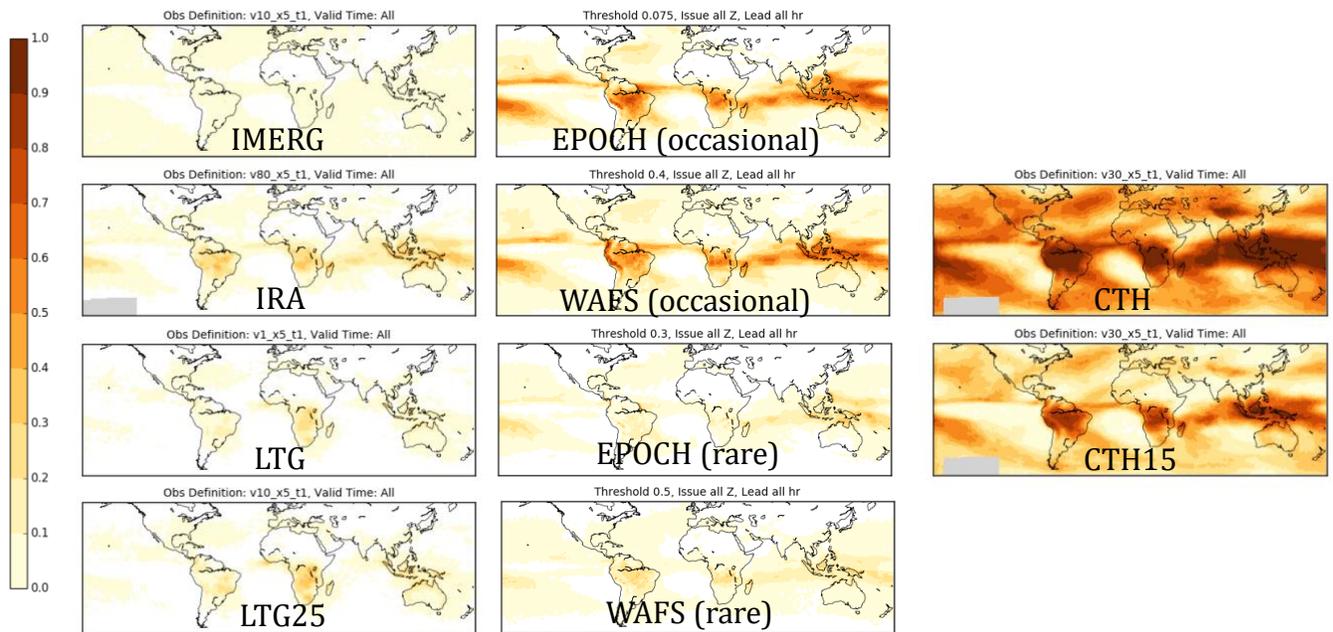
Figure 15. Global climatological maps of forecast and observation datasets. The left column shows observations used as lower bounds, while the right column shows upper bounds. The threshold used for each forecast (center) correspond to the "occasional" and "rare" forecast event rates. The observation definitions used a spatial coverage of 5%, temporal extent of 1 hr, and the primary intensity values given in Table 5.

5.2 PERFORMANCE METRICS

5.2.1 PERFORMANCE METRICS USING PRIMARY TRUTH DATASETS

With high spatial and temporal resolution, ground-based radar observations provide the best truth field for thunderstorm detection. In this section, performance metrics are calculated against the MRMS-ET. Figures showing results for the European maximum composite reflectivity can be found in Section 10.2.2 (Appendix B).

The MRMS-ET data, using a threshold of 30 kft to define observed "Yes" events, provided a truth field for the CONUS domain. The ROC curves for both EPOCH and WAFS against this data are shown in Figure 16. Each panel contains four curves, corresponding to different selection of the spatial and temporal requirements for defining a "Yes" observation. The bounding values of 5, 25% and 1, 4-hrs are presented to capture the range of outcomes. Contrary to the findings in Part I, WAFS had a higher Area Under Curve (AUC) than EPOCH for this assessment period for each spatial and temporal extent. The performance diagrams (Figure 17) show that EPOCH and WAFS have similar CSI values, when selecting the best performing threshold for each forecast product.
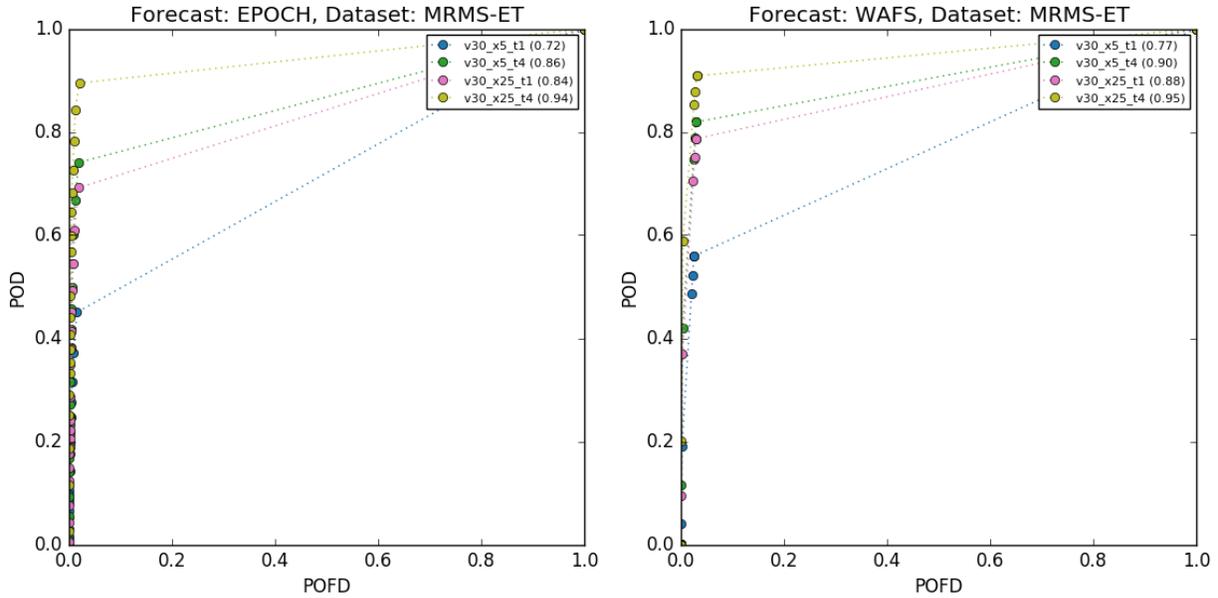
Figure 16. ROC curves for EPOCH (left) and WAFS (right) compared against MRMS Echo Top data using an intensity threshold of 30 kft and bounding spatial coverages and temporal extents. The Area Under Curve values obtained using each observation definition are included in the legend.
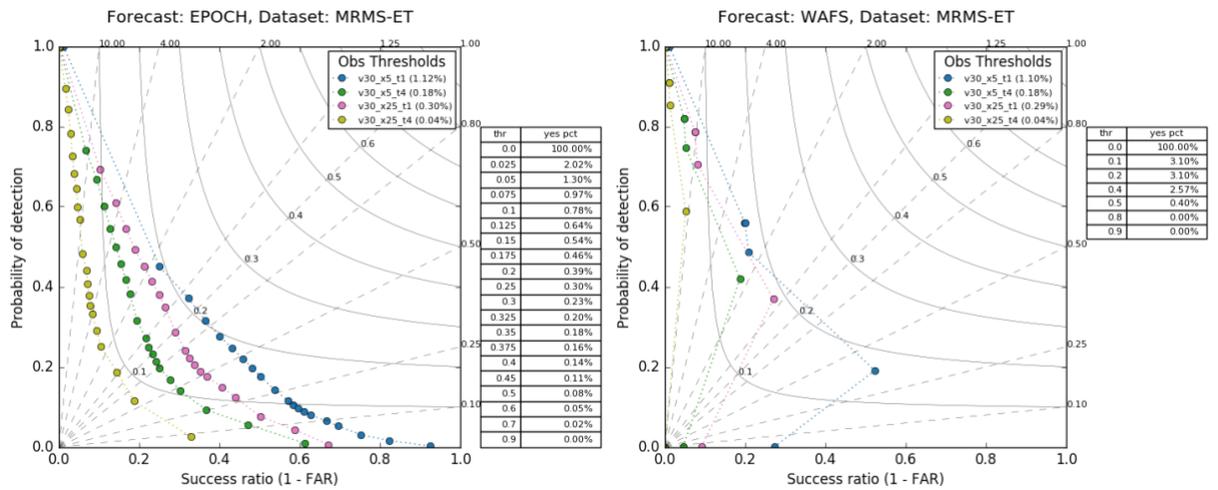

Figure 17. Performance diagrams for EPOCH (left) and WAFS (right) compared against MRMS Echo Top data using an intensity threshold of 30 kft and bounding spatial coverages and temporal extents. The dashed lines represent bias, while the solid curves give constant CSI values. The observed event frequencies using each observation definition are included in the legend. Also included is a table giving event frequencies by forecast threshold.

Both radar datasets were also used to compute the reliability of each forecast product, using a range of observation definitions to represent the truth field. Figure 18 shows the results using the MRMS-ET data with a threshold of 30 kft, and the same bounding spatial and temporal extents as used previously. EPOCH had higher observed relative frequencies than WAFS, consistent with the Part I findings, but the curves are rotated clockwise compared with the NH Summer season. Which forecast was more reliable depended on the choice of the observation definition, i.e. the scale of weather events considered relevant. Using the least-strict definition, which includes smaller, shorter-lived storms (bottom left), WAFS produced forecasts that

approach perfect reliability for several thresholds, but was near the no-skill line at the peak of the distribution (40%). EPOCH maintained some skill when using the 1-hr temporal threshold (left), but over-forecast when restricting observed events to those meeting the 4-hr threshold (right). The same trends were observed when computing the forecast reliability against the European radar data (see Figure 46, Appendix B).
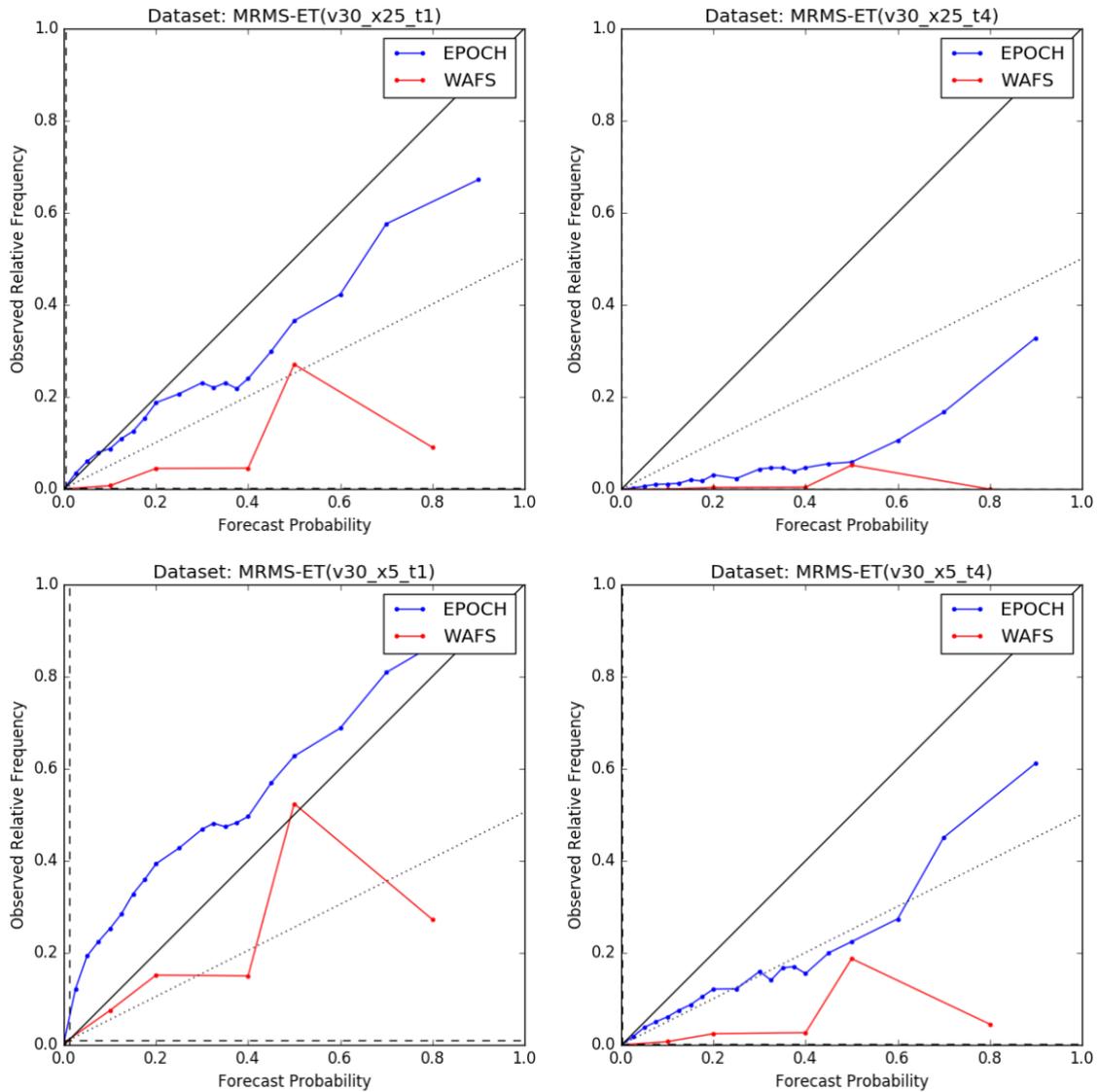

Figure 18. Reliability diagrams for EPOCH (blue) and WAFS (red) compared against MRMS echo top data. An observation intensity threshold of 30 kft is used along with the bounding spatial coverages (5%, bottom; 25% top) and temporal extents (1 hr, left; 4 hr, right). The black lines represent perfect reliability (solid), no skill (dotted), and climatology (dashed).

5.2.2 PERFORMANCE METRICS BY DATASET

Although a range of intensity thresholds were explored for each observation type, the results presented in this section will focus on a primary threshold that acts as either a truth field (for ground-based radar observations), or an upper or lower bound on the forecast. For a lower bound field, the forecasts were expected to produce "Yes" events whenever there is an observed "Yes" event, but an observed "No" event does not necessarily indicate that convection is absent. For an upper bound field, the forecasts should not produce a "Yes" event outside an event in the observations.
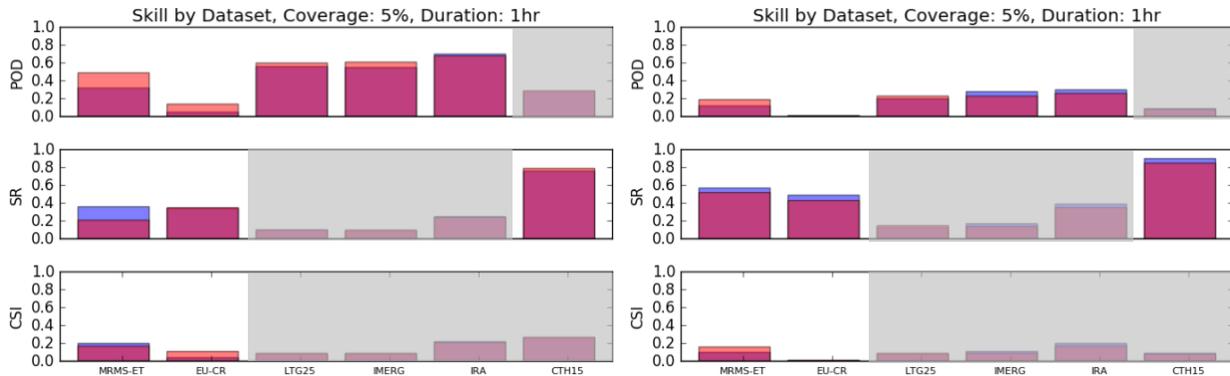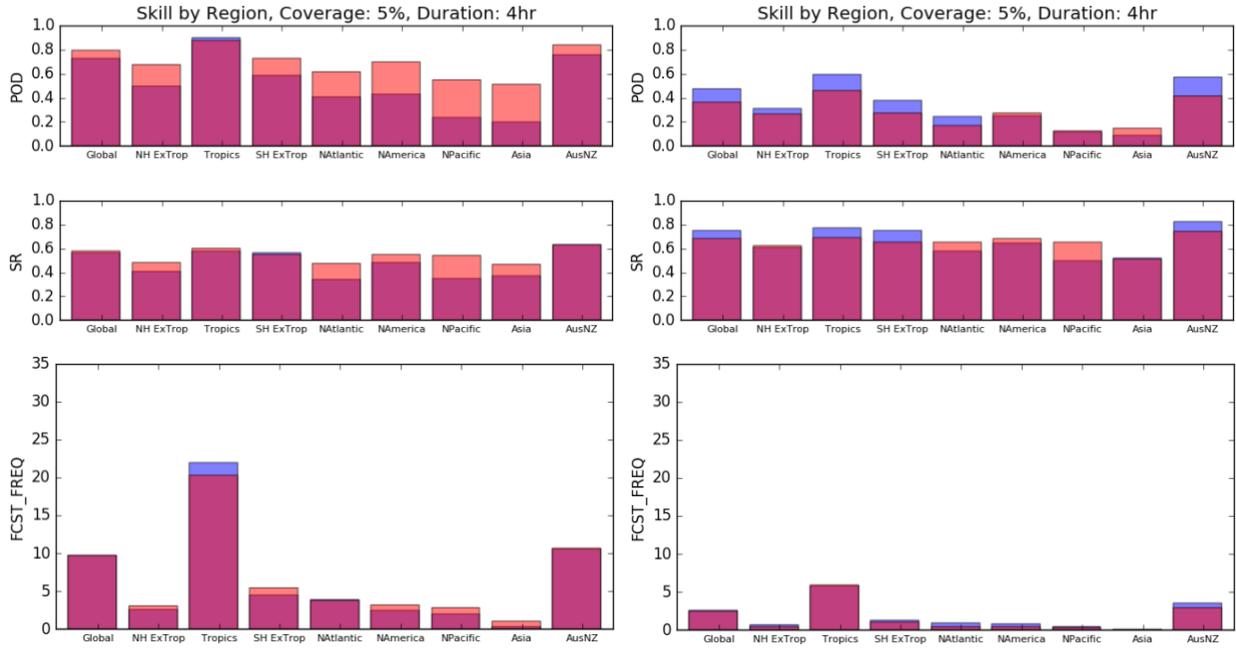

Figure 19. Performance metrics by dataset using the equalized forecast thresholds corresponding to "occasional" (left) and "rare" (right) event occurrence. The scores for EPOCH (blue) and WAFS (red) are overlaid using 50% transparency. Gray shading indicates equivocal scores from bounding datasets. The observation definition for each dataset uses 5% spatial coverage, 1-hr temporal extent, and the primary intensity thresholds given in Table 5.

Skill scores for each forecast threshold were examined for a range of observation definitions (see Figure 47 and Figure 48, Appendix B) and a representative case using the 5% coverage and 1-hr temporal extent is presented for brevity (Figure 19). The primary intensity used for each observation type is given in Table 5. WAFS scored a higher POD against all datasets except IRA at the "occasional" forecast threshold (Figure 19, left), and all except IRA and IMERG at the "rare" threshold (right). By contrast, EPOCH generally had a higher SR across the different observation datasets for both forecast thresholds, with one exception: the SR was slightly higher using CTH15 at the 5% and 1-hr coverage. WAFS earned a higher CSI against the European radar at the low forecast threshold as well as against the MRMS-ET at the "rare" threshold (consistent with the results presented in Section 5.2.1).

5.2.3 PERFORMANCE METRICS BY REGION

Skill by region was also evaluated for both EPOCH and WAFS using a variety of forecast thresholds and observation definitions. For results shown in this section, the POD was obtained using the IMERG threshold of 10 mm/hr, which provided a lower bound on the forecasts, while the SR was obtained using the 30-kft threshold of filtered cloud top height (CTH15), which acted as an upper bound. For conciseness, only the 5% spatial and 4-hr temporal coverage case is shown, but the patterns held across the other spatial and temporal extents. (To allow comparison, the results from Part I are presented in Appendix B, Figure 49).

Using the "occasional" forecast threshold (Figure 20, left), EPOCH achieved higher POD values than WAFS in the tropics, but a lower POD in all other regions and globally. At the same threshold, WAFS had a higher SR everywhere except the SH extratropics and Australia/New Zealand (AusNZ). Recall that results from the Part I assessment period showed that EPOCH earned a higher SR in every region, except the SH extratropics (see Figure 49 in Appendix B). Looking across both assessment periods, EPOCH generally

produced fewer false alarms in the summer hemisphere, while WAFS produced fewer in the winter hemisphere. At the "rare" forecast thresholds (Figure 49, right), EPOCH achieved a greater POD globally, while WAFS detected more events in the NH land regions. A similar pattern held when looking at SR, with EPOCH having fewer false alarms globally but WAFS having fewer in the NH extratropics.

The results for the additional regions added for Part II of the assessment are shown in Figure 21. At both forecast thresholds, WAFS produced more event over land, while EPOCH produced more in oceanic regions. At the "occasional" forecast threshold (Figure 21, left), the forecast with the higher event rate earned a lower POD but lower SR (more false alarms) for most regions, indicating that both forecasts suffered from spatial and/or temporal placement issues. Using the "rare" forecast threshold, WAFS had the same pattern of mixed score results for all land regions, while EPOCH outperformed WAFS in terms of both POD and SR in all oceanic regions.

Figure 20. Performance metrics by region using the equalized forecast thresholds corresponding to "occasional" (left) and "rare" (right) event occurrence. The POD (top) is obtained using IMERG ≥ 10mm/hr, while SR (middle) is obtained using CTH15 ≥ 30 kft. Forecast event frequencies (bottom) are given in percent. Scores using the 5% and 4-hr observed event thresholds are shown.
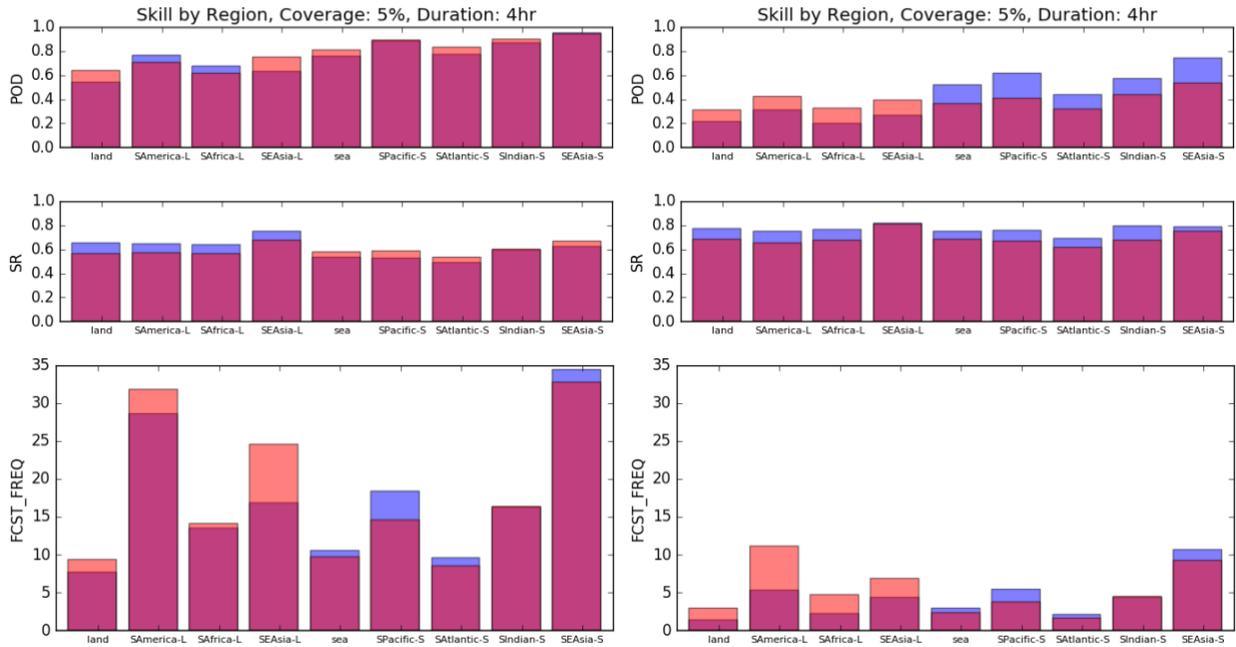

Figure 21. Same as Figure 22, except for the additional land (L) and sea (S) regions defined for Part II of the assessment.

Event frequencies were computed by latitude band to assess differences between the two forecast products. The Northern and Southern Hemisphere extratropics were split into two bands ranging from 20 to 40 degrees, and 40 to 60 degrees in latitude. Figure 23 shows the forecast frequencies for EPOCH and WAFS using the "occasional" forecast threshold and observed event frequencies for the primary observation definitions. EPOCH produced more events than WAFS in the tropics, with rates falling off sharply when moving to higher latitudes, such that its event frequency in the northernmost band was little more than half that in the WAFS forecasts, consistent with the Part I findings.
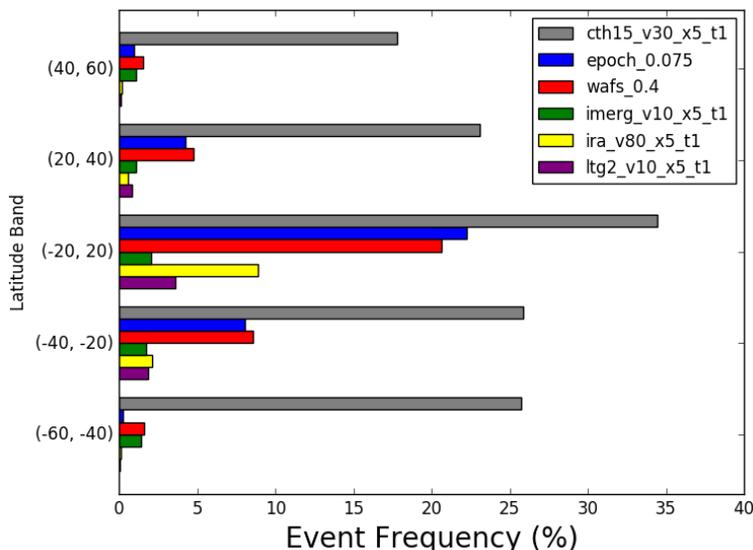
Figure 23. Event frequency by latitude band. The thresholds chosen for EPOCH (7.5%) and WAFS (40%) correspond to the "occasional" event occurrence. Observed event frequencies are shown for the primary intensity threshold for each observation dataset using the 5% and 1-hr observation definition. Negative latitudes correspond to the Southern Hemisphere.

5.2.4 PERFORMANCE METRICS BY VALID TIME

Results for Part I of the assessment showed that issuance and lead time had little impact on the skill of either forecast product, and that finding was substantiated by the Part II results (see Figure 50 in Appendix B). The analysis completed in Part I also showed notable differences in the two forecast products when examined by valid time for specific regions. In order to further investigate diurnal patterns, scores were computed by valid time in UTC, then compared using a "local valid time" calculated from the UTC offset of the central longitude for each region. Representative results are displayed in Figure 24, which shows the POD at the "occasional" threshold against the IRA dataset. For the oceanic regions (top), EPOCH and WAFS scored similar POD values for all local times, with a slight decrease for nighttime convection. In contrast, both forecast products had a notable diurnal pattern for the land regions (bottom), with higher POD values for the afternoon hours compared with overnight events. The timing of the peak POD for EPOCH was later in the evening compared with the afternoon peak seen for WAFS. The forecast bias by local valid time, shown in Figure 25, indicates that the increased POD obtained by EPOCH in the evening hours was achieved despite having a reduced forecast bias.
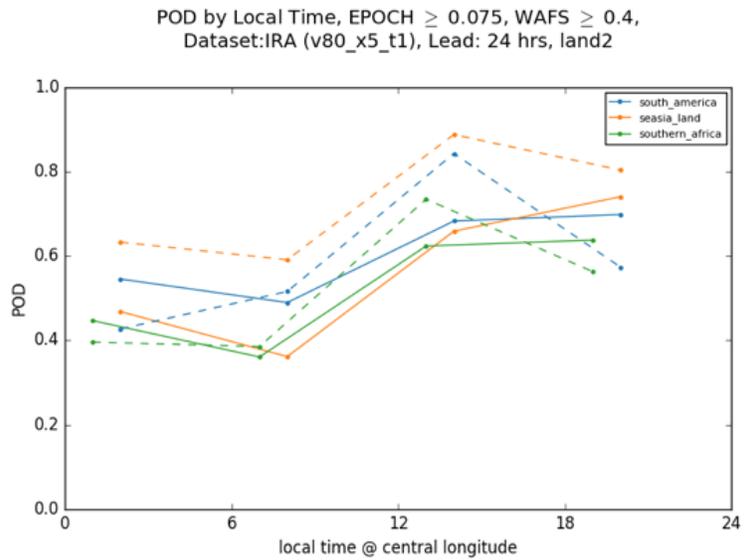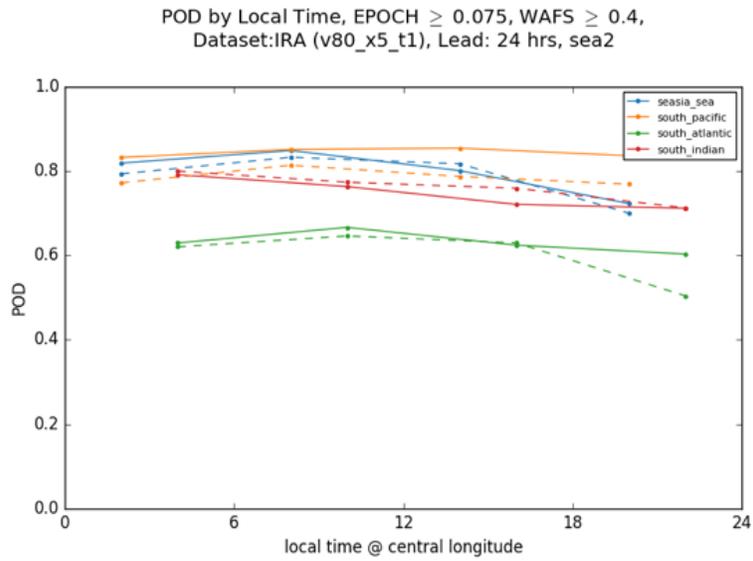
Figure 24. POD by local time by oceanic (top) and land (bottom) region for both EPOCH (solid) and WAFS (dashed) at the "occasional" forecast threshold using the IRA (v80_x5_t1) observation definition.
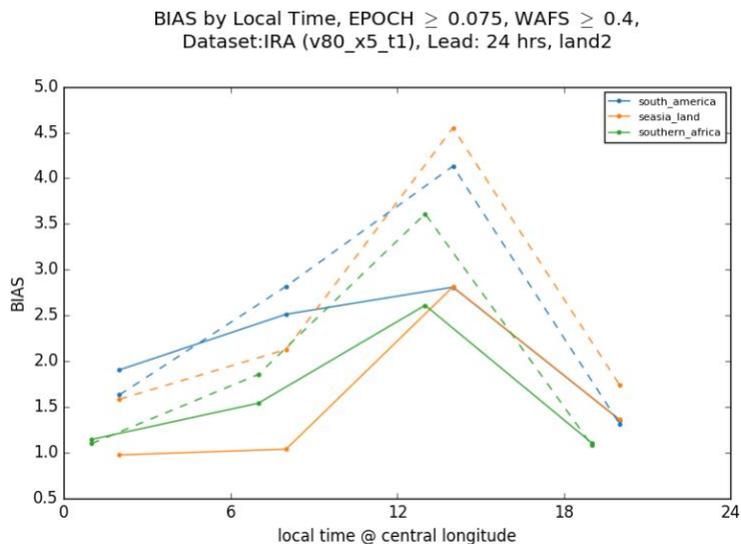
BIAS by Local Time, EPOCH $\geq$ 0.075, WAFS $\geq$ 0.4,
Dataset:IRA (v80_x5_t1), Lead: 24 hrs, land2

Figure 25. Same as Figure 24 (bottom), except for forecast bias.

5.3   CONSISTENCY

The consistency across lead times was calculated for EPOCH and WAFS using the equalized forecast thresholds (Figure 26). As seen in Part I, EPOCH was more consistent than WAFS at either forecast threshold. Both EPOCH and WAFS were more consistent at the "occasional" threshold, but the gap between the consistency of the two products widened at the "rare" threshold, with WAFS consistency also noticeably declining for increasing lead times.
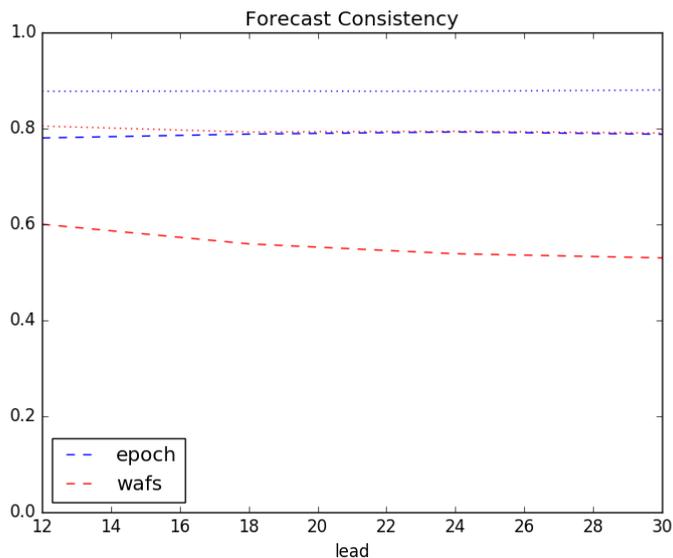
Figure 26. Forecast consistency for EPOCH (blue) and WAFS (red). The results for the equalized forecast event thresholds corresponding to "occasional" (dotted) and "rare" (dashed) event occurrence are shown.

5.4 ASSESSMENT OF CONVECTIVE CLOUD TOP FIELDS

5.4.1 FIELD CHARACTERISTICS

5.4.1.1 FORECAST FIELD DISTRIBUTIONS

Forecast distributions for each EPOCH CCT field (blue) and the WAFS Cb Tops (red) field are shown in Figure 27. To obtain corresponding forecast event frequencies for the two products, the cumulative distributions (not shown) of the three EPOCH convective cloud tops fields were matched to that of corresponding WAFS Cb Tops threshold. The resulting forecast thresholds and event frequencies are given in Table 12. Note that the best-matching EPOCH thresholds still gave slight differences in event frequencies for 30 and 35 kft, and in both cases the EPOCH frequency was slightly higher.
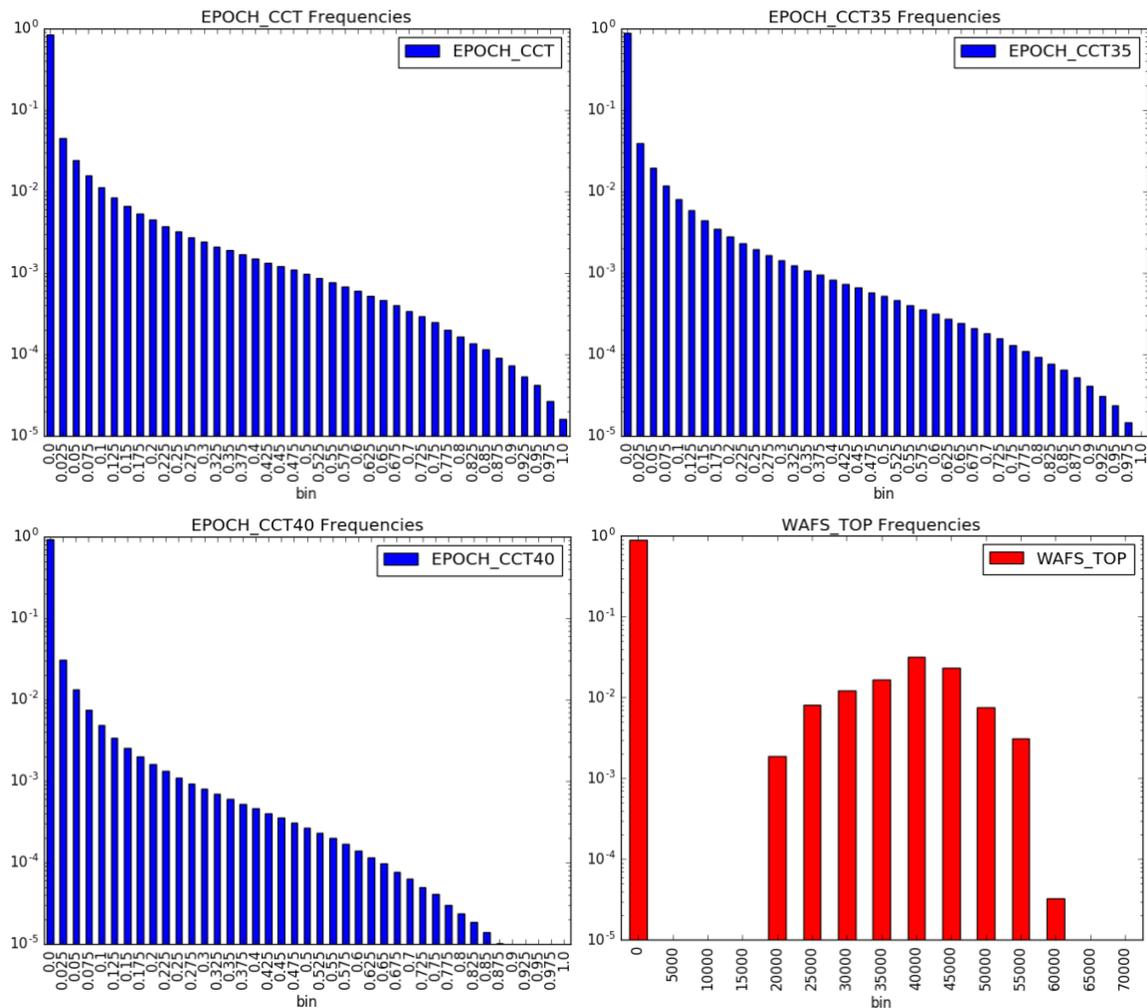
Figure 27. Frequency distributions for the convective cloud top forecast fields. EPOCH fields are shown in blue: 30 kft (upper left), 35 kft (upper right), 40 kft (Bottom left). The WAFS Cb tops distribution is shown in red (bottom right).

Table 12. Frequency-matched thresholds for the CCT forecast fields.

CCT Threshold	WAFS Tops Frequency	EPOCH CCT Threshold	EPOCH CCT Frequency
30 kft	9.5%	5.0%	10.5%
35 kft	8.3%	5.0%	7.0%
40 kft	7.5%	2.5%	7.5%

5.4.1.2 CLIMATOLOGICAL MAPS

Climatological maps showing the event frequencies for convective cloud tops exceeding 35 and 40 kft are shown in Figure 28. As found in Part I, the overall geographic patterns were similar to those seen for the main forecast fields (Figure 15), except that Cb tops exceeding either 35 or 40 kft were more common in WAFS through most of the tropics, despite EPOCH placing a similar number or more storms there overall.
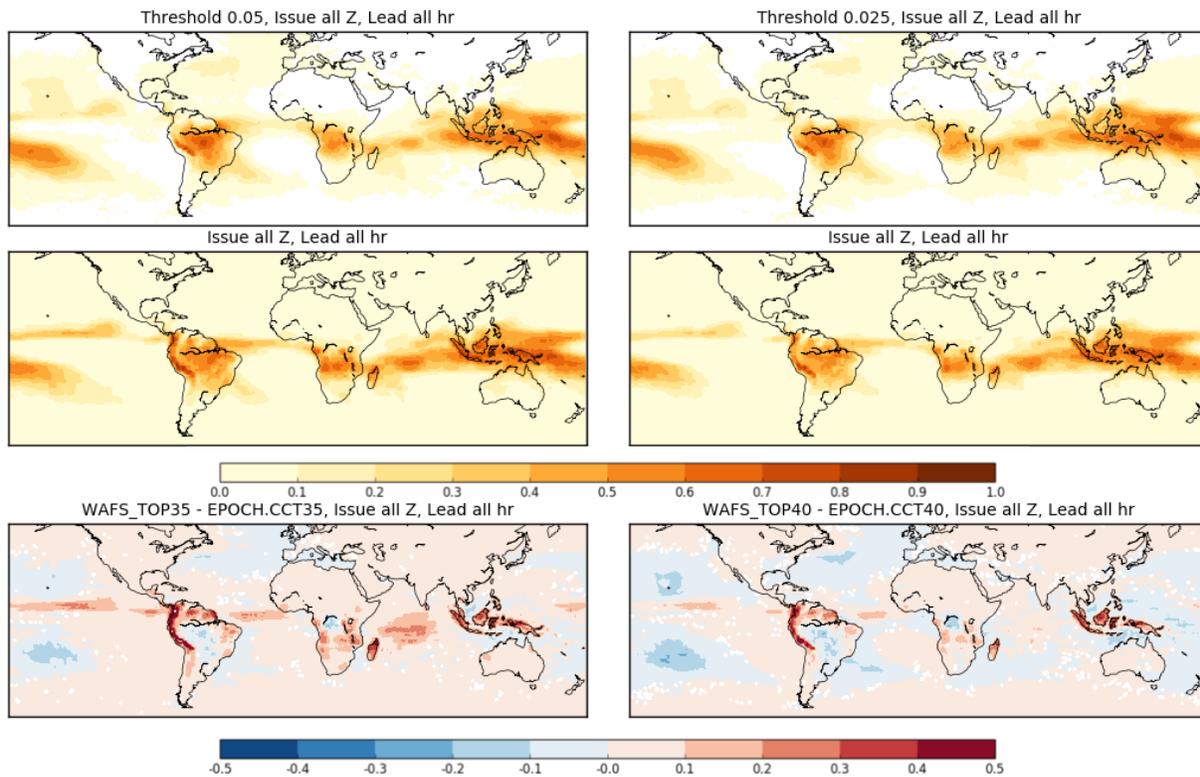
Figure 28. Forecast climatological maps of CCT exceeding 35 kft (left) and 40 kft (right) for EPOCH (top), WAFS (middle), and WAFS-EPOCH (bottom; red=more WAFS, blue=more EPOCH). The thresholds used for each forecast are given in Table 12.

5.4.2 PERFORMANCE METRICS

In Part I, skill scores for the convective cloud top fields were computed against the MRMS Echo Top data, which is limited to the CONUS domain. For Part II, the skill scores were also computed against the filtered cloud top height (CTH15). Performance diagrams for each altitude threshold are shown in Figure 29. In each case, the forecast altitude was used for the Echo Top (top) or Cloud Top Height (bottom) threshold, coupled with the set of bounding spatial and temporal extents. When compared to the MRMS-ET dataset (top), WAFS generally scored a higher POD and lower SR than EPOCH at the 30 kft (left) and 35 kft (middle) thresholds. At 40 kft, the two products performed similarly for the 5%, 1hr threshold, but WAFS

retained a higher POD for larger, longer-lived storms. When compared to the CTH15 dataset, EPOCH and WAFS had similar skill at each of the three altitude thresholds.
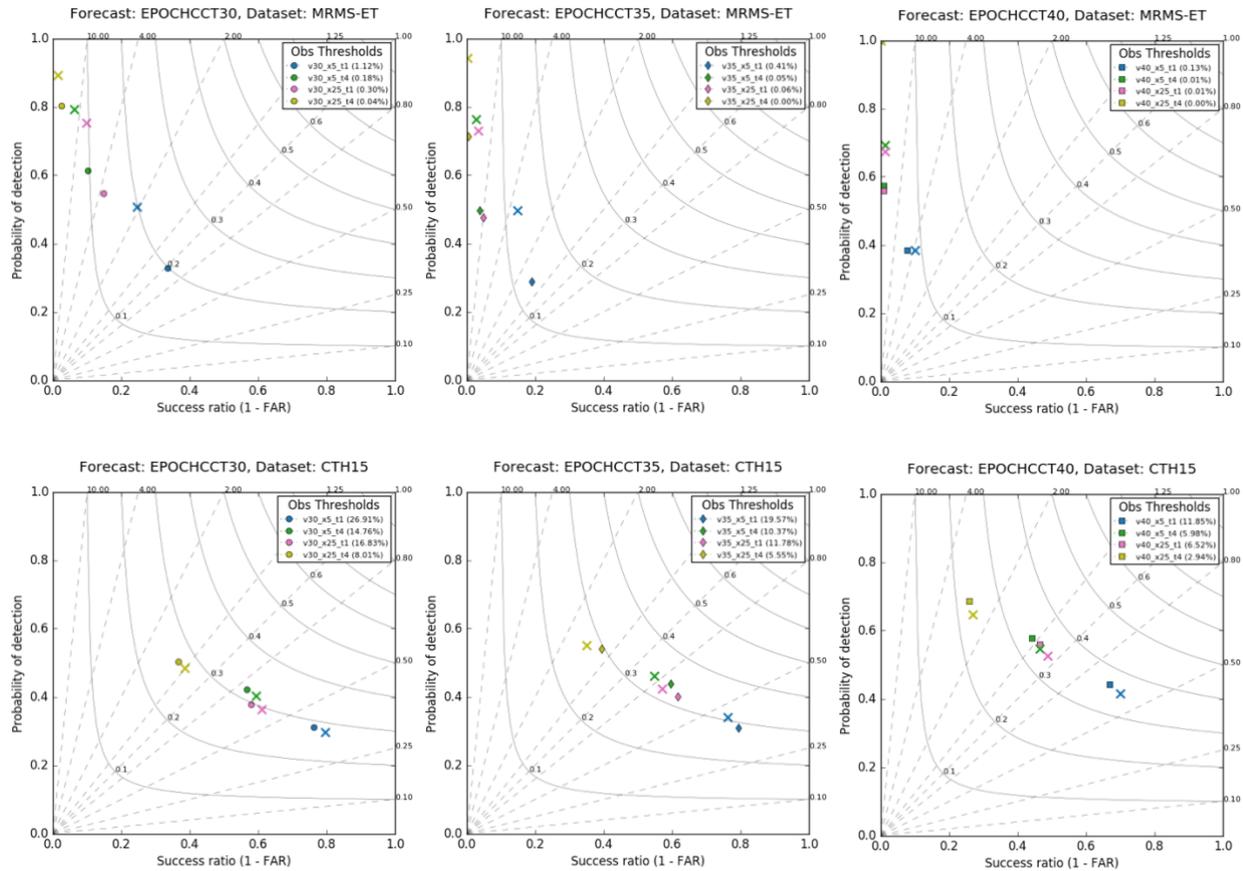

Figure 29. Performance diagrams for convective cloud top fields in EPOCH (filled markers) and WAFS Cb tops (X markers) against MRMS-ET (top) and CTH15 (bottom) data for altitude thresholds of 30 kf (left), 35 kft (middle), and 40 kft (right).

6 CASE STUDIES

Two case studies were examined for this assessment. One case focused on convection over South America, and the other on convection over Asia. The cases were chosen to demonstrate the behavior of the products in both the summer and winter hemisphere. Results for both cases are presented using the 24-h forecast lead time, based on the motivation for and expected use of the product.

6.1 SOUTH AMERICA

On 20-21 March 2018, a heavy rainstorm, producing lightning and strong winds, occurred in the city of São Paulo, Brazil. Satellite imagery (Figure 30) showed the presence of afternoon convection over land.

Development of the storm from isolated showers to organized convection can be seen in the radar data (Figure 31).
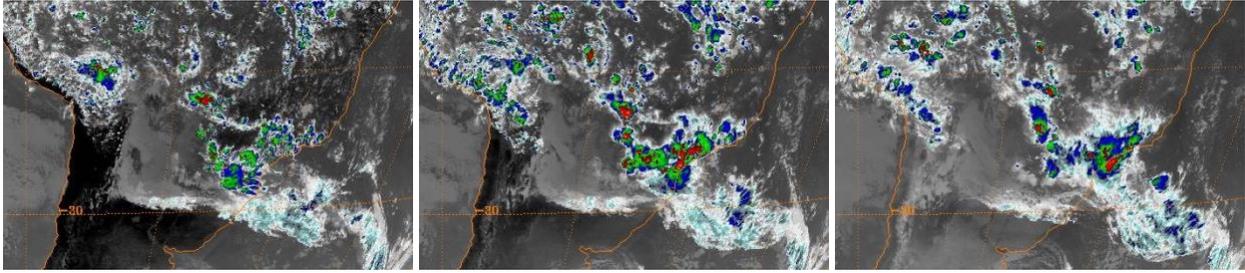

Figure 30. Enhanced IR images of part of South America, including southern Brazil, at 1745, 2045, and 2345 UTC, on 20 Mar 2018.

Figure 31. Composite reflectivity in the vicinity of São Paulo, Brazil on 20 March 2018 at 1800 (left), 2000 (middle) and 2200 (right) UTC. Data obtained from SOS CHUVA database.
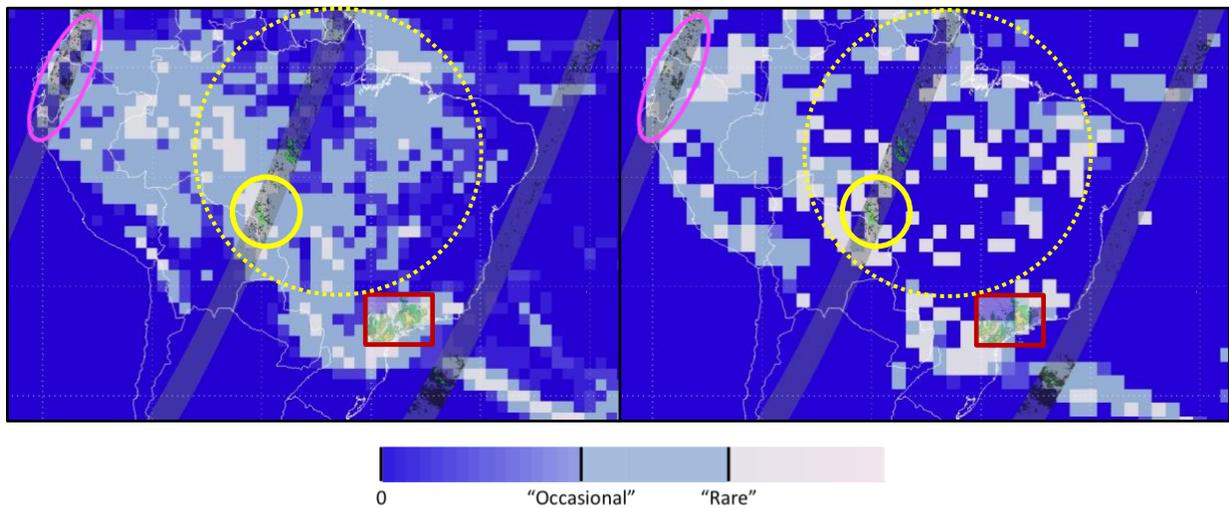

Figure 32. The 24-h forecasts of EPOCH thunderstorm likelihood (left) and WAFS Cb horizontal extent (right) valid for 1800-2400 UTC, 20 March 2018. The main color scale differentiates forecast values exceeding the "occasional" (light blue) and "rare" (grey) thresholds. The red box contains an overlay of the 2200 UTC radar image from Figure 31. The swaths show the storm top height observed by the GPM Ku-radar using green to represent tops exceeding 30 kft.

The 24-hr forecasts for both EPOCH and WAFS are shown in Figure 32. Both forecasts captured the deep, organized convection over southern Brazil. The storm top height observed by the GPM Ku-radar shows low tops (< 30 kft) over central Ecuador and surrounding areas along the Andes (solid magenta oval). In this region, WAFS produced a broad area exceeding the "occasional" threshold; whereas EPOCH had

pockets where this threshold was not met. The GPM radar detected tops exceeding 30 kft in the swath across inland Brazil. Both forecast products exceed the "rare" threshold in a region coincident with the tops near the Bolivia-Brazil border (solid yellow circle). However, the two products differed across much of the Amazon Basin (dashed yellow circle), where EPOCH produced broad areas exceeding the "occasional" threshold, while the WAFS forecast produced isolated pockets exceeding the "rare" threshold. Both forecasts represent the convection around São Paolo as the more organized, MCS-like structure seen in the 2200 UTC radar image (Figure 31), with broad areas exceeding the "rare" threshold.

6.2   ASIA

On 25 Feb 2018, a Shaheen Airlines flight from Dubai to Islamabad suffered extensive hail damage, struck by nearly 700 hail stones (Aviation Herald, 2018). This case was chosen to highlight convection in the winter hemisphere which was sufficient to be hazardous to aviation. Although the aviation impact was a factor in selecting this particular case, the characteristics of the two forecasts in the surrounding region were examined independent of the (unknown) flight path in order to provide a broader perspective. Satellite imagery (Figure 33) showed thunderstorms associated with a cold front over the Persian Gulf, as well as areas of weaker activity over northern parts of India.

For the activity in the Persian Gulf region (Figure 34, solid oval), lightning counts (left) show most strikes were detected to the west over Saudi Arabia and Iran, while the accumulated precipitation (right) shows the peak rainfall occurring over the Gulf and extending inland to the northeast over Iran. Both forecast products (Figure 35) captured the core of convection over the Persian Gulf, but WAFS (right) extended the region exceeding the "occasional" threshold further inland into Saudi Arabia and Iran than EPOCH (left).

A second area of convection was detected over the northern part of India (Figure 34, dashed oval). Although these storms produced lower rainfall rates (right), there was lightning detected within the region (left). The GPM radar also detected high storm tops (>30 kft) near the India-Nepal border (Figure 35, green). WAFS captured small areas exceeding the "rare" threshold in this vicinity (Figure 35, right), whereas EPOCH did not produce any likelihoods meeting the "occasional" threshold (left).

As an aside, this case also illustrates the challenge of observing global convection. The satellite imagery (Figure 33) indicated inclement weather over a broader area than either the lightning or accumulated rainfall field (Figure 34), while the latter two data sources identified areas largely displaced from each other. In addition, while neither the planned or actual route of the flight is known, there was little in either the rainfall or lightning field in the area of what one would expect the route to be. This is particularly true since the flight continued to Islamabad, suggesting that the incident likely occurred later in flight and not with the weather near Dubai.
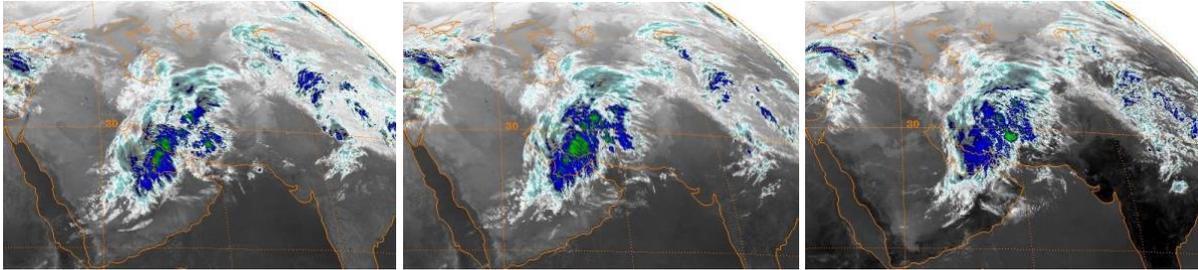
Figure 33. Enhanced IR images of a region including the Persian Gulf and India at 0012, 0312, and 0612 UTC, 25 Feb 2018.

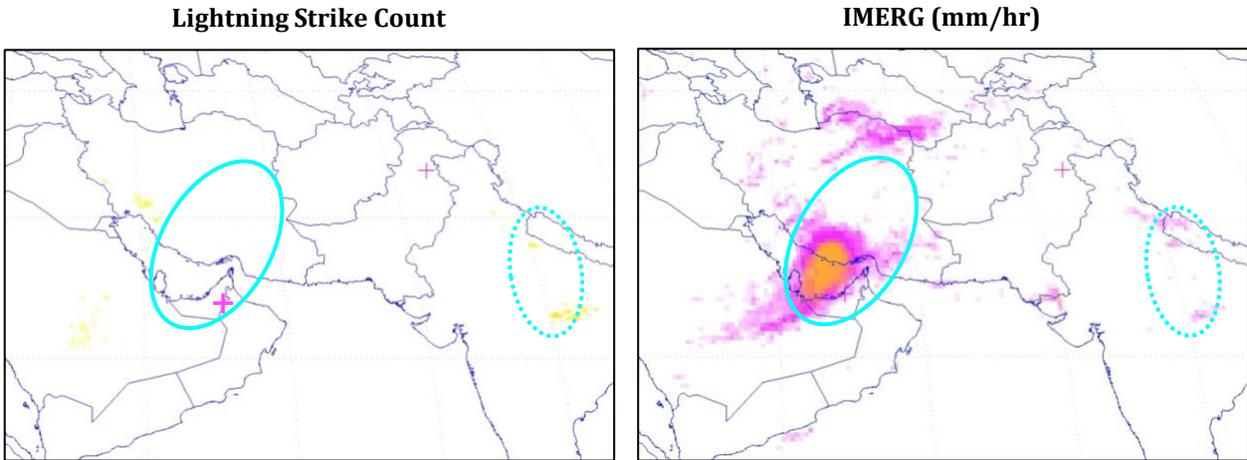**Lightning Strike Count**   **IMERG (mm/hr)**

Figure 34. Lightning strike count (left) and IMERG accumulated precipitation (right) for 2018-02-25 0000 – 0600 UTC. The cyan ovals indicate regions of interest, while the + symbols indicate the origin and destination airports of the Shaheen Airlines flight.
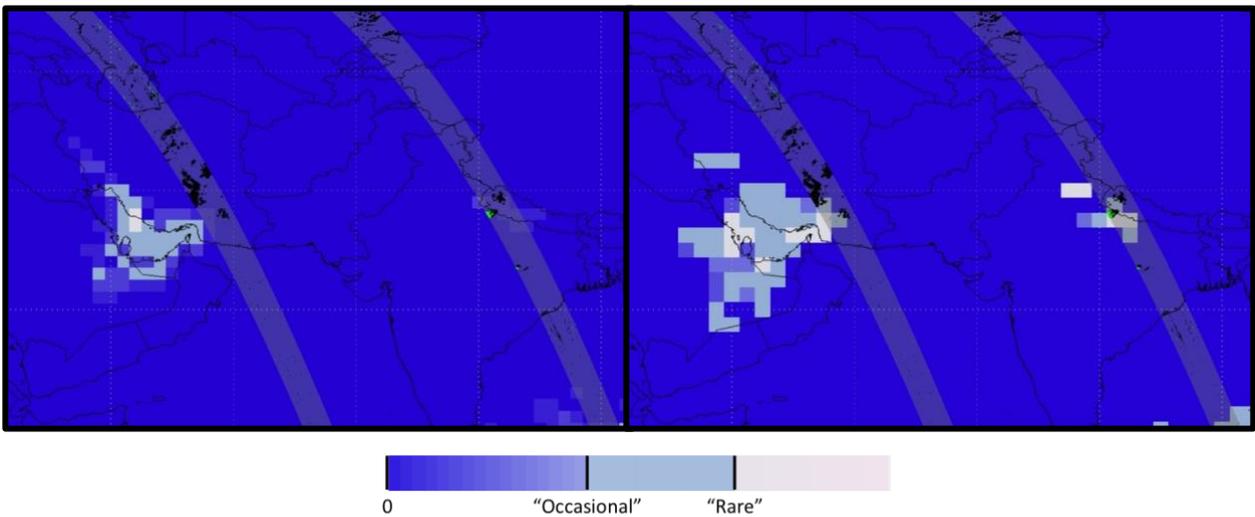

Figure 35. The 24-h forecasts of EPOCH (left) and WAFS Cb (right) valid for 0000-0600 UTC, 25 Feb 2018. The main color scale differentiates forecast values exceeding the "occasional" (light blue) and "rare" (grey) thresholds. The swaths show the storm top height observed by the GPM Ku-radar using green to represent tops exceeding 30 kft.

7 SUMMARY

While there are no direct observations against which to evaluate the global performance of EPOCH forecasts, proxies for thunderstorm occurrence and convective cloud top observations were identified for use in this assessment. The verification of EPOCH employed a diverse set of observations, including: two radar data sets, lightning, and global satellite cloud and precipitation products, which are less directly linked to identification of tall thunderstorms. The EPOCH products were assessed and compared with the baseline WAFS Cb forecasts through the application of climatological mapping, statistical metrics from contingency tables, and case studies.

Many findings in this report confirmed results from Part I of the assessment. While the global performance of EPOCH is similar to that of the existing WAFS Cb product, the skill of the two forecasts is dependent on the strength and scale of weather events considered relevant. EPOCH generally performed better than WAFS using the "rare" forecast threshold, while relative performance at the "occasional" forecast threshold depended on the dataset and observed event definition. EPOCH was found to produce more consistent forecasts than WAFS, with a greater difference at the "rare" forecast threshold.

Notable regional differences in the two products were identified. As seen Part I of the assessment, a stronger latitudinal drop-off was present in EPOCH event rates than in WAFS. EPOCH also had greater skill than WAFS in the global tropics. The relative performance in the extratropics depended on the season, with EPOCH having fewer false alarms in the summer extratropics, while WAFS had fewer in the winter.

EPOCH was found to produce more events over the oceans, while WAFS forecasted more events over land. Consistent with the global results, WAFS was found to perform better over oceanic regions when using the "occasional" forecast threshold; while EPOCH performed better at the "rare" threshold. In contrast, the performance over land was more mixed, where WAFS scored a higher POD but lower SR over global land areas at both forecast thresholds. EPOCH outperformed WAFS over South America and southern Africa for the "occasional" threshold, achieving a higher POD despite having a lower event rate. EPOCH also produced more events over oceanic regions, but scored a lower POD for the global oceans, indicating that both forecast products suffered from placement issues.

The impact of the diurnal cycle on forecast skill was shown to differ between land and oceanic regions. While little variation by local valid time was seen for the oceanic regions, a clear pattern was seen in both EPOCH and WAFS for the land regions, where both forecast products scored higher POD values in the afternoon hours compared with overnight events. In Part I of the assessment, climatological studies over land regions indicated that WAFS tended to initiate convection early. This finding was substantiated by the higher forecast bias for WAFS in the late morning and early afternoon for each land region defined in this assessment. For the late evening, EPOCH scored a higher POD, despite having a similar or lesser forecast bias than WAFS over these same regions.

The performance of the forecast products in identifying convective cloud tops at or above 30 kft was found to be similar when using the filtered cloud top height as the truth dataset.

The case studies corroborated the global characteristics. For the South America case, both forecast products captured the convection over Sao Paulo, Brazil, but WAFS placed more convection near the Andes, and had more isolated pockets exceeding the "rare" threshold over the Amazon basin, compared with the broad

areas exceeding "occasional" threshold seen in EPOCH. For the Asia case, both products captured the core of convection over the Persian Gulf, but WAFS captured a larger inland area, and a smaller region of activity over northern India that was missed by EPOCH.

8 ACKNOWLEDGEMENTS

9 REFERENCES

Aviation Herald, 2018: http://avherald.com/h?article=4b6af03f&opt=4096

Bedka, K. and K. Khlopenkov, 2016: A Probabilistic Multispectral Pattern Recognition Method for Detection of Overshooting Cloud Tops Using Passive Satellite Imager Observations. *J. Appl. Meteor. Clim.,* **55**, 1983-2005.

Hamada A. and Y. N. Takayabu, 2016: Improvements in Detection of Light Precipitation with the Global Precipitation Measurement Dual-Frequency Precipitation Radar (GPM DPR). *J. Atmos. Oceanic Technol.,* **33**, 653–667.

Huffman, G. J., Bolvin, D.,T., Braithwaite, D., Joyce, R., Kidd, C., Nelkin, E.J., and X. Pingping, 2015: Algorithm Theoretical Basis Document (ATBD) Version 4.5: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) [Available at http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf]

ICAO, 2012: Guidance on the Harmonized WAFS Grids for Cumulonimbus Cloud, Icing and Turbulence Forecasts, Version 2.5.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydromet*., **5**, 487-503.

Jurkovic P.M., N. Strelec Mahovic, and D. Pocakal, 2015: Lightning, overshooting top and hail characterisitics for strong convective storms in Central Europe. *Atm. Res.,* **161-162,** 153-168.

Khaiyer, M. M., et al., 2006: Derivation of improved surface and TOA broadband fluxes using CERES-derived narrowband-to-broadband coefficients. *Proc. AMS 12th Conf. Atmos. Radiation*, Madison, WI, July 10-14, CD- ROM, P3.5.

Lakshmanan, V., K. Hondl, C. K. Potvin, and D. Preignitz, 2013: An improved method for estimating radar echo-top height. *Wea. Forecasting*, **28,** 481-488.

Mace, G. G., T. P. Ackerman, P. Minnis, and D. F. Young, 1998: Cirrus layer microphysical properties derived from surface-based millimeter radar and infrared interferometer data. *J. Geophys. Res.,* 103, 23,207-23,216.

Min, Q, P. Minnis, and M. M. Khaiyer, 2004: Comparison of cirrus optical depths from GOES-8 and surface measurements. *J. Geophys. Res*. 109, D20119, 10.1029/2003JD004390.

Minnis, P., and Co-authors, 2008: Near-real time cloud retrievals from operational and research meteorological satellites. *Proc. SPIE Europe Remote Sens*. 2008, Cardiff, Wales, ID, 15-18 September, 7107, No. 2, 8pp.

Roebber, P.J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.

Said, R. K.*,* M. B. Cohen*, and* U. S. Inan*,* 2013: Highly intense lightning over the oceans: Estimated peak currents from global GLD360 observation.*, J. Geophys. Res. Atmos*.*,* **118***,* 6905–6915.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630

SOS CHUVA, 2018: http://soschuva.cptec.inpe.br/soschuva/

Stone, K., Pinto, K., Steiner, M., and C. Kessinger, 2016: Ensemble Prediction of Oceanic Convective Hazards (EPOCH). *Presentation given to the FAA on 27-09-2016.*

WMO Joint Working Group on Forecast Verification, 2018: Forecast Verification. Accessed 26 Feb 2018, http://www.cawcr.gov.au/projects/verification/

10 APPENDICES

10.1 APPENDIX A: SUPPLEMENTAL TABLES

Table 13. Data coverage periods for NASA satellite cloud top height product.

Satellite	Cloud Product Resolution	Coverage Start	Coverage End
Meteosat-10	9 km	2017-12-01 0000 UTC	2018-02-09 1800 UTC
Meteosat-11	6 km	2018-02-09 1900 UTC	2018-03-31 2300 UTC
Meteosat-8	9 km	2017-12-01 0000 UTC	2018-02-09 1700 UTC
Meteosat-8	6 km	2018-02-09 1800 UTC	2018-03-31 2300 UTC
Himowari-8	8 km	2017-12-01 0000 UTC	2018-02-09 2100 UTC
Himowari-8	6 km	2018-02-09 2200 UTC	2018-03-31 2300 UTC
GOES-15 (West)	8 km	2017-12-01 0000 UTC	2018-03-31 2300 UTC
GOES-13 (East)	8 km	2017-11-30 2345 UTC	2017-12-22 1945 UTC
GOES-16 (East)	8 km	2017-12-22 2100 UTC	2018-02-09 2100 UTC
GOES-16 (East)	6 km	2018-02-09 2100 UTC	2018-03-31 2300 UTC

Table 14. Regional focus area definitions.

Region Name	North Limit	South Limit	West Limit	East Limit
WMO Area 141 – N America	60	25	215	310
WMO Area 143 – Asia	65	25	60	145
Australia / New Zealand	-10	-55	90	180
WMO area 145 – N Pacific	75	30	120	240
N Atlantic (Polar Stereographic)	50.1419716	25.3162384	243.1301	356.7594975
South Pacific (Sea Mask)	20	-40	150	277
South Atlantic (Sea Mask)	20	-40	-70	20
South Indian (Sea Mask)	20	-40	30	105
Southeast Asia (Sea Mask)	30	-10	95	155
Southeast Asia (Land Mask)	30	-10	95	155
South America (Land Mask)	15	-60	277	327
Southern Africa (Land Mask)	20	-40	-20	55

Table 15. Equalized forecast thresholds for the Part I assessment period (Jun 2017 – Sep 2017).

Terminology	EPOCH Threshold	WAFS Threshold	EPOCH Frequency	WAFS Frequency
"Occasional"	7.5%	40%	10.8%	10.8%
"Rare"	32.5%	50%	2.6%	2.6%

10.2 APPENDIX B: SUPPLEMENTAL FIGURES

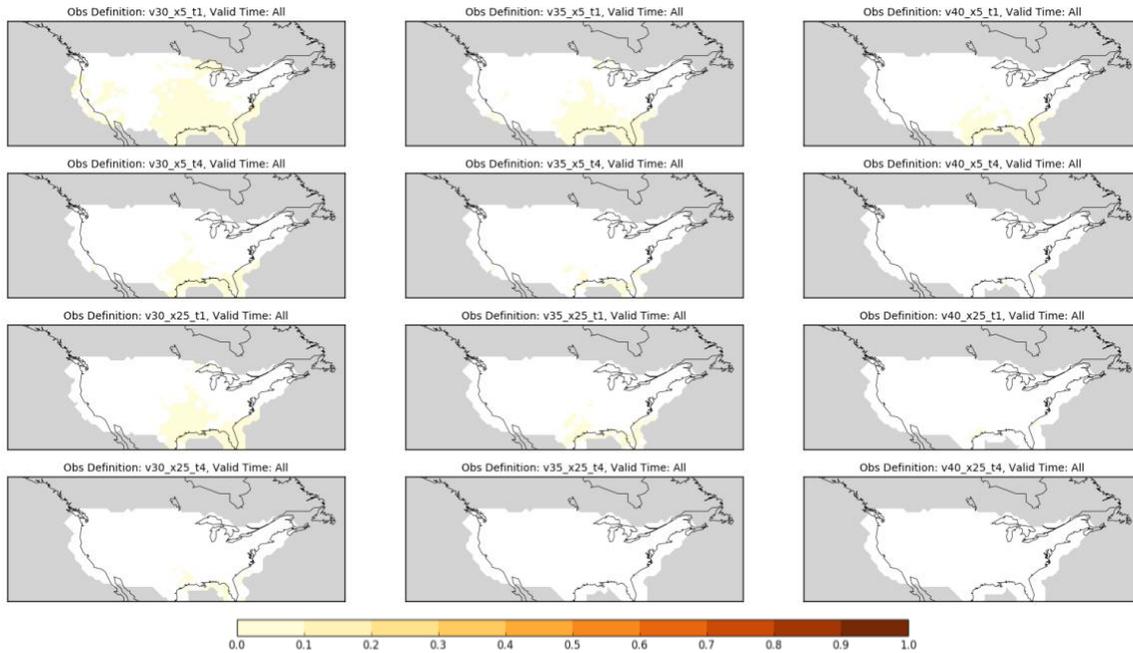### 10.2.1 CLIMATOLOGICAL MAPS OF OBSERVATIONAL DATASETS

Figure 36. MRMS Echo Top climatological maps. Intensity thresholds of 30 (left), 35 (middle), and 40 (right) kft are shown for the bounding spatial and temporal extent thresholds, from top to bottom: 5% and 1 hr, 5% and 4 hr, 25% and 1 hr, 25% and 4 hr.
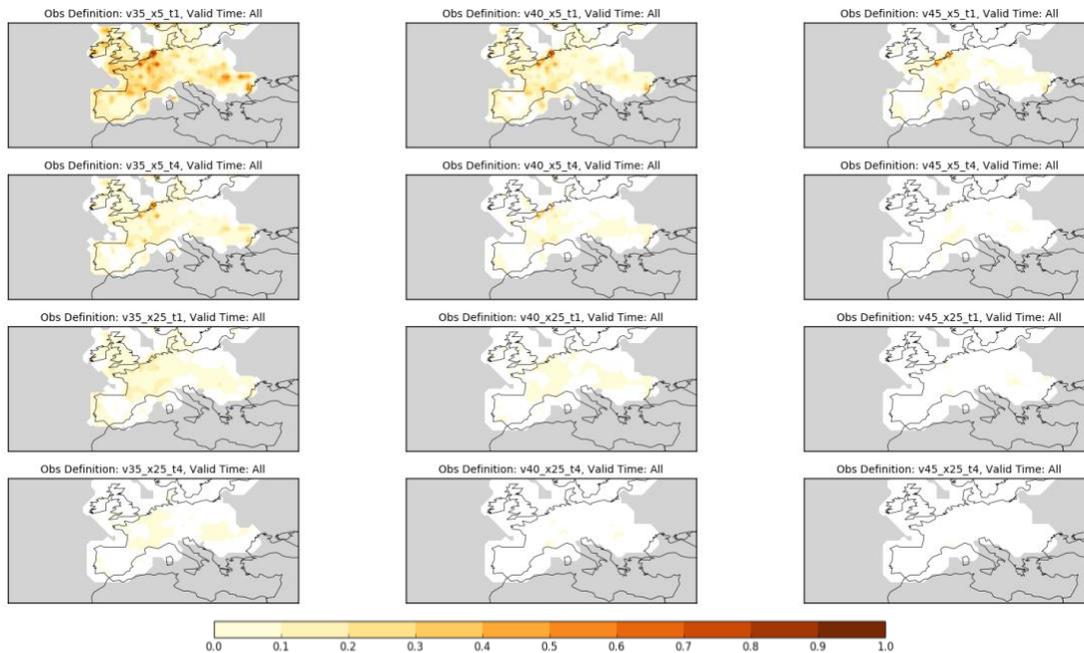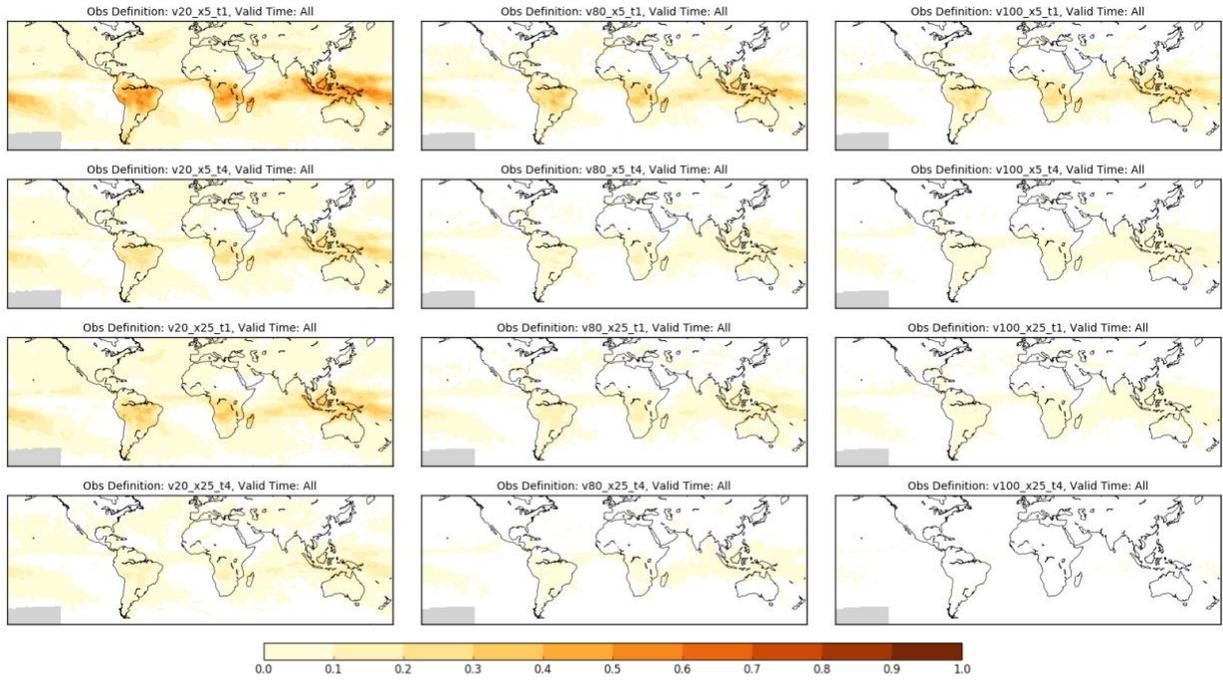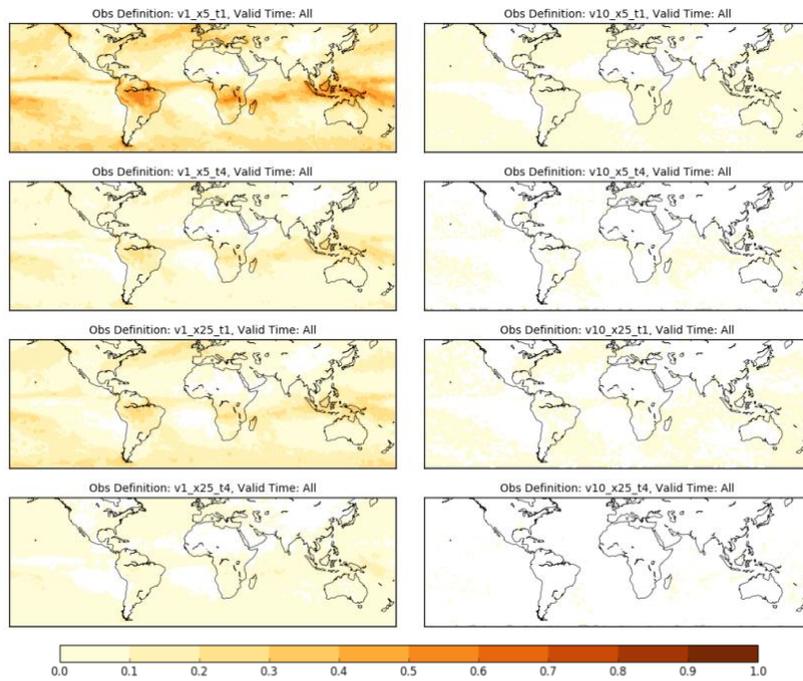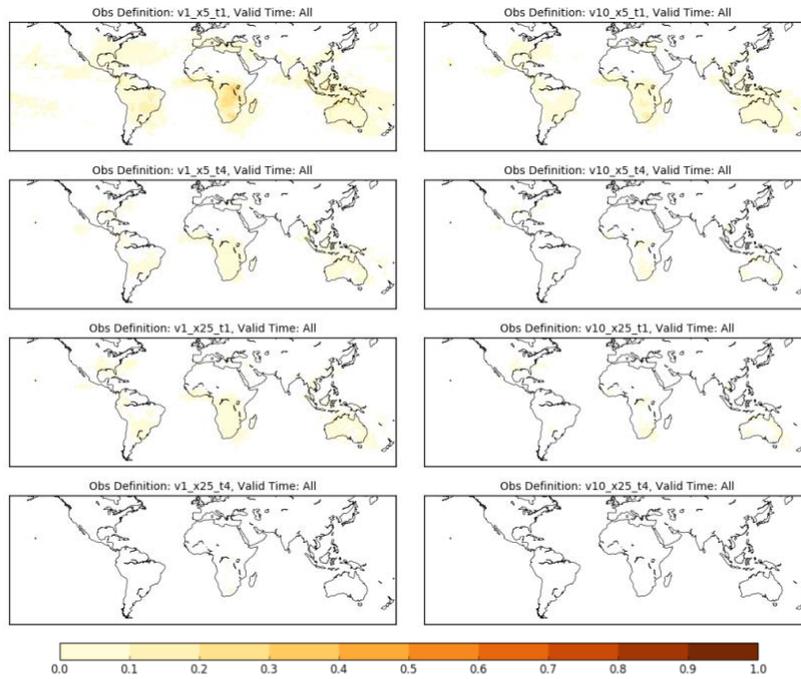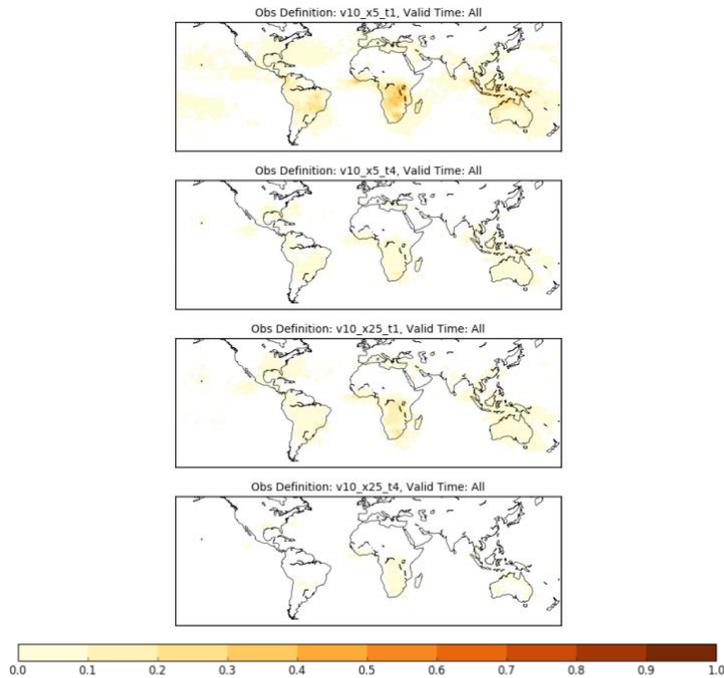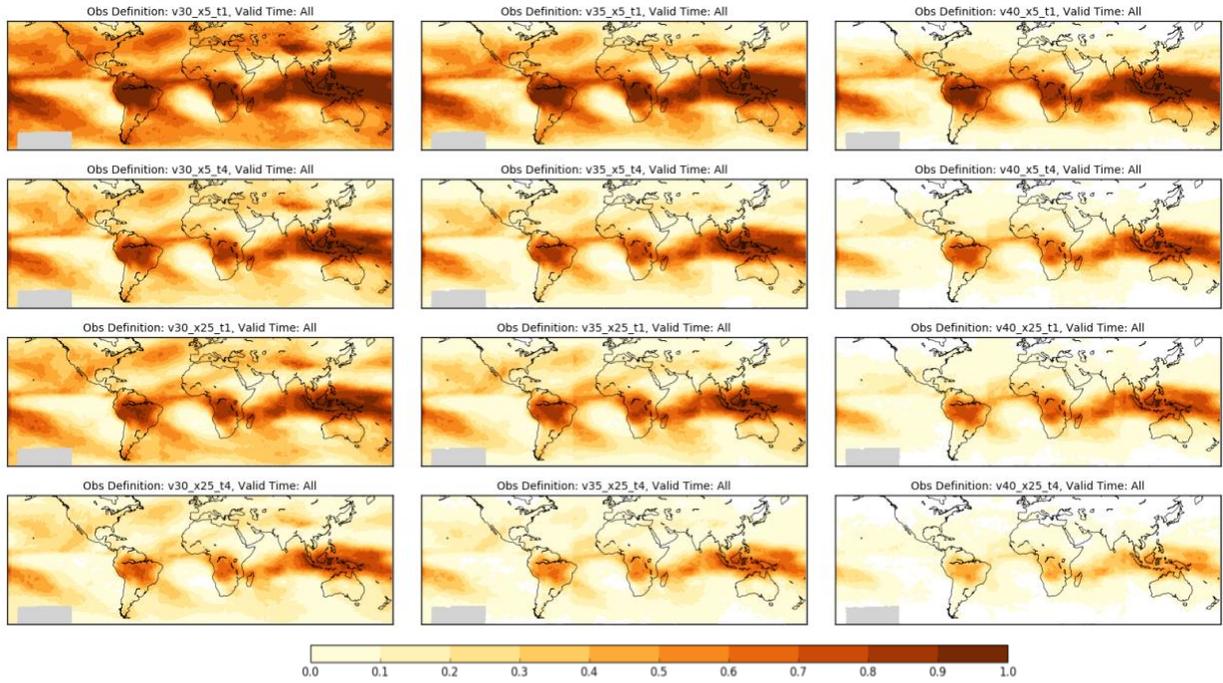

Figure 37. EU Reflectivity climatological maps. Intensity thresholds of 35 (left), 40 (middle), and 45 (right) dBZ are shown for the bounding spatial and temporal extent thresholds, from top to bottom: 5% and 1 hr, 5% and 4 hr, 25% and 1 hr, 25% and 4 hr.

Figure 38. IR Anvil Detection climatological maps. Intensity thresholds of 20 (left), 80 (middle), and 100 (right) are shown for the bounding spatial and temporal extent thresholds, from top to bottom: 5% and 1 hr, 5% and 4 hr, 25% and 1 hr, 25% and 4 hr.

Figure 39. IMERG climatological maps. Intensity thresholds of 1 (left) and 10 (right) mm/hr are shown for the bounding spatial and temporal extent thresholds, from top to bottom: 5% and 1 hr, 5% and 4 hr, 25% and 1 hr, 25% and 4 hr.

Figure 40. Lightning climatological maps. Intensity thresholds of 1 (left) and 10 (right) strikes are shown for the bounding spatial and temporal extent thresholds, from top to bottom: 5% and 1 hr, 5% and 4 hr, 25% and 1 hr, 25% and 4 hr.

Figure 41. Same as Figure 40, except for lightning within a search radius of 0.25 degrees (LTG25) at a single intensity threshold of 10 strikes.

Figure 42. Cloud Top Height climatological maps. Intensity thresholds of 30 (left), 35 (middle), and 40 (right) kft are shown for the bounding spatial and temporal extent thresholds, from top to bottom: 5% and 1 hr, 5% and 4 hr, 25% and 1 hr, 25% and 4 hr.
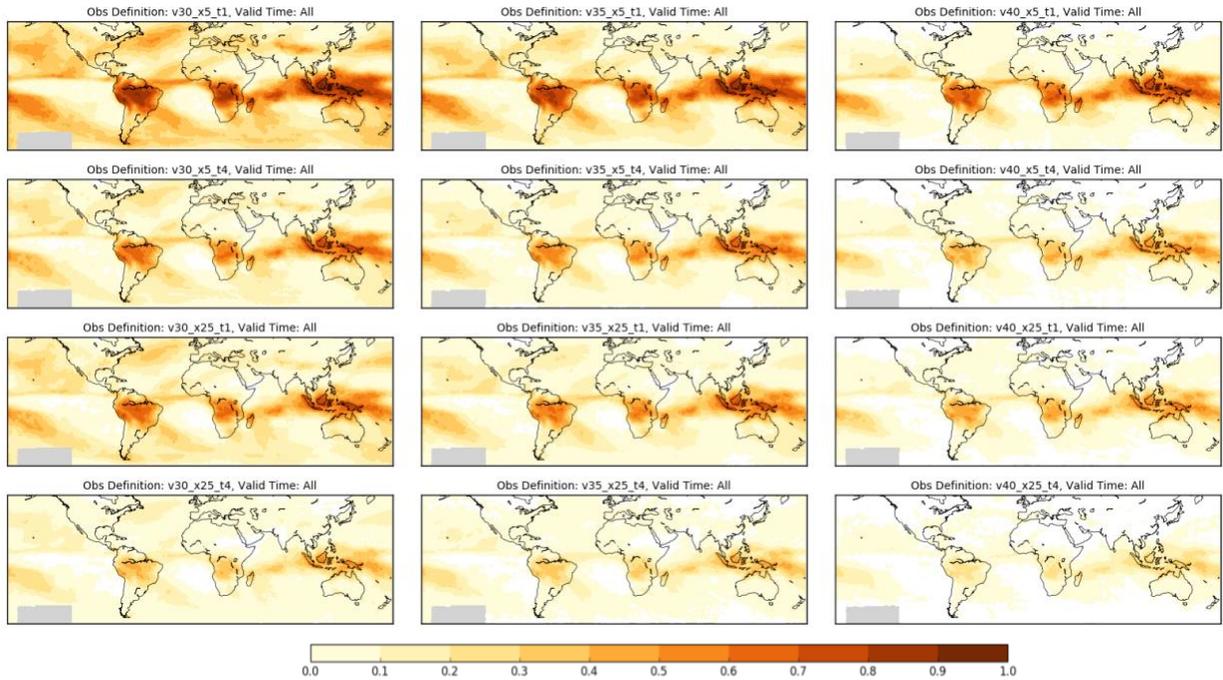
Figure 43. Same as Figure 42, except for the filtered cloud top height with minimum 15 kft cloud depth (CTH15).
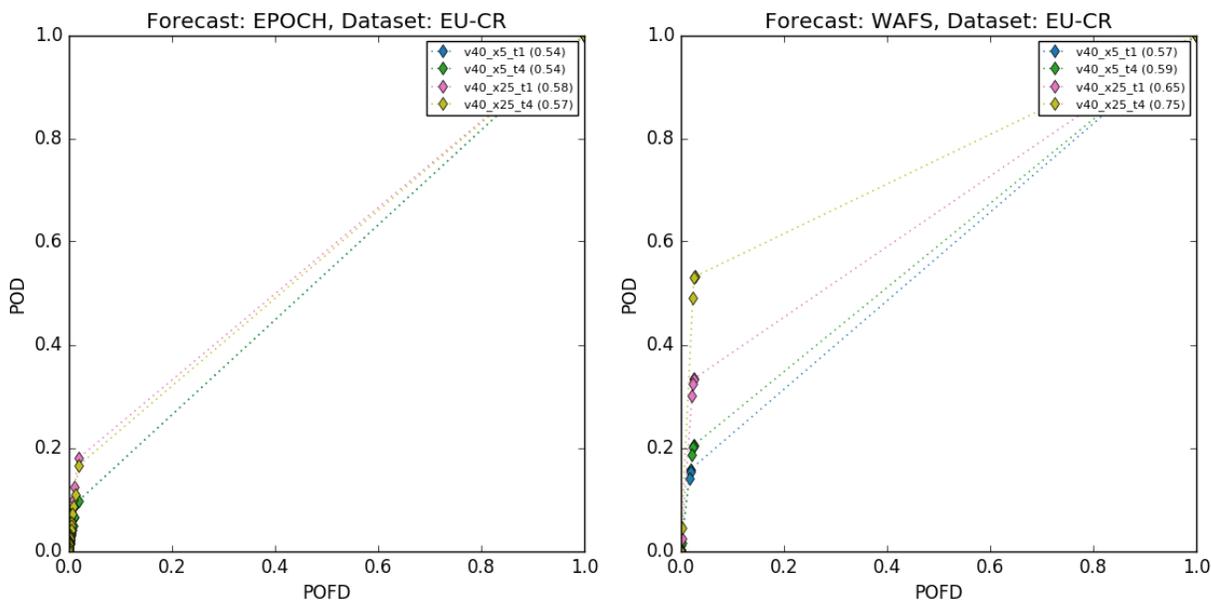
10.2.2 PERFORMANCE METRICS

Figure 44. ROC curves for EPOCH (left) and WAFS (right) compared against EU Composite Reflectivity data using an intensity threshold of 40 dBZ and bounding spatial coverages and temporal extents. The Area Under Curve values obtained using each observation definition are included in the legend.
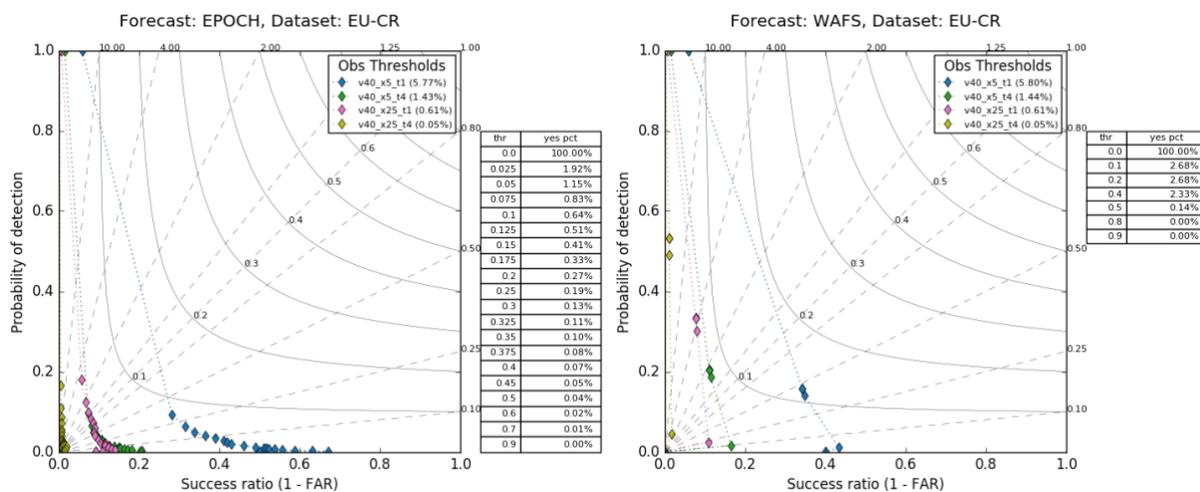

Figure 45. Performance diagrams for EPOCH (left) and WAFS (right) compared against EU Composite Reflectivity data using an intensity threshold of 40 dBZ and bounding spatial coverages and temporal extents. The dashed lines represent bias, while the solid curves give constant CSI values. The observed event frequencies using each observation definition are included in the legend. Also included is a table giving event frequencies by forecast threshold.
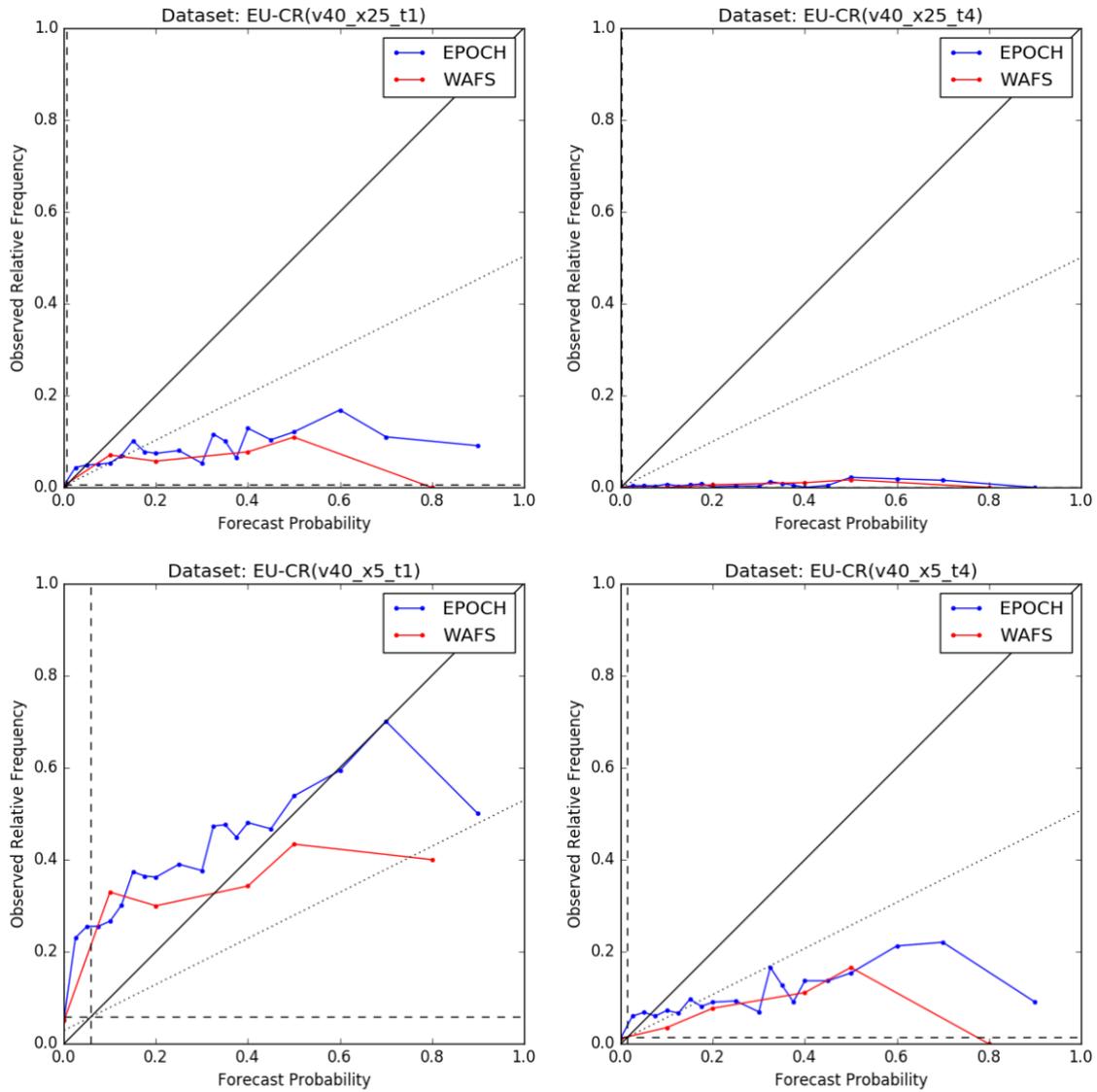
47

Figure 46. Reliability diagrams for EPOCH (blue) and WAFS (red) compared against EU Composite Reflectivity data. An observation intensity threshold of 40 dBZ is used along with the bounding spatial coverages (5%, bottom; 25% top) and temporal extents (1 hr, left; 4 hr, right). The black lines represent perfect reliability (solid), no skill (dotted), and climatology (dashed).
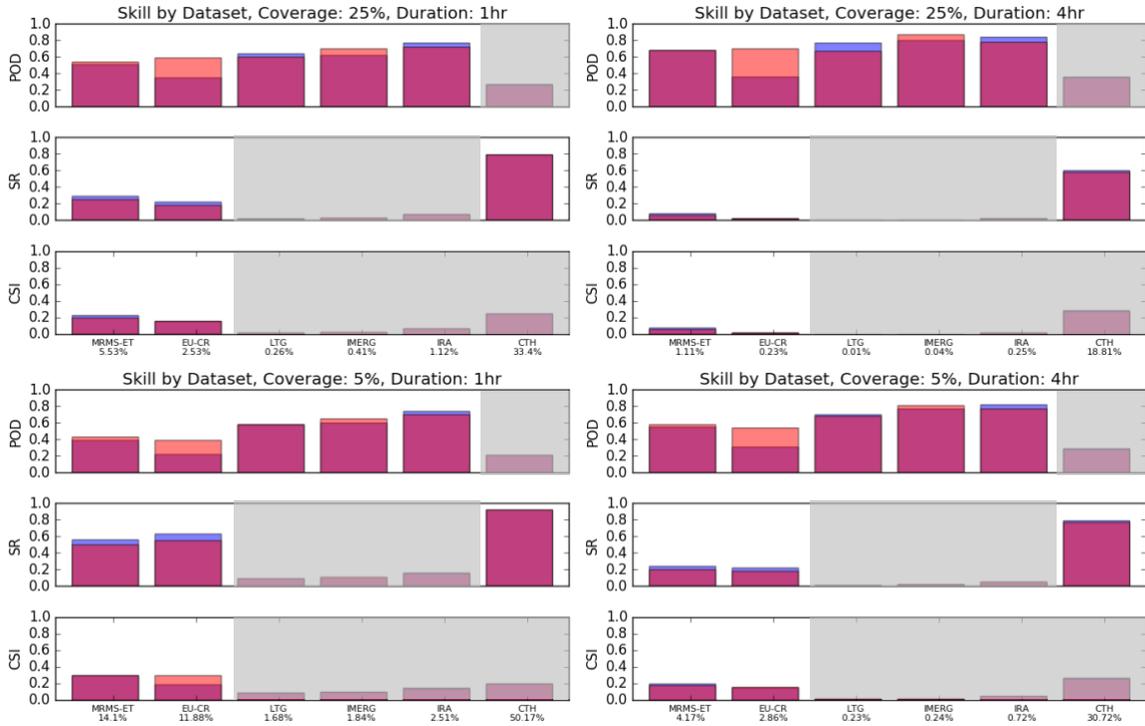
Figure 47. Performance metrics by dataset for the "occasional" forecast threshold. The scores for EPOCH (blue) and WAFS (red) are overlaid using 50% transparency. Gray shading indicates equivocal scores from bounding datasets. Spatial coverages of 5% (bottom) and 25% (top) and temporal extents of 1 hr (left) and 4 hr (right) are shown. The primary intensity thresholds for each dataset are given in Table 5.
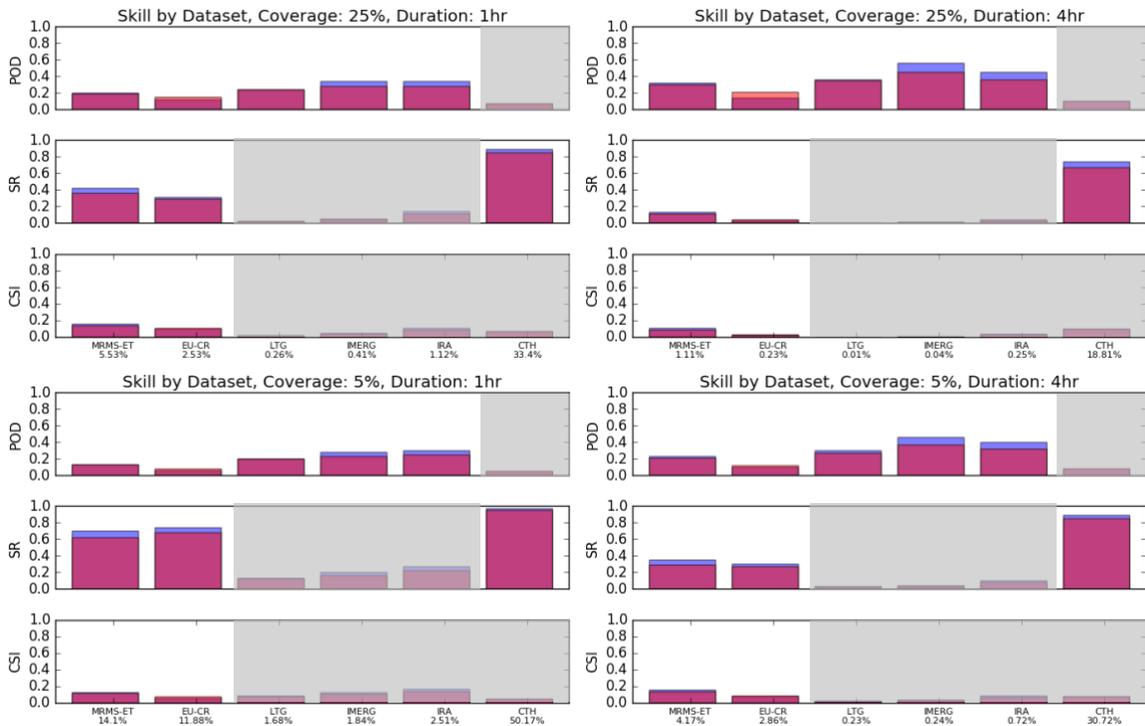

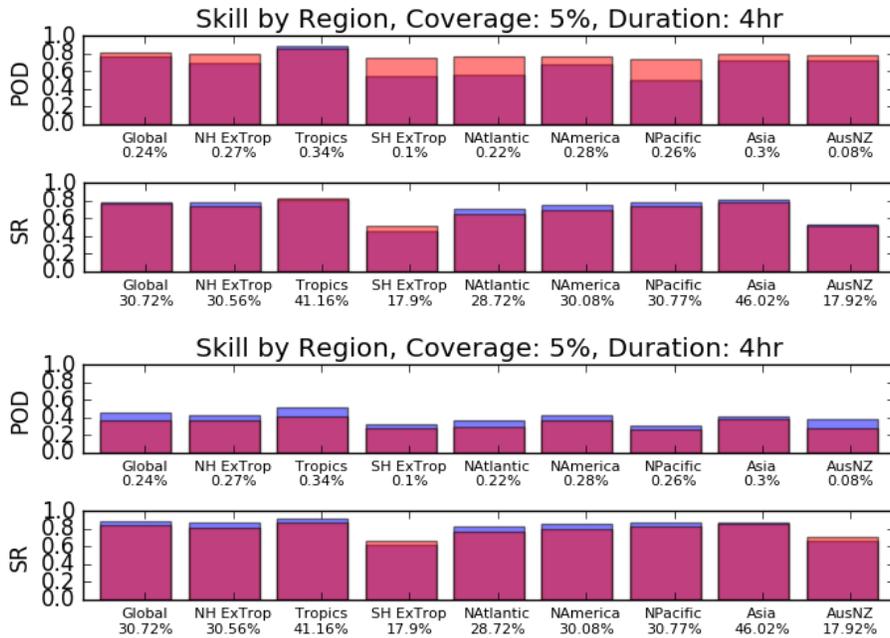**Figure 48. Same as Figure 47, except for the "rare" forecast threshold.**

Figure 49. Performance metrics results by region from the Part I assessment period (2017-06-01 to 2017-09-30) using the equalized forecast thresholds corresponding to "occasional" (top) and "rare" (bottom) forecast thresholds. The POD was obtained using IMERG ≥ 10 mm/hr, while SR was obtained using CTH ≥ 30 kft. In both cases, 5% spatial coverage and 4-hr temporal extent thresholds are shown.
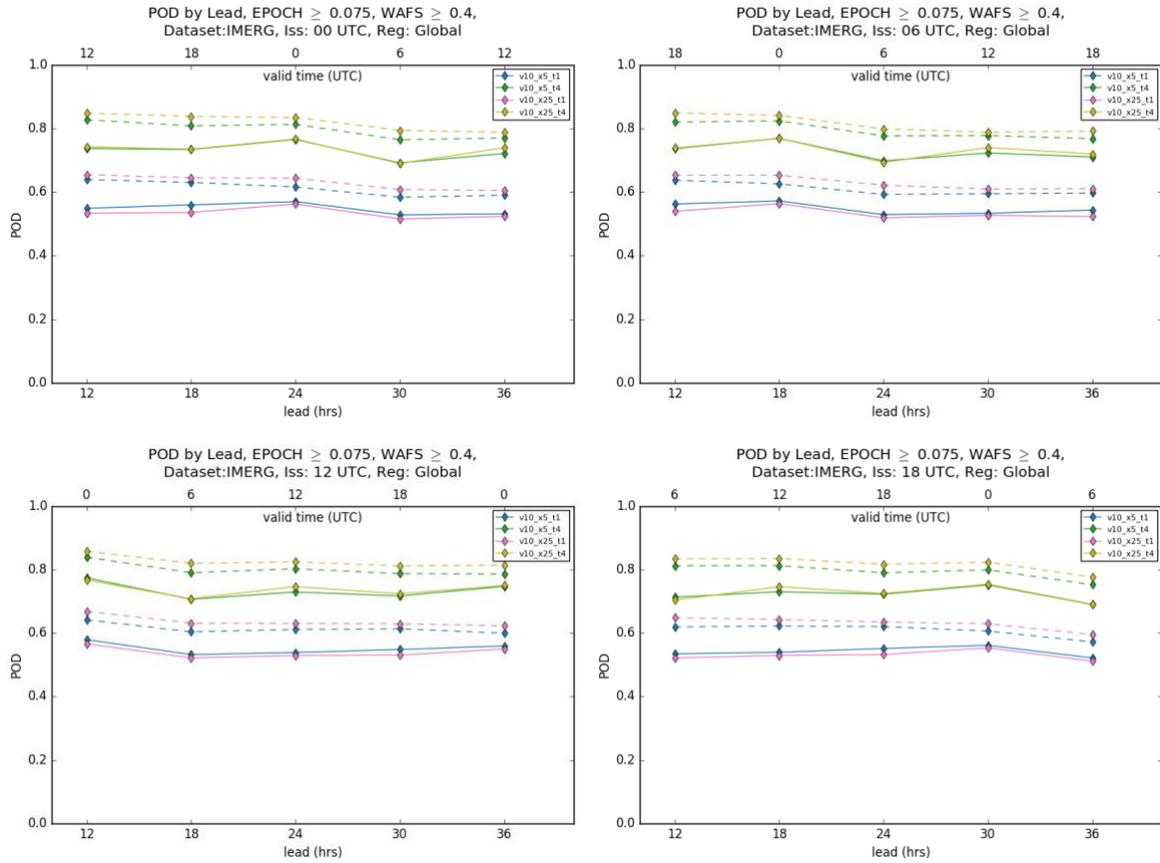
Figure 50. POD by issuance and lead time for EPOCH (solid) and WAFS (dashed) at the "occasional" threshold against IMERG ≥ 10 mm/hr for a range of spatial and temporal extents (colors). Each sub-diagram represents the results for a different issuance time: 00 UTC (top left), 06 UTC (top right), 12 UTC (bottom left), 18 UTC (bottom right).