Improving Short-Term QPF using Geostationary Satellite All-Sky Infrared Radiances: Real-Time Ensemble Data Assimilation and Forecast during the PRECIP 2020 and 2021 Experiments

Yunji Zhang[1*], Xingchao Chen[1], Michael M. Bell[2]

[1] *Center for Advanced Data Assimilation and Predictability Techniques, and Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania*

[2] *Department of Atmospheric Science, The Colorado State University, Fort Collins, Colorado*

Submitted to *Weather and Forecasting* as an article

Original Submission 29 August 2022

Revised 18 December 2022

Revised 31 January 2023

*Corresponding author: Yunji Zhang, yuz31@psu.edu

Early Online Release: This preliminary version has been accepted for publication in *Weather and Forecasting*, may be fully cited, and has been assigned DOI 10.1175/WAF-D-22-0156.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

The Prediction of Rainfall Extremes Campaign In the Pacific (PRECIP) aims to improve our understanding of extreme rainfall processes in the East Asian summer monsoon. A convection-permitting ensemble-based data assimilation and forecast system (the PSU WRF-EnKF system) was run in real-time in the summers of 2020 to 2021 in advance of the 2022 field campaign, assimilating all-sky infrared (IR) radiances from the geostationary Himawari-8 and GOES-16 satellites, and providing 48-hour ensemble forecasts every day for weather briefings and discussions. This is the first time that all-sky IR data assimilation has been performed in a real-time forecast system at a convection-permitting resolution for several seasons. Compared with retrospective forecasts that exclude all-sky IR radiances, rainfall predictions are statistically significantly improved out to at least 4–6 hours for the real-time forecasts, which is comparable to the timescale of improvements gained from assimilating observations from the dense ground-based Doppler weather radar network. The assimilation of all-sky IR radiances also reduced the forecast errors of large-scale environments and helped to maintain a more reasonable ensemble spread compared with the counterpart experiments that didn't assimilate all-sky IR radiances. The results indicate strong potential for improving routine short-term quantitative precipitation forecasts using these high-spatiotemporal-resolution satellite observations in the future.

Significance Statement

During the summers of 2020–2021, the PSU WRF-EnKF data assimilation and forecast system was run in real-time in advance of the 2022 Prediction of Rainfall Extremes Campaign In the Pacific (PRECIP), assimilating all-sky (clear-sky and cloudy) infrared radiances from the geostationary satellites into a numerical weather prediction model and providing ensemble forecasts. This study presents the first-of-its-kind systematic evaluation of the impacts of assimilating all-sky infrared radiances on short-term qualitative precipitation forecasts using multi-year, multi-region, real-time ensemble forecasts. Results suggest that rainfall forecasts are improved out to at least 4–6 hours with the assimilation of all-sky infrared radiances, comparable to the influence of assimilating radar observations, with benefits in forecasting large-scale environments and representing atmospheric uncertainties as well.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

1. Introduction

Extreme rainfall and its associated hazards, such as flash flooding and landslides, are impactful weather phenomena that threaten human lives and properties worldwide. Improving the short-term forecasts of extreme rainfall, i.e., quantitative precipitation forecasts (QPFs) at 0–12 hours forecast lead time, is therefore crucial. However, our knowledge of the fundamental processes that lead to extreme rainfall is incomplete and our ability to forecast these events remains limited.

To improve our understanding of the dynamical, thermodynamical, and microphysical processes associated with extreme rainfall as well as their predictability, the Prediction of Rainfall Extremes Campaign In the Pacific (PRECIP) was proposed to take place in Taiwan and Japan from late May to early August of 2020 as a collaborative effort across multiple universities and institutions internationally. The field phase of PRECIP was designed to observe extreme rainfall events ranging from local rainstorms, Meiyu fronts, to typhoons, in the moisture-rich environment of the East Asian monsoon region. Although the primary field phase of PRECIP was postponed to 2022 due to the COVID-19 pandemic, some activities were carried out during the same period of 2020 and 2021 summers in Taiwan and Japan as pilot studies; additionally, the Preparatory Rockies Experiment for the Campaign in the Pacific ("PRE"-CIP) was carried out from mid-July to mid-August of 2021 in Colorado, United States, during the North American monsoon season, providing datasets complementary to those gathered in Taiwan and Japan. Hereafter, we will refer to experiments in East Asia during the 2020 and 2021 summers as PRECIP2020 and PRECIP2021, and experiments in Colorado during the 2021 summer as "PRE"-CIP2021.

The holistic PRECIP field campaign was designed to have both observational and modeling components. The modeling component in the 2020 and 2021 summer seasons included 1) real-time convection-permitting deterministic forecasts using the Model for Prediction Across Scales (MPAS; Skamarock et al. 2012) with a global variable-resolution domain that uses a 3-km convection-permitting grid spacing in Taiwan and surrounding area and a 15-km grid spacing elsewhere; and 2) real-time convection-permitting ensemble forecasts using the Pennsylvania State University (PSU) Weather Research and Forecasting (WRF) ensemble Kalman filter (EnKF) regional data assimilation and forecast system (the PSU WRF-EnKF system; Zhang et al. 2009, Weng and Zhang 2012). The PSU WRF-EnKF system was run in real-time for PRECIP2020, PRECIP2021, and "PRE"-CIP2021 in advance of the 2022 primary field campaign. This study focuses on the 2020 and 2021 forecasts which were conducted with limited field observations due

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

to the COVID pandemic. The results from the 2022 observational field campaign and associated modeling components will be reported in a future study.

One of the highlights of the real-time PSU WRF-EnKF forecasts for PRECIP is the assimilation of satellite all-sky infrared (IR) brightness temperatures (BTs; used interchangeably with "radiance" hereafter) from the Advanced Himawari Imager (AHI) onboard the Himawari-8 geostationary satellite of Japan and the Advanced Baseline Imager (ABI) onboard the GOES-16 geostationary satellite of the United States. Improving QPF remains one of the biggest challenges of numerical weather predictions (NWPs). Assimilations of Doppler weather radar observations have been shown to improve short-term QPF at 0–6 hours forecast lead time (e.g., Xiao and Sun 2007; Aksoy et al. 2010; Clark 2012; Johnson et al. 2015; Surcel et al. 2015; Yussouf et al. 2016; Schwartz et al. 2021). Compared with Doppler weather radars that are mostly located inland and can only observe the rainfall after precipitation hydrometeors are formed, infrared sensors onboard geostationary satellites can detect the formation of clouds before the formation of precipitation with relatively high spatiotemporal resolutions over both ocean and land.

Currently, only clear-sky IR BTs are assimilated in operational NWP models at coarse resolutions (>10 km; Geer et al. 2018) due to the complexities associated with all-sky (i.e., clear-sky and cloudy) IR BT assimilations (e.g., Geer and Bauer 2011). Geer et al. (2018) shows that the European Centre for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), the National Center for Environmental Prediction (NCEP), the United Kingdom (UK) Met Office, and Météo France are assimilating clear-sky IR BTs from the geostationary satellites into their global models, and JMA and NCEP are developing the capability to assimilate all-sky IR BTs for their global models. Aside from all-sky IR BTs from geostationary satellites, operational NWP centers are also actively developing the capability of assimilating all-sky IR BTs from hyperspectral infrared sounders onboard low-Earth-orbiting (LEO) satellites, which have much more moisture-sensitive channels compared with the moisture-sensitive IR image channels of AHI and ABI, as well as all-sky microwave (MW) BTs from LEO satellites, which are sensitive to different types of hydrometeors at different spectral bands. However, the temporal resolutions of the LEO satellites are relatively coarse, such that they are suitable for the 3–6-hour assimilation time windows of the global models, but they are not able to capture the rapid development of the convective-to-mesoscale rainfall systems. On the other hand, ground-based Doppler weather radars have adequate spatiotemporal resolutions to resolve convection, but their limited spatial

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

coverage prohibits their monitoring of rainfall systems far from land. Besides, the blocking effects of topography, trees, and buildings further limit ground-based radars' ability to detect storms over some regions. IR BTs from geostationary satellites can provide continuous, seamless observations of rainfall systems with high spatiotemporal resolutions.

There have been significant enhancements in our ability to ingest all-sky IR BTs from geostationary satellites into convection-permitting regional models using ensemble-based data assimilation techniques (such as EnKF) in recent years. Numerous studies utilizing convection-permitting models have already shown that the assimilation of all-sky IR BTs can improve predictions of tropical cyclones (TCs; F. Zhang et al. 2016, 2019; Honda et al. 2018a; Minamide and Zhang 2018; Minamide et al. 2020; Hartman et al. 2021), severe thunderstorms (Y. Zhang et al. 2018, 2019; Jones et al. 2020), and other convective systems (Honda et al. 2018b; Otkin and Potthast 2019; Sawada et al. 2019; Chan et al. 2020b). However, their potential impacts on the short-term QPF (0–12-hour) have not been systematically explored using long-term and real-time data assimilation experiments.

In this study, we show that all-sky IR BTs have the potential to improve rainfall predictions in different regions through advanced ensemble-based data assimilation techniques. To the best of our knowledge, the real-time operation of the PSU WRF-EnKF system during PRECIP is the first time that a regional, convection-permitting, ensemble-based data assimilation and forecast system running in real-time has assimilated all-sky IR BTs. This study does not aim at improving our capability to assimilate all-sky IR BTs, but rather presents a systematic evaluation of the potential of improving the short-term QPF in real-time at both convective grey-zone and convection-permitting resolutions. The rainfall associated with Meiyu/Baiu fronts in China, Korea, and Japan in the warm season of 2020 is also above the climate average (e.g., Takaya et al. 2020), leading to devastating floods. This also provides a good opportunity to evaluate the impacts of all-sky IR data assimilation on the real-time forecasts of extreme rainfall.

2. Configurations of the real-time PSU WRF-EnKF system

2.1 Configurations for PRECIP2020

The PSU WRF-EnKF system configured for the PRECIP experiments combines the Advanced Research WRF (ARW/WRF; Skamarock et al. 2019) model, version 4.2, the ensemble square root filter (EnSRF; Houtekamer and Mitchell 2001) variation of EnKF, and the Community Radiative

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Transfer Model (CRTM; Han et al. 2006), version 2.3.0. For PRECIP2020, the system is configured with a 450×400×50 domain with a horizontal grid spacing of 9 km, covering East Asia and part of the western North Pacific, and a one-way nested 3-km domain of 300×300×50 grids covering Taiwan and adjacent regions (Fig. 1a). The hybrid sigma-pressure vertical coordinate is used to deal with the complex terrain over the Taiwan and Luzon islands and the Indochinese Peninsula, and the highest model level is located at 50 hPa. Employed physical parameterization schemes include the aerosol-aware version of the Thompson microphysics scheme (Thompson and Eidhammer 2014), the revised MM5 Monin-Obukhov scheme for surface layer processes (Jiménez et al. 2012), the Noah land-surface model (Tewari et al. 2004), the YSU scheme for PBL processes (Hong et al. 2006), and the RRTMG scheme for longwave and shortwave radiations (Iacono et al. 2008). No cumulus scheme is applied in both domains.

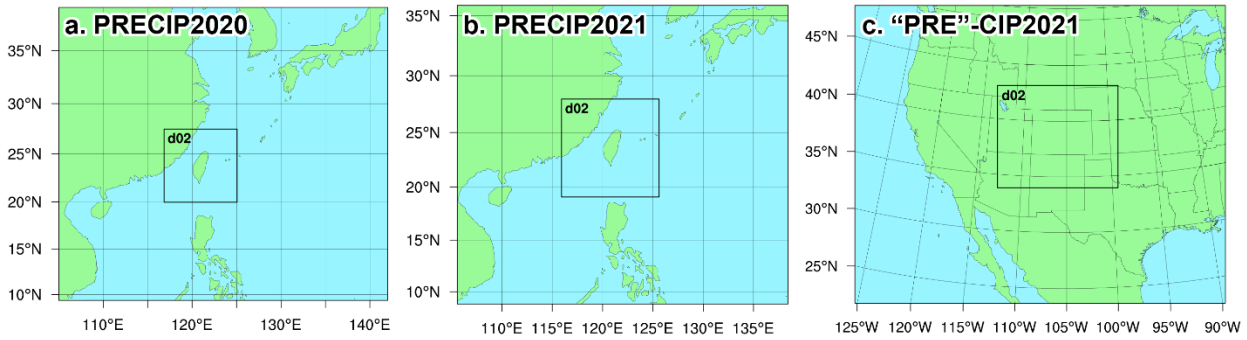

Figure 1. *PSU WRF-EnKF domain configurations for (a) PRECIP2020, (b) PRECIP2021, and (c) "PRE"-CIP2021 experiments.*

The PSU WRF-EnKF system uses 40 ensemble members, which is a balance between available computational resources and the need to reduce sampling errors and under-dispersive ensembles that result from limited ensemble members. Similar ensemble sizes of around 40 have been successfully used for other studies assimilating all-sky IR BTs using EnKF at convection-permitting resolutions (e.g., Otkin and Potthast 2019; Y. Zhang et al. 2018, 2019; Jones et al. 2020). The system also uses the Adaptive Observation Error Inflation (AOEI; Minamide and Zhang 2017) and the Adaptive Observation Error Background Inflation (ABEI; Minamide and Zhang 2018) to treat the nonlinearities and non-Gaussianities associated with the IR BTs, which have been proven to improve the assimilation of all-sky IR BTs by numerous studies (e.g., F. Zhang et al. 2019; Minamide and Zhang 2017, 2018, 2019; Y. Zhang et al. 2018, 2019, 2021a;

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Minamide et al. 2020; Chan et al. 2020b; Minamide and Posselt 2021; Hartman et al. 2021). The relaxation-to-prior-perturbation (RTPP; Zhang et al. 2004) method is used to maintain ensemble spread and combines 80% of the prior perturbations and 20% of the posterior perturbations.

The system was run in real-time from 20 May to 10 August 2020 once daily for PRECIP2020 (83 days in total). Initial conditions of the 40 ensemble members at 0000 UTC every day were generated using the 20-member Global Ensemble Forecasting System (GEFS) analyses at 0000 UTC of the same day, and 20-member 6-hour GEFS forecasts from 1800 UTC of the previous day valid also at 0000 UTC of the current day. Values at GEFS grids are horizontally and vertically interpolated to the WRF grids. After a spin-up period of 3 hours, hourly EnKF cycling data assimilations were performed till 1200 UTC (i.e., 10 data assimilation cycles in total). Assimilated observations including conventional surface, rawinsonde, and commercial aircraft observations from the Global Telecommunication System (GTS) acquired through the Meteorological Assimilation Data Ingest System (MADIS) and all-sky IR BTs from the 7.3-μm channel 10 (lower-tropospheric water vapor channel) of Himawari-8's AHI. The raw IR BTs were thinned to a horizontal spacing of 0.2° (roughly 20 km) for the outer domain and 0.04° (roughly 4 km) for the inner domain, and they were assimilated with no vertical localization and horizontal radii of influence (ROI; in terms of cut-off distance where the observations' impacts reduce to zero) of 100 km and 40 km for the outer and inner domains. No additional quality control or bias correction procedures were performed following previous practices (F. Zhang et al. 2019; Minamide and Zhang 2017, 2018, 2019; Y. Zhang et al. 2018, 2019, 2021a; Minamide et al. 2020; Chan et al. 2020b; Minamide and Posselt 2021; Hartman et al. 2021) due to the inability for a regional model to separate errors in model fields from systematic biases associated with the parameterization schemes and the radiative transfer model. After the final analysis at 1200 UTC of each day, a 40-member 48-hour ensemble forecast was carried out.

The daily 48-hour ensemble forecasts described above will be referred to as the "RT2020" ("real-time 2020") forecasts hereafter. As a comparison, a parallel experiment following the same manner as RT2020 but excluding IR BTs during cycling EnKF was performed retrospectively after the field campaign, and its ensemble forecasts will be referred to as the "NoIR2020" forecasts hereafter. Our purpose is to create "twin" forecasts that only differ from the real-time forecasts by not assimilating all-sky IR BTs, therefore the system configurations are kept identical. Due to limited extreme rainfall events near Taiwan during the pilot study resulting from the unexpectedly

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

calm western North Pacific in 2020, as well as limited computational resources, the retrospective NoIR2020 experiment only uses the 9-km outer domain for both the data assimilations and the ensemble forecasts, and the verification of the RT2020 and NoIR2020 forecasts will be focused on the 9-km domain as well. Although the 9-km grid spacing is unable to resolve isolated convective cells, this model resolution can capture the primary characteristics of mesoscale systems that contribute most monsoonal precipitation, and it has been widely used in simulating tropical and monsoonal rainfall with adequate skill (e.g., Wang et al. 2015; Chen et al. 2018, 2021, 2022; Chen and Zhang 2019; He et al. 2019; Ou et al. 2020; Ruppert and Chen 2020). Limited by computational resources, we also did not perform additional retrospective experiments to compare the impact of clear-sky versus all-sky IR BTs. Okamoto et al. (2019) showed that compared with when only clear-sky IR BTs are assimilated, the assimilation of all-sky IR BTs leads to reduced errors against independent conventional observations and satellite retrievals and improved rainfall forecasts.

2.2 Configurations for PRECIP2021 and "PRE"-CIP2021

Based on preliminary analyses performed after PRECIP2020, several adjustments were made to the PSU WRF-EnKF system for its PRECIP2021 and the "PRE"-CIP2021 (assimilating channel 10 IR BTs from GOES-16's ABI, which has an identical 7.3-μm wavelength to channel 10 of Himawari-8's AHI) real-time runs. Unless otherwise noted, all the changed configurations described below are applied to both the PRECIP2021 and the "PRE"-CIP2021 domains. The differences in the system configurations are also summarized in Table 1.

Table 1. *Differences in the configurations of the PSU WRF-EnKF system during the PRECIP2020, PRECIP2021, and "PRE"-CIP2021 experiments (see Section 2.2 for details).*

		PRECIP2020 (Taiwan)	PRECIP2021 (Taiwan)	"PRE"-CIP2021 (Colorado)
WRF Config	ARW-WRF Version	4.2	4.3	4.3
	9-km Domain Size	450×400×50	400×360×50	400×320×50
	3-km Domain Size	300×300×50	354×354×50	390×330×50
	Microphysics	Thompson aerosol	Thompson aerosol	Thompson
	Surface	Monin-Obukhov	Monin-Obukhov	MYNN
	Land Surface Model	Noah	Noah	RUC

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

	PBL	YSU	YSU	MYNN 2.5
	SKEB & SPPT	No	Yes	Yes
	Forecast Length	40-member 48-h	40-member 12-h & 20-member 48-h	40-member 12-h & 20-member 48-h
EnKF Config	Satellite & Sensor	Himawari-8 AHI	Himawari-8 AHI	GOES-16 ABI
	9-km-domain BT ROI	100 km	40 km	40 km
	Multiplicative Inflation	No	1%	1%

The latest version 4.3 of ARW/WRF was used instead of version 4.2 for PRECIP2020. The domain configurations for PRECIP2021 were adjusted (Fig. 1b) with the area covered by the 3-km domain increased by 40% to 354×354 compared with the PRECIP2020 configuration of 300×300 and the area covered by the 9-km domain slightly decreased to 400×360 to compensate the increased 3-km computational costs. The 9-km domain (400×320×50) of "PRE"-CIP2021 covers the western 2/3 of the contiguous United States (CONUS) and the 3-km domain (390×330×50) covers Colorado and the surrounding area. While PRECIP2021 inherits the same set of physical parameterization schemes from PRECIP2020, the "PRE"-CIP2021 runs use a slightly different combination of parameterization schemes, including the Thompson et al. (2008) microphysics, the Mellor-Yamada-Nakanishi-Niino (MYNN) surface scheme (Nakanishi and Niino 2004), the RUC land surface model (Benjamin et al. 2004), and the MYNN 2.5-level TKE scheme for PBL processes (Nakanishi and Niino 2004). These schemes are also used by the operational High-Resolution Rapid Refresh (HRRR) model and are expected to provide satisfactory performance for convection over CONUS at a 3-km horizontal grid spacing.

One outstanding issue associated with the PRECIP2020 forecasts is that the ensemble forecasts are generally under-dispersive (see Section 4 for more details), which is also an issue that has been frequently observed in regional ensemble prediction systems (e.g., Schumacher and Clark 2014; Schwartz et al. 2014; Hagelin et al. 2017). To combat the ensemble under-dispersiveness, we applied stochastic kinetic energy backscatter (SKEB; Mason and Thompson 1992) and stochastically perturbed physics tendencies (SPPT; Buizza et al. 1999) schemes for the PRECIP2021 and "PRE"-CIP2021 runs with different random seeds assigned for each of the ensemble members, as well as multiplicative covariance inflation of prior perturbations for EnKF (Anderson and Anderson 1999) by increasing the magnitudes of EnKF prior perturbations by 1% before assimilating observations for all EnKF cycles. The impacts of applying these additional

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

methods are examined in Section 4. The horizontal ROI for the outer 9-km domain is also reduced from 100 km to 40 km based on our recent study on the structure of correlations between all-sky IR BTs and atmospheric states (Zhang et al. 2022a).

Additionally, preliminary quantitative verifications based on the PRECIP2020 forecasts show that QPF is generally improved out to 6 to 12 hours when all-sky IR BTs are assimilated (see Section 4 for more details). Therefore, for PRECIP2021 and "PRE"-CIP2021, the size of the ensemble was reduced from 40 to 20 after the first 12 hours, providing a combination of 40-member 12-hour ensemble forecasts and 20-member 48-hour ensemble forecasts for each run. Clark et al. (2011) and Schwartz et al. (2014, 2019) showed that ensemble forecasts with 10–20 members can adequately quantify the uncertainties associated with the forecasts.

The real-time data assimilation and forecast experiments for PRECIP2021 (Taiwan) and "PRE"-CIP2021 (Colorado) will be referred to as RT2021TW and RT2021CO, respectively. To facilitate similar verifications on the impact of assimilating all-sky IR BTs, we also performed retrospective "NoIR2021CO" runs. Different from the NoIR2020 runs, the NoIR2021CO runs implemented both domains of RT2021CO to properly resolve the circulations of localized convective storms that are frequently observed during the "PRE"-CIP2021 period in the United States, and for "PRE"-CIP2021 we will only evaluate the 3-km forecasts. Since NoIR2020 can be used to evaluate the potential impacts of all-sky IR on the rainfall forecast during the East Asian summer monsoon, we decided not to perform NoIR2021TW due to limited computational resources, and RT2021TW will not be evaluated in this current study. Both RT2021CO and NoIR2021CO were initialized twice daily at 0000 UTC and 1200 UTC from 15 July 2021 to 15 August 2021 (33 days and 66 forecasts in total). Combined with RT2020 and NoIR2020 from PRECIP2020, they will provide more comprehensive evaluations on the impacts of assimilating all-sky IR BTs with two different regions at two different years.

3. Verification metrics

Two datasets are used to evaluate the impact of assimilating all-sky IR BTs on QPF. For the 9-km PRECIP2020 forecasts of RT2020 and NoIR2020, we use the 0.1°×0.1° Integrated Multi-Satellite Retrievals for GPM (IMERG; Huffman et al. 2019) rainfall estimates produced by the Global Precipitation Measurement (GPM) project that combines multiple microwave sensors from the GPM constellation satellites, as well as all-sky infrared radiances and surface rain gauges.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Although certain deficiencies exist (e.g., Lee et al. 2019, Chen et al. 2022), IMERG is one the most reliable estimations of global precipitation with high spatiotemporal resolutions and seamless coverage over both land and sea. For the 3-km "PRE"-CIP2021 forecasts of RT2021CO and NoIR2021CO, we use the Stage IV rainfall estimates (Lin and Mitchell 2005), which have a 4-km grid spacing and are generated by the NCEP's Environmental Modeling Center (EMC) using ground-based radars and rain gauges. In addition to QPF verifications, rawinsonde observations from the MADIS are used to evaluate the forecasts of large-scale environments. Due to the requirement of the workflow of the real-time system, no rawinsonde observations are assimilated for the last EnKF cycles (at 0000 UTC or 1200 UTC) because these observations came in much later than the time required to provide the ensemble forecasts, therefore the ensemble forecasts at 0-hour forecast lead time (identical to the analysis of the last EnKF cycle at 0000 UTC or 1200 UTC) are independent of verified rawinsondes. The WRF outputs are interpolated to the IMERG or Stage IV grid and rawinsonde locations using linear interpolation.

The primary focus of this study is QPF, which is evaluated using the equitable threat score (ETS; Wilks 2011), the area under the receiver-operating-characteristic (ROC) curve (AUC; Marzban 2004), and the fraction skill score (FSS; Roberts and Lean 2008). Information from the entire ensemble is used in the calculations of all three metrics.

ETS is a pointwise, deterministic metric, with higher scores representing more accurate forecasts. Using a 2-by-2 contingency table that determines the hits, misses, and false alarms of a deterministic forecast compared with the observations, ETS is formulated as

$$ETS = \frac{hits - hits_{random}}{hits + misses + false\ alarms - hits_{random}},$$

where $hits_{random} = (hits + misses) \times (hits + false\ alarms)/total\ events$ represents the reference value of hits for a completely random forecast. To include the information from the entire ensemble, the probability-matched mean (PMM; Ebert 2001) is used to calculate ETS instead of the arithmetic mean. The PMM preserves the cumulative distribution function (CDF) of rainfall of the entire ensemble while keeping the horizontal structure of the ensemble arithmetic mean at the same time, and the arithmetic mean is not able to present the extreme values of the entire ensemble.

AUC is a pointwise, probabilistic metric evaluating the resolution of the forecasts on the occurrence of events exceeding a certain threshold, with higher scores representing better discriminations between events and non-events. A 2-by-2 contingency table is first generated

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

considering whether the forecast probability of a certain rainfall amount exceeds a certain probability threshold; this is different from the contingency table for ETS, which considers whether a deterministic forecast exceeds a certain rainfall threshold. Then, the ROC curve is generated by plotting the probability of detection, $hits/(hits + misses)$, against the probability of false detection, $false\ alarms/(correct\ negatives + false\ alarms)$, for a set of probability thresholds ranging from 0 to 1 using the forecast probability contingency table. Lastly, AUC is calculated based on the ROC curve using the trapezoidal approximation.

FSS is a neighborhood, probabilistic metric based on neighborhood ensemble probability (NEP; Schwartz and Sobash 2017), with higher scores representing more accurate rainfall forecasts at a certain neighborhood radius (also called the horizontal or spatial scale of a given FSS calculation). NEPs at each grid point given a certain rainfall threshold and a certain neighborhood radius (spatial scale) for the forecasts and the corresponding neighborhood probability (NP; Schwartz and Sobash 2017) of the observations are first calculated, then FSS for a certain combination of rainfall threshold and neighborhood radius is formulated as

$$FSS = 1 - \frac{\sum_{i=1}^{N}\left(NEP_i^f - NP_i^o\right)^2}{\sum_{i=1}^{N}\left[\left(NEP_i^f\right)^2 + (NP_i^o)^2\right]},$$

where NEP_i^f and NP_i^o represents NEP of the forecasts and NP of the observations at the i-th of the total N grid points. A smaller difference between NEP and NP results in a higher FSS value. When a larger neighborhood radius is used, it is more likely for the NEPs of the forecasts and the observations to overlap, therefore FSS often increases with increasing neighborhood radius when the rainfall threshold is fixed.

When calculating ETS, AUC, and FSS, we adopted percentile thresholds that vary with time instead of fixed physical thresholds. The percentile thresholds can help correct some model biases and facilitate verifications that focus on the forecast skills of spatial patterns of given rainfall occurrence frequencies (Dey et al. 2014; Gowan et al. 2018; Schwartz 2019). For each day's forecasts at each forecast lead time, the physical thresholds are established by calculating the 95th, 97th, 99th, 99.5th, and 99.9th percentiles of the rainfall values produced by the entire ensemble, and ETS, AUC, and FSS are calculated against physical thresholds that are similarly established for the IMERG and Stage IV estimates that cover the same region of the verified model domain.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

For ETS, because PMM preserves the rainfall CDF of the entire ensemble, the physical values of PMM's percentiles are identical to those of the entire ensemble.

Environmental conditions are evaluated using root-mean-square errors (RMSEs) and ensemble standard deviations (STDs). RMSEs are verified against rawinsonde observations, and STDs are calculated against the ensemble mean. Smaller RMSEs indicate smaller errors, while STDs that are comparable to the corresponding RMSEs – neither too large nor too small – suggest that the ensemble spread adequately represents the uncertainty of the atmosphere.

We use the Wilcoxon signed rank test (Wilks 2011) for the statistical significance test, which does not assume any specific form of the distributions. For each rainfall threshold and/or forecast lead time, we first calculate the differences between the above-mentioned metrics obtained using RT forecasts and NoIR forecasts (we focus more on the one-on-one paired differences between these two forecasts, rather than the differences between their overall distributions), then we apply the Wilcoxon signed rank test to the collected differences (with sample sizes of 83 for PRECIP2020, i.e., RT2020 and NoIR2020, and 66 for "PRE"-CIP2021, i.e., RT2021CO and NoIR2021CO) to examine whether these differences are statistically significantly greater or smaller than 0. All statistical significance tests are performed for the 95% and 99% confidence levels.

Due to the significant computational costs to run CRTM and limited storage space that requires the outputs of the ensemble forecasts to be post-processed shortly after they are generated to keep only the most essential model fields, no systematic verifications of simulated IR BTs against AHI or ABI observations are performed.

4. PRECIP2020 Verification

We first present the evaluations of the forecasts during the PRECIP2020 experiments in the East Asia and western North Pacific region, namely, RT2020 and NoIR2020 forecasts. All the results in this section are obtained using the forecasts from the 9-km outer domain.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

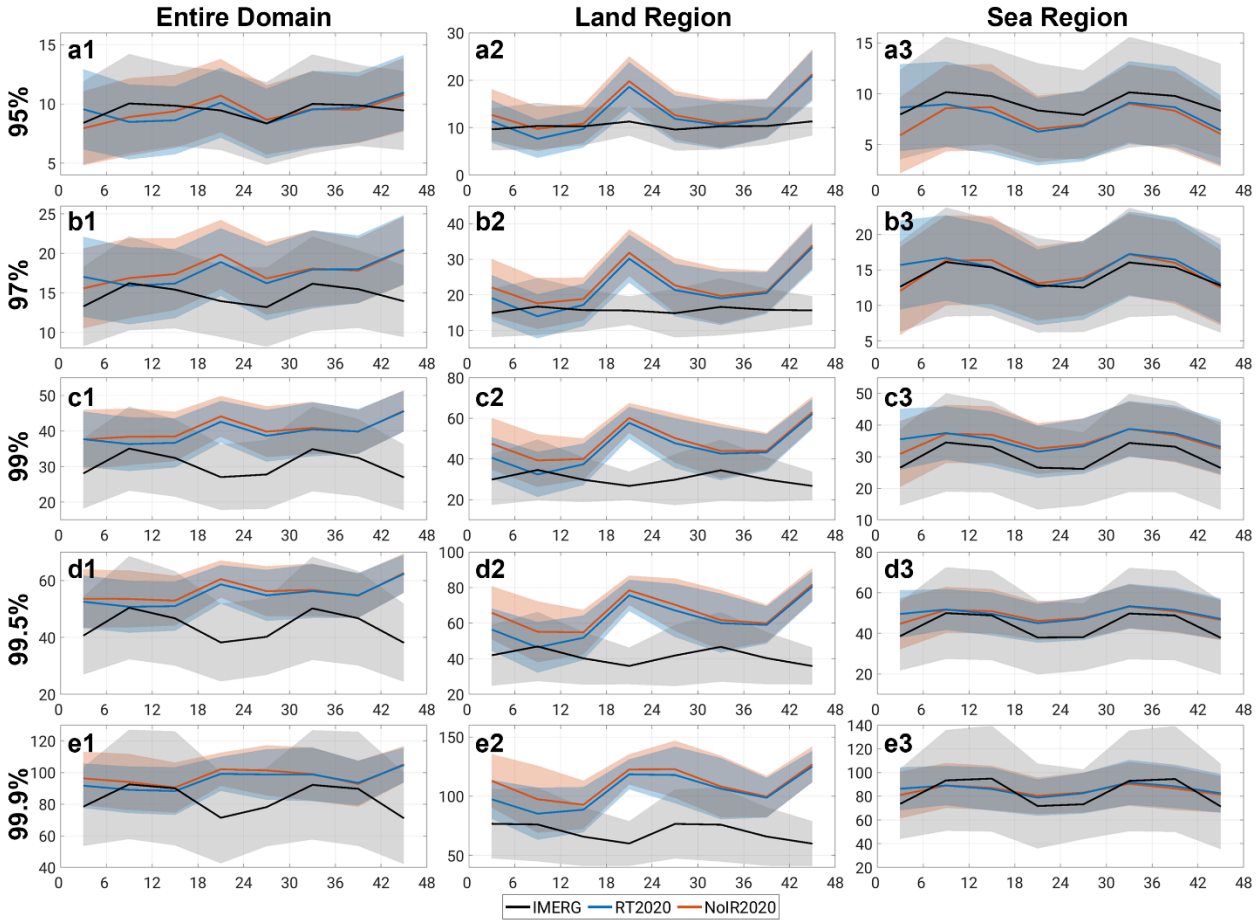

Figure 2*. Distributions of physical thresholds with respect to forecast lead times at the (row a) 95th, (row b) 97th, (row c) 99th, (row d) 99.5th, and (row e) 99.9th percentiles over (first column) the entire domain, (second column) the land region, and (third column) the sea region. Solid lines are the mean values of the observations or the forecasts. Shadings mark the range of one standard deviation. Values from the entire ensemble are used for the forecasts.*

Figure 2 shows the distributions of physical thresholds at different percentiles with respect to forecast lead times for the RT2020 and NoIR2020 forecasts and the IMERG estimates. The mean values of the physical thresholds of RT2020 and NoIR2020 are comparable with those of IMERG at the 95th percentile (Fig. 2a1), but they become higher than the IMERG's physical values when moving towards higher percentiles, suggesting a slight overprediction of rainfall at these percentiles (Fig. 2b1–e1). An outstanding characteristic is the offset of the diurnal cycles in the forecasts' physical thresholds compared with those from IMERG. If we divide the model domain into the land region and the sea region, calculate the physical thresholds of the percentiles within

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

each region, and compare the temporal evolutions of the forecasts' physical thresholds at these two regions with the corresponding IMERG thresholds at these two regions, it is apparent that the forecasts generally have a good match with IMERG (both diurnal cycles and magnitudes of the thresholds) at the sea region (Fig. 2a3–e3), while notable discrepancies occur at the land region (Fig.2a2–e2). At the land region, the 18–24-hour and 42–48-hour forecast diurnal peaks correspond to the local afternoon of 0600–1200 UTC (for Taiwan, local time = UTC + 8), while the 6–12-hour and 30–36-hour IMERG diurnal peaks correspond to the local late night to the early morning of 1800–0000 UTC. This late-night to early-morning rainfall peak is frequently observed in the coastal regions of East Asia (Chen et al. 2009; Xu and Zipser 2011). It is often associated with land-sea breezes, low-level/boundary-layer jet, and their interactions (Chen et al. 2016; Du and Rotunno 2018; Du and Chen 2019), and the numerical models nowadays still lack sufficient skills in simulating these mechanisms even with convection-permitting grid spacings, resulting in unsatisfactory skills in forecasting late night to early morning rainfall (e.g., Huang et al. 2022; Lu et al. 2022). The necessary horizontal and vertical resolution to accurately capture this late-night to early-morning rainfall peak in East Asian coastal regions still deserves further studies.

Using percentile thresholds can partly mitigate these systematic model biases. RT2020 forecasts show generally higher ETSs than NoIR2020 forecasts for the first 24 hours, although their differences diminish with the increasing forecast lead time or the increasing percentile threshold (Fig. 3a). These improvements are significant at the 99% confidence level for all percentiles except for the 99.9th percentile at the first 6 hours, and significant at least at the 95% confidence level except for the 95th percentile at 6–12 hours (Fig. 3a). The improvements last slightly longer at the land region (improvements significant at the 99% confidence level for percentiles up to the 99th for 6–12 hours; Fig. 3b) than the sea region (mostly improvements are only significant for 0–6 hours; Fig. 3c). Additionally, RT2020's AUCs are higher than those of NoIR2020 and are significant for most percentile thresholds for 0–12 hours; the significance also extends to 48 hours for the 95th and 97th percentiles (Fig. 3d–f). This suggests that the RT2020 forecasts better discriminate the rainfall occurrence than the NoIR2020 forecasts, and the improvements can be further extended with additional calibrations and bias corrections (although percentile thresholds already serve as crude bias correction processes).

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

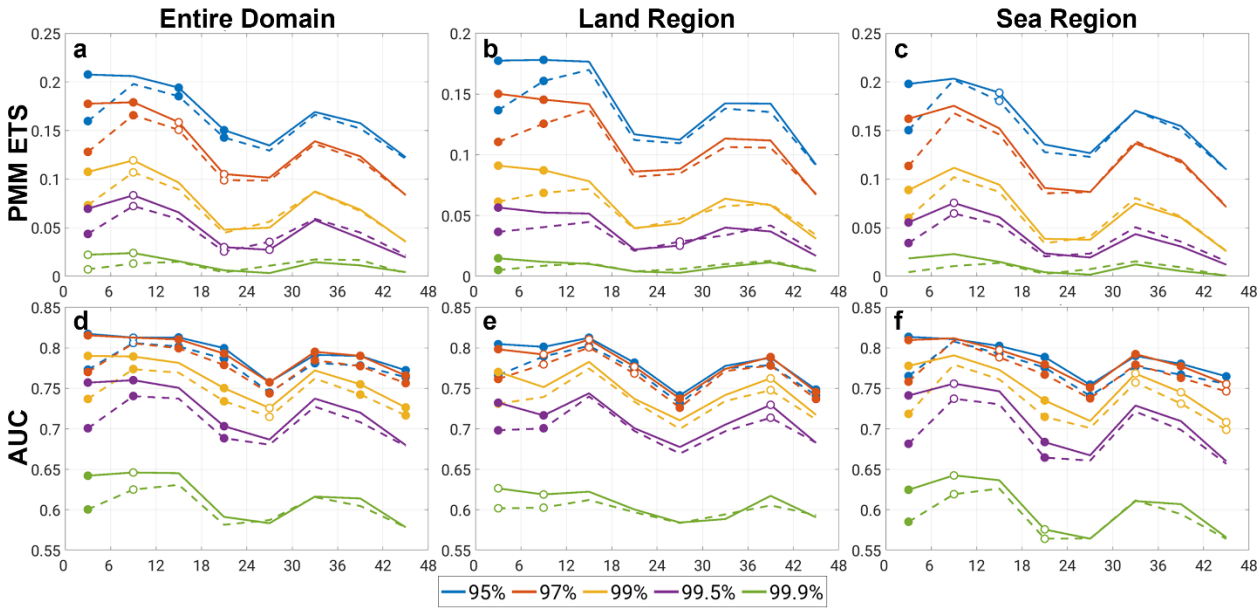

Figure 3*. (First row) ETS of the PMM and (second row) AUC for the PRECIP2020 experiments over (first column) the entire domain, (second column) the land region, and (third column) the sea region. Solid and dashed lines represent scores of RT2020 and NoIR2020, respectively. Open circles and filled circles indicate that the differences between the two experiments' scores are statistically significant at the 95% and 99% confidence levels, respectively.*

As previously explained, ETS and AUC are both pointwise metrics. Neighborhood verifications using FSS suggest that the assimilation of all-sky IR BTs improves short-term rainfall forecasts at both small and large spatial scales (neighborhood radii) too: similar to ETS and AUC, RT2020's FSSs are higher than those of NoIR2020 for 0–6-hour rainfall forecasts for all percentile thresholds at all horizontal scales from 20 km to 200 km, regardless of whether it is over the land region or the sea region (Fig. 4). Almost all these multiscale improvements at 0–6-hour lead time are significant at the 99% confidence level (only a few at 95% level).

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

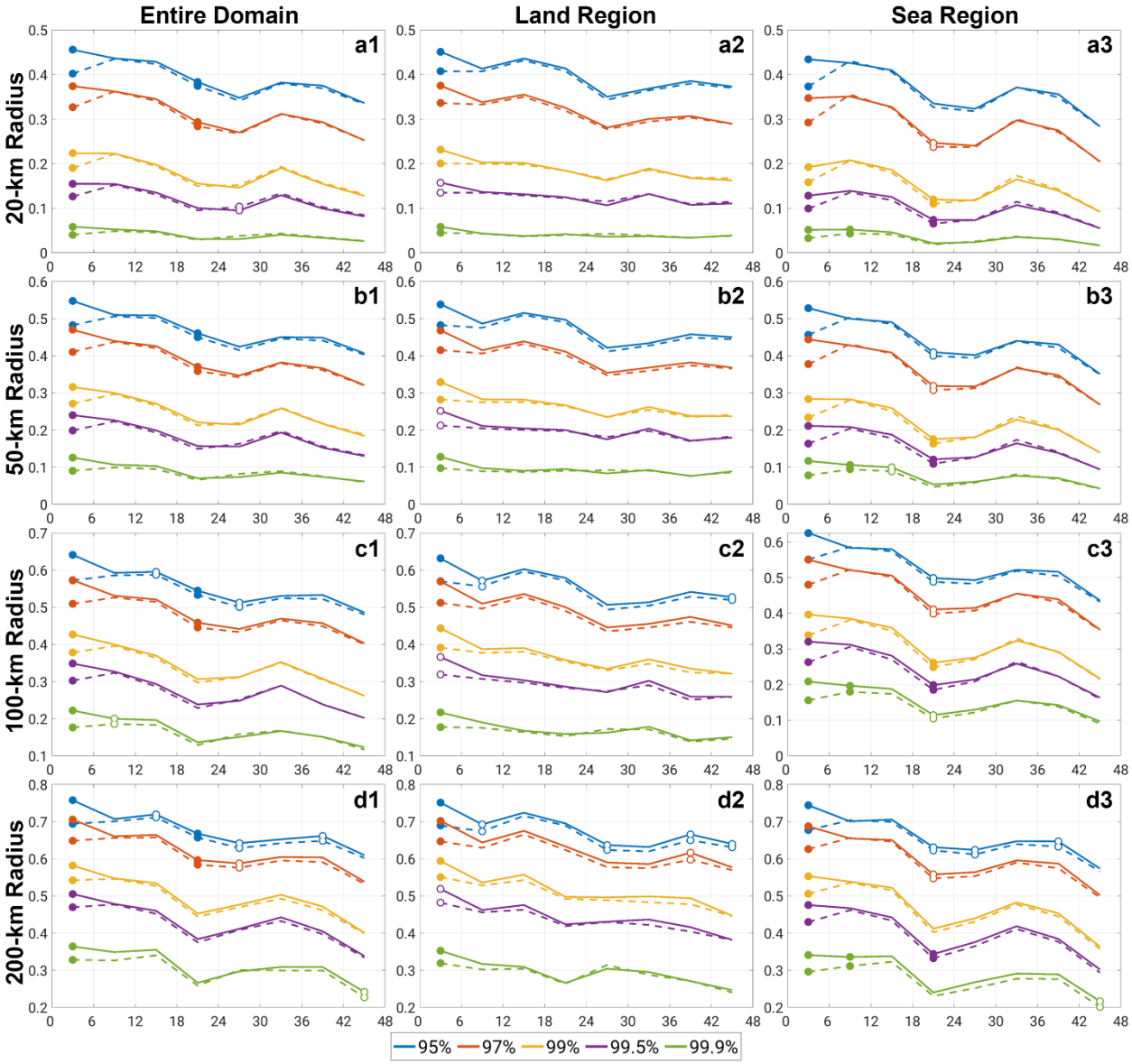

Figure 4*. FSS for the PRECIP2020 experiments with a neighborhood radius of (row a) 20 km, (row b) 50 km, (row c) 100 km, and (row d) 200 km over (first column) the entire domain, (second column) the land region, and (third column) the sea region. Solid and dashed lines represent scores of RT2020 and NoIR2020, respectively. Open circles and filled circles indicate that the differences between the two experiments' scores are statistically significant at the 95% and 99% confidence levels, respectively.*

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

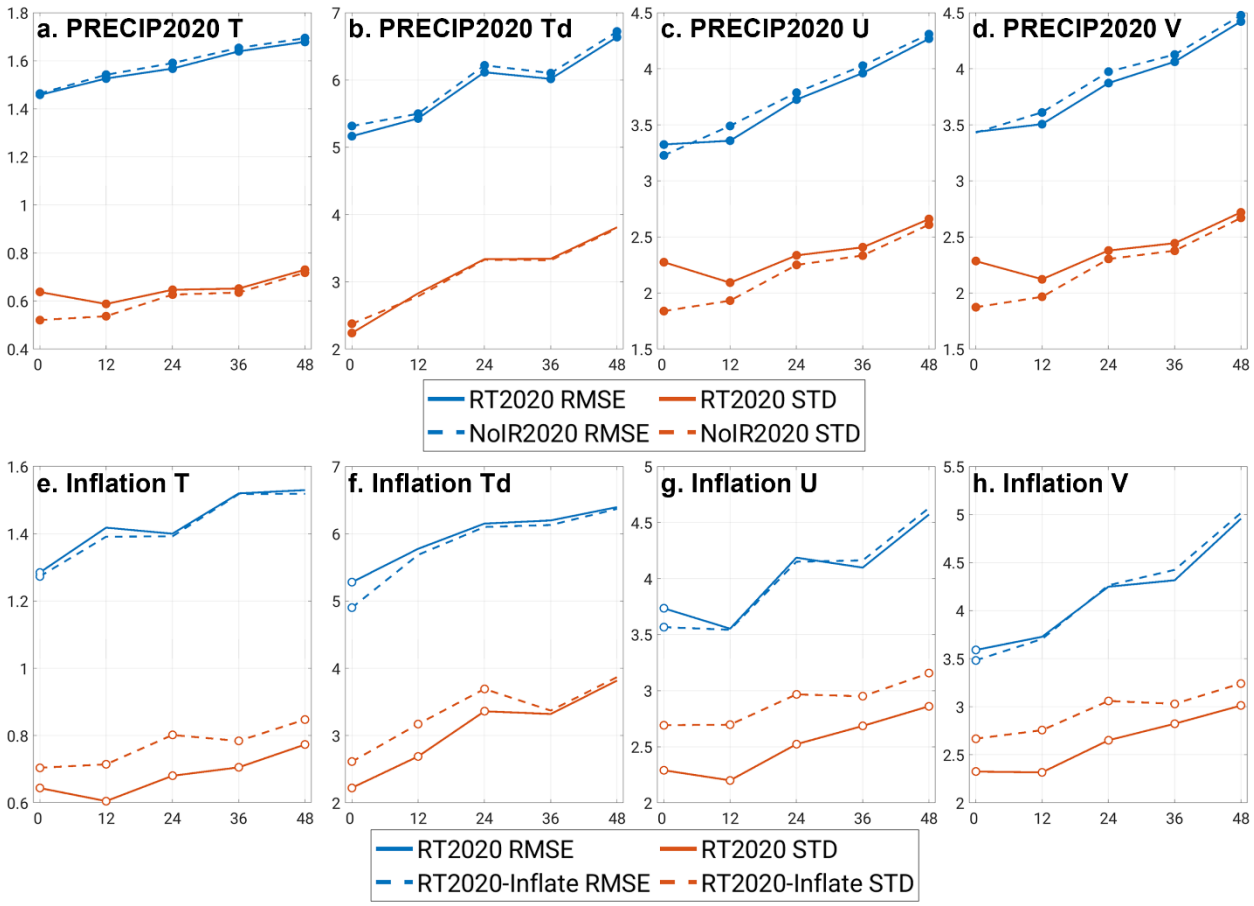

Figure 5*. Comparisons of RMSEs and STDs for (first column) T, (second column) Td, (third column) U, and (fourth column) V between (first row) the RT2020 and NoIR2020 forecasts, and (second row) the RT2020-Inflate and corresponding RT2020 forecasts that cover the same period. Open circles and filled circles indicate that the differences between the two experiments' scores are statistically significant at the 95% and 99% confidence levels, respectively.*

Besides improved rainfall forecasts, RT2020 forecasts also represent the environments and their uncertainties better than the NoIR2020 forecasts. Statistically significant reductions in RMSEs (at the 99% confidence level) occur for temperature and dew point (Fig. 5a, b) in the RT2020 forecasts compared with the NoIR2020 forecasts. For the wind fields (Fig. 5c, d), it is noted that the error at 0-hour forecast lead time (i.e., final EnKF analysis) is statistically significantly increased for U-wind in the RT2020 forecasts (Fig. 5c). Some recent studies also suggest that assimilating all-sky IR BTs might slightly degrade large-scale circulations in the analysis (Hartman et al. 2022). However, RT2020's RMSEs of wind fields become statistically

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

significantly lower than NoIR2020 beyond 12 hours (at the 99% confidence level). This improvement of the environments spanning the entire 48-hour forecast period also extends across the entire troposphere (Fig. 6a–d). Consistent with the overall average shown in Fig. 5, at 0-hour, the RT2020's RMSEs in temperature and V-wind component are generally mixed with improvements and degradations compared with NoIR2020's RMSEs, while for the U-wind component RT2020's RMSEs are generally higher than those of NoIR2020. However, RT2020's RMSEs become overwhelmingly lower than those of NoIR2020 beyond 12 hours across the entire troposphere with only occasional degradations. This suggests that although the improvements to rainfall forecasts resulting from assimilating all-sky IR BTs are generally limited to the first 6–12 hours, the improvements to the environments last longer.

RT2020 forecasts also provide statistically significant enhancement in ensemble STDs (at the 99% confidence level) compared with the NoIR2020 forecasts (Fig. 5a–d), except for the dew point which is only significant at 0-hour. The greater STDs in RT2020 forecasts also extend throughout the entire troposphere (Fig. 6e–h), except for a reduction in STDs at the upper troposphere for dew point (Fig. 6f). Although the algorithms of EnSRF make the ensemble spread smaller when more observations are assimilated, RT2020 forecasts have stronger and more active updrafts and downdrafts compared with NoIR2020 (figure not shown), and the enhanced deep convective activities in RT2020 forecasts can lead to their greater ensemble spread. When comparing STDs with RMSEs, STDs of both RT2020 and NoIR2020 forecasts are smaller than their corresponding RMSEs, suggesting that the ensemble forecasts are under-dispersive, which is an issue commonly observed in ensemble forecasts using a regional model (e.g., Schumacher and Clark 2014, Schwartz et al. 2014, Hagelin et al. 2017). Nonetheless, due to the smaller RMSEs and larger STDs in the RT2020 forecasts, it is less under-dispersive than the NoIR2020 forecasts. This indicates that the RT2020 forecasts better capture the uncertainty of the atmospheric states than the NoIR forecasts.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

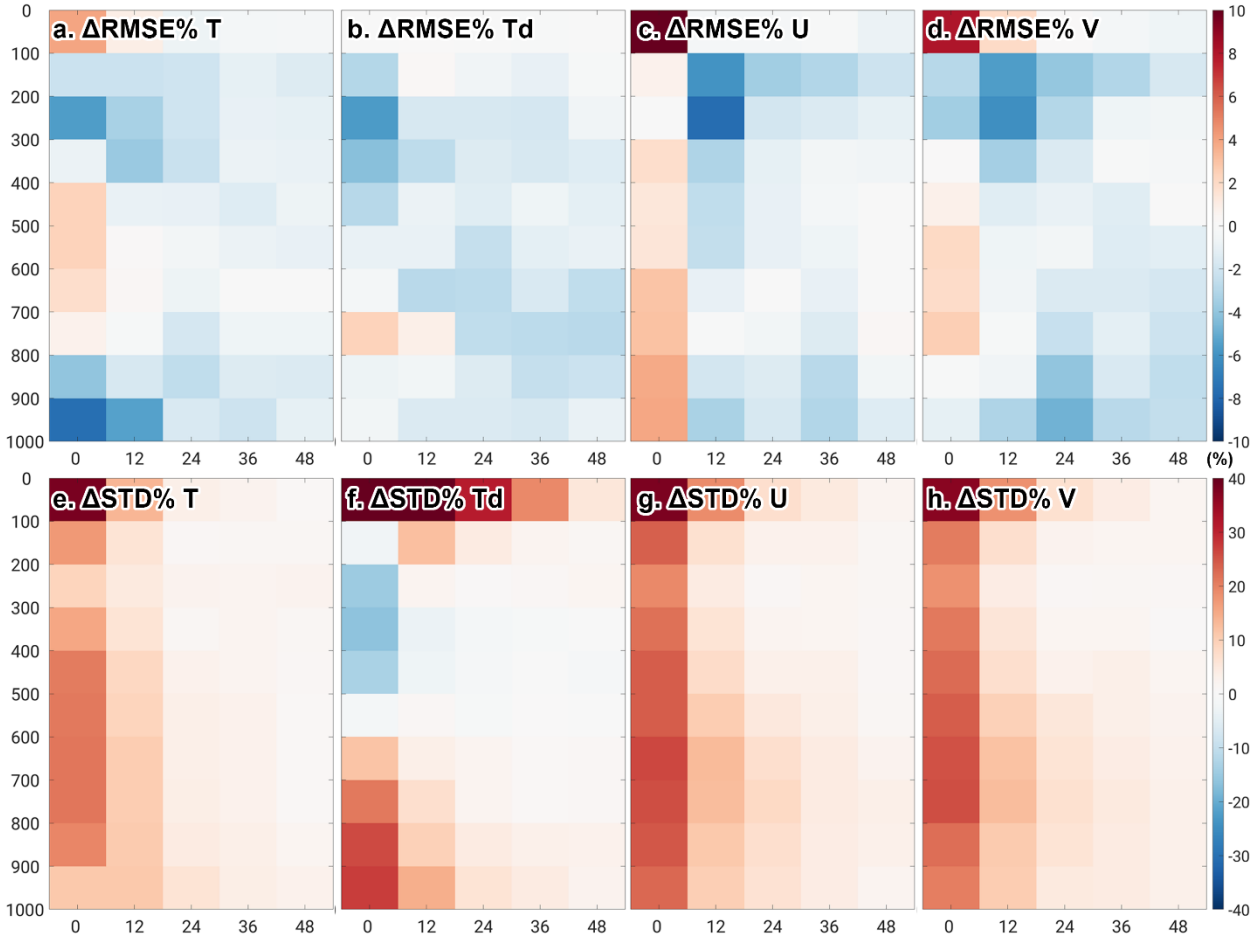

Figure 6. *Vertical distribution of relative changes (in percentage) of (first row) RMSEs and (second row) STDs of the RT2020 forecasts with respect to the NoIR2020 forecasts for (first column) T, (second column) Td, (third column) U, and (fourth column) V.*

As mentioned in Section 2, we added SKEP, SPPT, and multiplicative covariance inflation in the real-time forecasts of PRECIP2021 to combat the under-dispersive ensemble forecasts. Although RT2021TW forecasts show notably larger STDs than RT2020 forecasts (figure not shown), this is not a fair comparison because the synoptic activities in these two years are different from a climate perspective. Therefore, for 6 days of 4 July 2020 to 9 July 2020 when Kyushu, Japan, was hit by devastating heavy rainfall, we rerun the RT2020 forecasts with SKEB, SPPT, and multiplicative covariance inflation (referred to as "RT2020-Inflate" hereafter; only 9-km domain is used so that the results are consistent with the rest of this section). Figures 5e–h and 7 show the changes in RT2020-Inflate's RMSEs and STDs relative to those from RT2020 that cover the same period. It is apparent that applying these methods leads to overwhelming STD increases

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

throughout the entire 48-hour forecast periods (Fig. 5e–h) and this behavior persistently occurs for all 6 days of forecasts, except for occasional STD decreases beyond 36 hours in dew point temperature (Figs. 5f, 7f) and one forecast in V wind component (Fig. 7h). In the meantime, RMSEs of RT2020-Inflate are also slightly lower than those of RT2020 at 0-hour forecast (i.e., final EnKF analysis), although for forecast beyond 12 hours they present no difference (Figs. 5e–h, 7a–d). The persistent STD increases at all variables in almost all forecast days in this short-period sensitivity test suggest that these methods helped to maintain a more reasonable ensemble spread as expected. Given the formulation of EnKF that the analysis will be closer to the observations with greater background ensemble spread, the reduced RMSEs are also not surprising, even though the observations used for verifications are not assimilated.

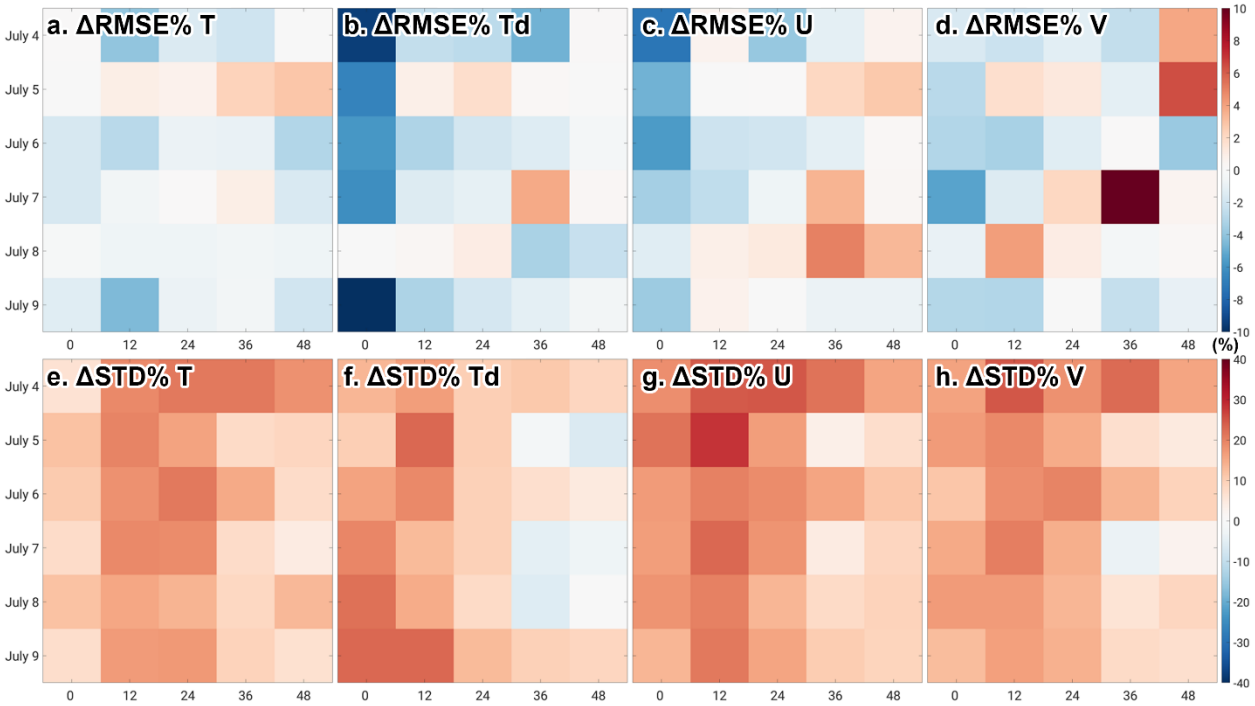

Figure 7*. Day-by-day relative changes (in percentage) of (first row) RMSEs and (second row) STDs of the RT2020-Inflate forecasts with respect to the RT2020 forecasts that cover the same period for (first column) T, (second column) Td, (third column) U, and (fourth column) V.*

5. "PRE"-CIP2021 Verifications

This section presents the evaluations of the forecasts during the "PRE"-CIP2021 experiments in Colorado, namely, RT2021CO and NoIR2021CO. All the results in this section are obtained

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

using the forecasts from the 3-km, convection-permitting inner domain. Since most improvements in rainfall forecast appear during the first 12 hours, we have also provided a different version of Figs. 8–11 with the first 12 hours stretched in the Supplementary Material as Figs. S1–S4.

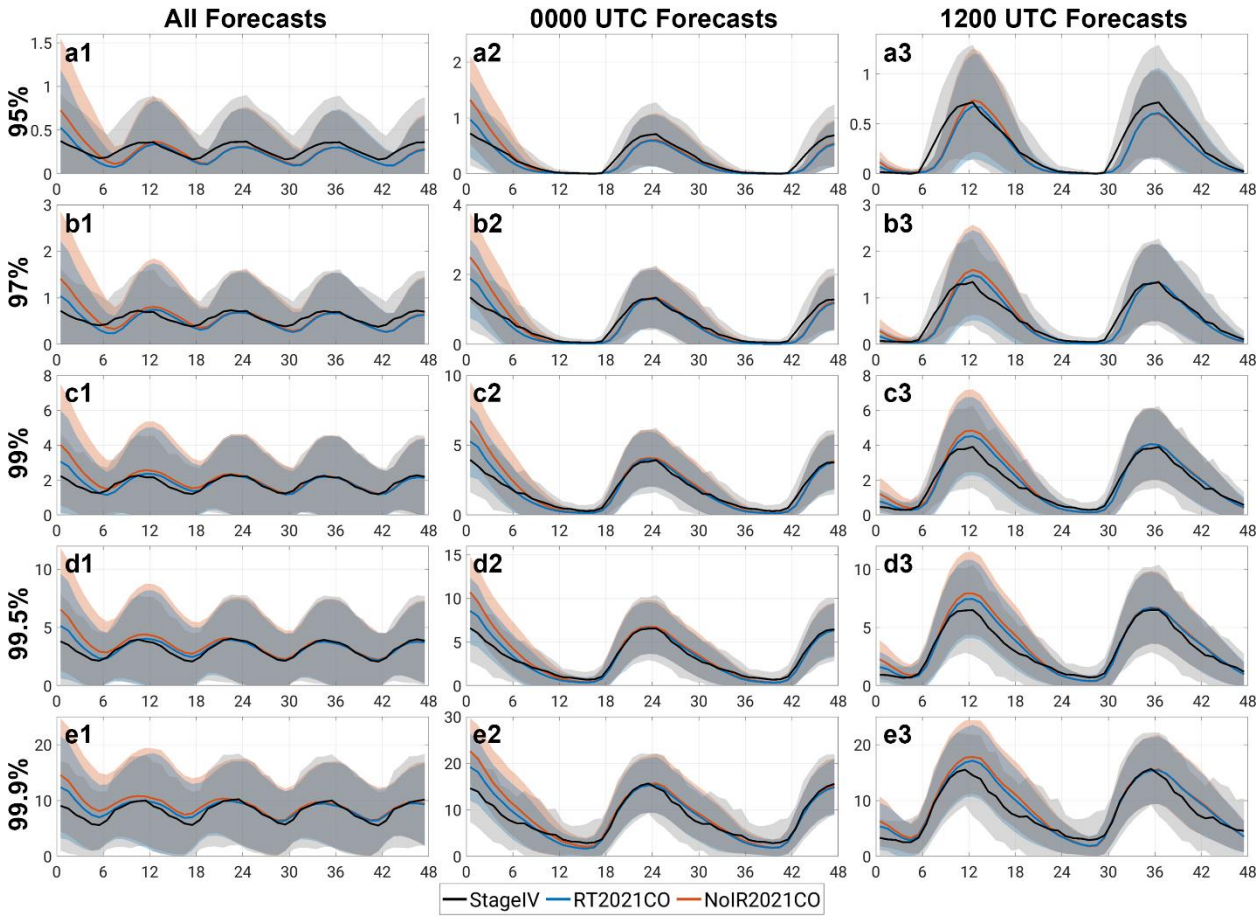

Figure 8. *Similar to Figure 2, but for the "PRE"-CIP2021 experiments containing (first column) all forecasts, (second column) only 0000 UTC forecasts, and (third column) only 1200 UTC forecasts.*

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

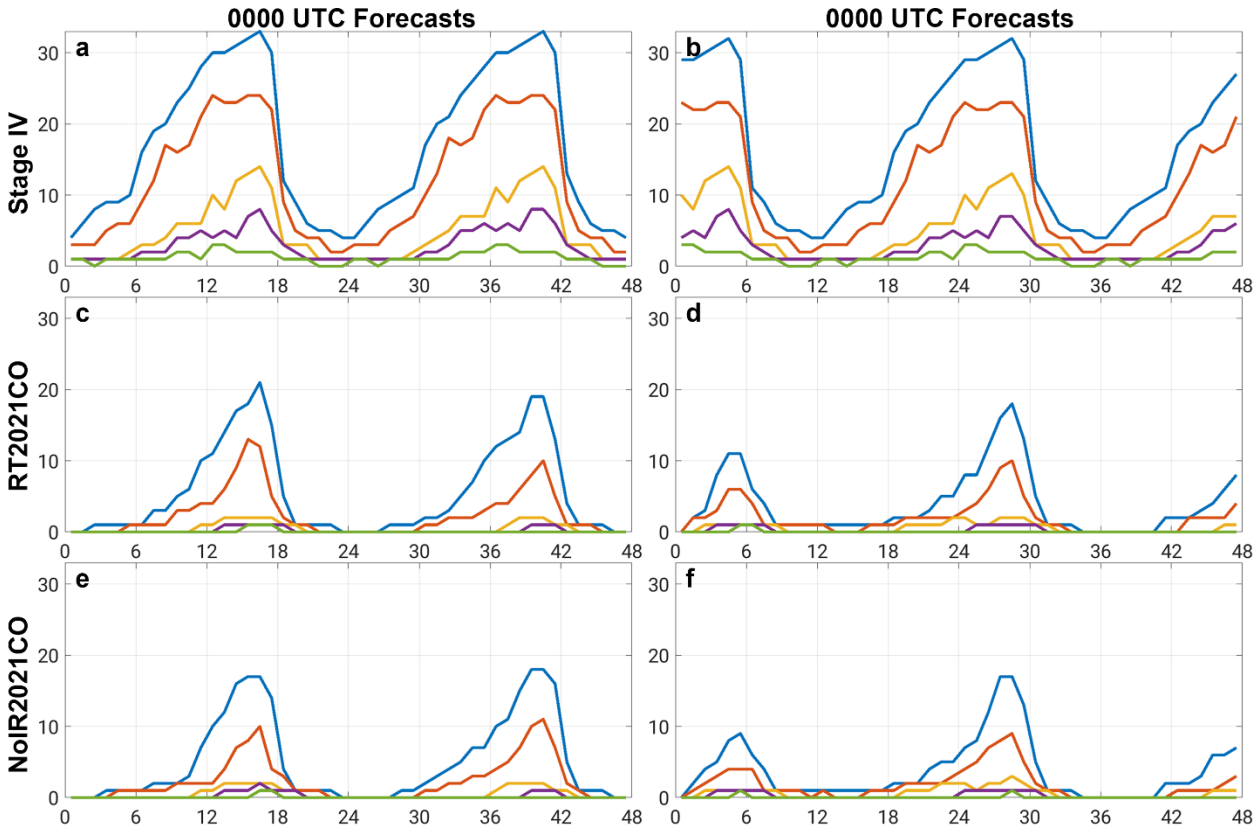

Figure 9*. Number of days that the physical value of a given percentile threshold equals zero for (first row) Stage IV estimates, (second row) RT2021CO forecasts, and (third row) NoIR2021CO containing (first column) only the 0000 UTC forecasts, and (second column) only the 1200 UTC forecasts.*

Figure 8 shows the distributions of the percentile thresholds' physical values for the "PRE"-CIP2021 experiments, as well as the values established separately for the 0000 UTC and 1200 UTC forecasts. Overall, the physical values of the forecasts are in good agreement with those from Stage IV. NoIR2021CO tends to produce more rainfall for the first 12 hours compared with Stage IV, but this overprediction is partly mitigated in RT2021CO. The onset of diurnal rainfall in RT2021CO and NoIR2021CO also tends to be slightly later (~ 1 hour) than that in Stage IV at the 95th and 97th percentiles (Fig. 8a, b), but this delay disappears at higher percentile thresholds. Figure 8 also shows that occasionally some percentile's mean values, or even the upper quartile values, become zero. This is because very little rainfall is observed or predicted in the 3 km domain, particularly in the late night to morning hours (12–18-hour and 36–42-hour lead times for the 0000 UTC forecasts, and 0–6-hour and 24–30-hour lead times for the 1200 UTC forecasts). Therefore,

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

we calculated the number of days that a certain percentile threshold becomes zero, and they are shown in Figure 9. Note that there are 33 forecasts for both the 0000 and 1200 UTC forecasts. Therefore, when the number of zero-value-threshold day for a given percentile at a given forecast lead-time reaches 33, there is no forecast within the entire "PRE"-CIP2021 period that this given percentile exceeds 0 mm at this forecast lead-time. We can see a steady increase in the number of zero-value-threshold days for all percentiles, starting from the early night and peaking around noon, both in the observations (Fig. 9a, b) and in the forecasts (Fig. 9c–f). The forecasts have fewer zero-value-threshold days than the observations. Nonetheless, if a certain percentile threshold equals zero at a certain forecast lead time either in the observations or in the forecasts, this forecast at this particular time must be excluded in ETS, AUC, and FSS calculations, because a physical value of zero no longer represents the corresponding percentile threshold. This will lead to a much smaller sample size – or even complete removal of all forecasts – at certain times.

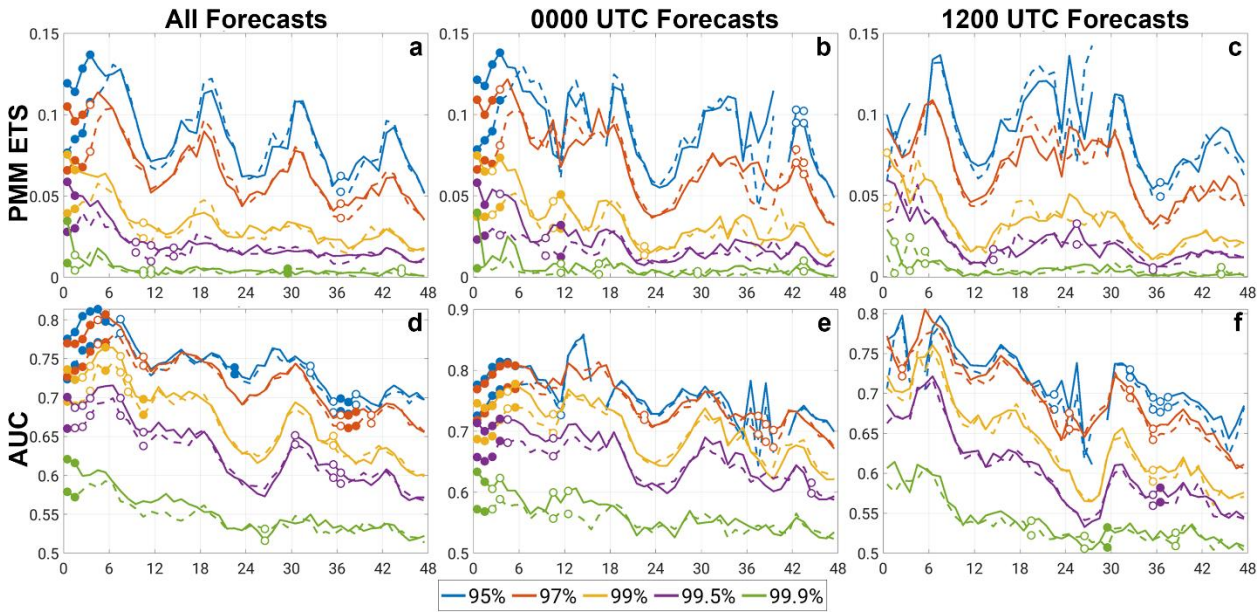

Figure 10. *Similar to Figure 3, but for the "PRE"-CIP2021 experiments containing (first column) all forecasts, (second column) only 0000 UTC forecasts, and (third column) only 1200 UTC forecasts.*

Figure 10 shows the ETSs and AUCs for all forecasts from the "PRE"-CIP2021 experiment, as well as calculated separately for the 0000 UTC and 1200 UTC forecasts using their percentiles. Consistent with evaluations in Section 4 for PRECIP2020, RT2021CO forecasts have higher ETS

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

than NoIR2021CO for the first 6–8 hours, with the first 2–4 hours achieving at least the 95% confidence level of statistical significance (Fig. 10a). If we examine only the 0000 UTC forecasts, there are consistent 0–4-hour statistically significant improvements in ETS except for the 99.9th percentile (Fig. 10b). For the 1200 UTC forecasts, the sample sizes at the beginning of the forecasts are too small (Fig. 9b), making the improvements in RT2021CO compared with NoIR2021CO in the first 2–3 hours unreliable (Fig. 10c). For AUC, RT2021CO forecasts are generally higher for the first 6–18 hours (Fig. 10d). If we examine only the 0000 UTC forecasts, RT2021CO's AUCs show statistically significant improvements at the 95%–99% confidence levels for the first 4–6 hours compared with NoIR2021CO's AUCs (Fig. 10e), suggesting that we can further improve the performance of RT2021CO forecasts with adequate calibrations and bias corrections.

The FSSs for the "PRE"-CIP2021 experiment in Figure 11 generally show characteristics that are consistent with ETSs and AUCs. When all forecasts are aggregated together, RT2021CO forecasts show significantly higher FSSs than NoIR2021CO forecasts for the first 3–6 hours (Fig. 11a1–d1), depending on the percentile and the spatial scale (neighborhood radius). If we only consider the 0000 UTC forecasts, RT2021CO forecasts significantly outperform NoIR2021CO forecasts for the first 4–6 hours, with the majority of the statistical significance tests exceeding the 99% confidence level (Fig. 11a2–d2). RT2021CO's 1200 UTC forecasts also show slightly higher FSSs for the first 2–3 hours than those in NoIR2021CO (Fig. 11a3–d3), but the sample size is too small to draw any conclusions.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

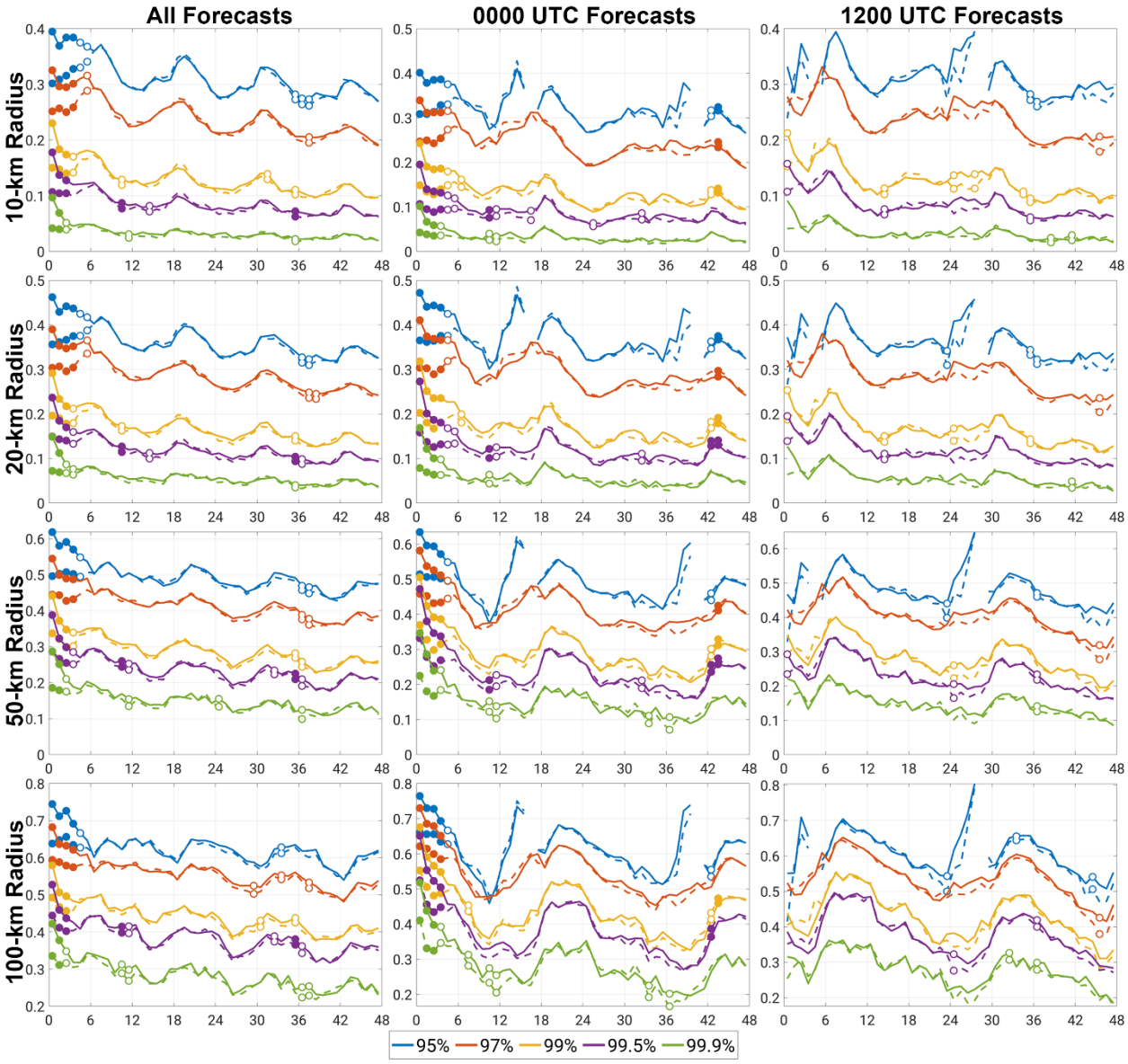

Figure 11. *FSS for the "PRE"-CIP2020 experiments with a neighborhood radius of (row a) 10 km, (row b) 20 km, (row c) 50 km, and (row d) 100 km for (first column) all forecasts, (second column) only 0000 UTC forecasts, and (third column) only 1200 UTC forecasts. Solid and dashed lines represent scores of RT2021CO and NoIR2021CO, respectively. Open circles and filled circles indicate that the differences between the two experiments' scores are statistically significant at the 95% and 99% confidence levels, respectively.*

The number of rawinsonde stations within the 3-km "PRE"-CIP2021 domain is much smaller than that within the 9-km PRECIP2020 domain (8 vs. 70–80). There are also very few observations

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

available below 900 hPa considering the topography of this domain, therefore RT2021CO and NoIR2021CO's differences in RMSEs and STDs below 900 hPa should be examined with caution. In general, RT2021CO's RMSEs are slightly lower than those of NoIR2021CO, although these differences are only occasionally statistically significant at the 95% or 99% confidence levels (Fig. 12). Temperature and dew point generally exhibit slightly reduced RMSEs in RT2021CO compared with NoIR2021CO across the whole troposphere (Fig. 13a, b), while for U and V wind components the RMSE changes are more mixed (Fig. 13c, d). On the other hand, we still see overall significantly increased STDs in RT2021CO compared with NoIR2021CO throughout the entire 48-hour forecast period (Fig. 12) and the entire troposphere (Fig. 13e–h), with the only exception being dew point at 0 hours (final EnKF analysis; Fig. 12b) primarily at the upper troposphere (Fig. 13f). This slight reduction of dew point STDs at the final EnKF analysis at the upper troposphere when all-sky IR BTs are assimilated is also observed in the verifications for PRECIP2020 (Fig. 6f) and the reason deserves further studies.

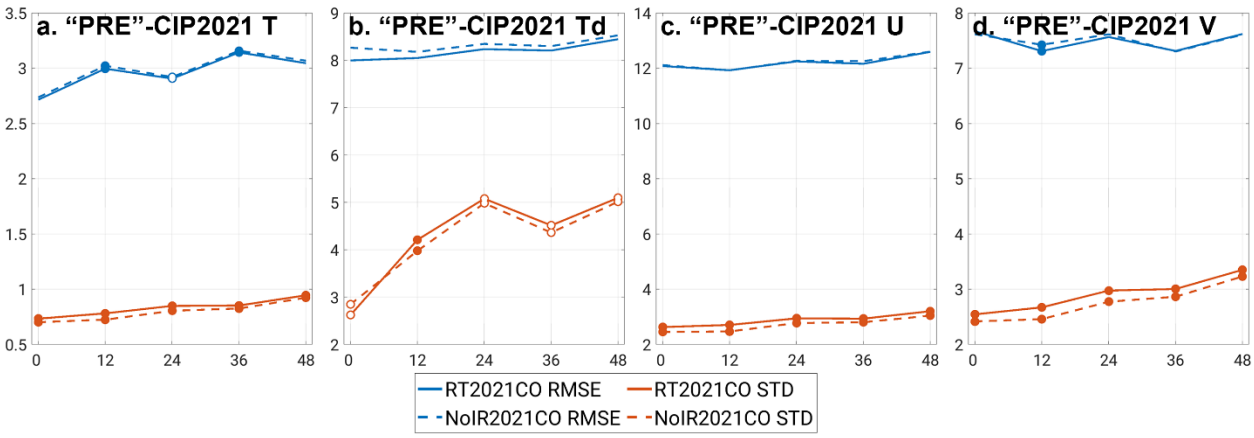

Figure 12. *Comparisons of RMSEs and STDs for (a) T, (b) Td, (c) U, and (d) V between the RT2021CO and NoIR2021CO forecasts. Open circles and filled circles indicate that the differences between the two experiments' scores are statistically significant at the 95% and 99% confidence levels, respectively.*

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

\

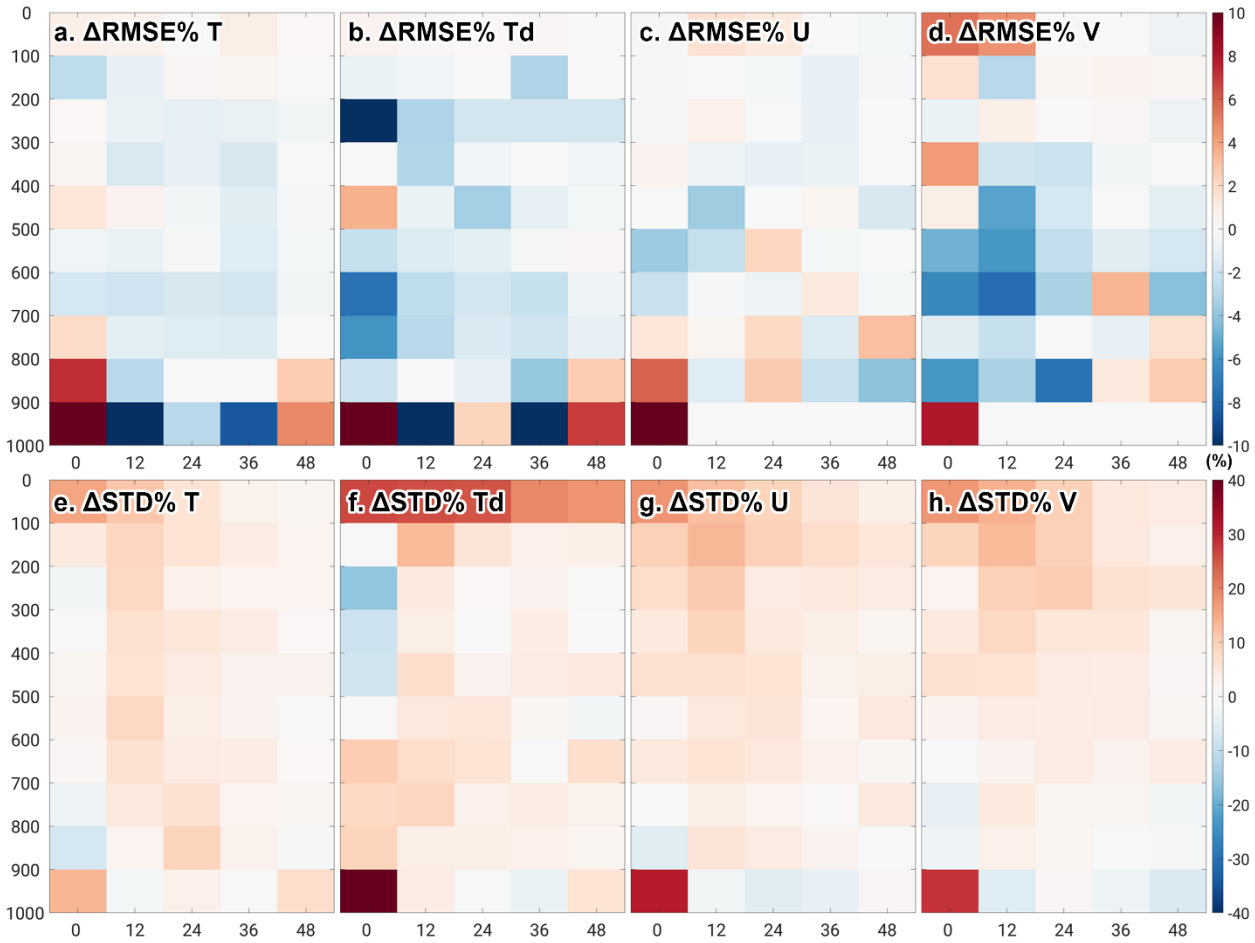

Figure 13. *Similar to Figure 6, but for the "PRE"-CIP2021 experiments.*

6. Concluding Remarks

During the summers of 2020 and 2021, the PSU WRF-EnKF data assimilation and forecast system was run in real-time in support of the Prediction of Rainfall Extremes Campaign In the Pacific (PRECIP), assimilating all-sky infrared radiances from the Himawari-8 geostationary satellite of Japan and the GOES-16 satellite of the United States, and provided ensemble forecasts every day for weather briefing and discussions of the field campaign. By comparing with retrospective forecasts excluding all-sky infrared radiances in the EnKF data assimilation cycles, this study presents the first systematic evaluation of the impact of assimilating all-sky infrared radiances for the quantitative precipitation forecasts (QPF) using real-time ensemble forecasts over several seasons at different regions.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Pointwise metrics of equitable threat score (ETS) and area under the receiver-operating-characteristic (ROC) curve (AUC) as well as neighborhood metric of fraction skill score (FSS) indicate that the rainfall forecasts are improved for at least 4–6 hours lead time when all-sky infrared radiances are assimilated for different rainfall percentile thresholds and spatial scales (neighborhood radii). The 2020 summer forecasts in East Asia and West Pacific also show 6–12-h QPF improvements for some thresholds. This timescale of 6 to 12 hours of rainfall forecast improvements is similar to what can be achieved with the assimilation of observations from Doppler weather radars (e.g., Xiao and Sun 2007; Aksoy et al. 2010; Clark 2012; Johnson et al. 2015; Surcel et al. 2015; Yussouf et al. 2016; Schwartz et al. 2021), and also likely governed by the limited predictability of convective-to-mesoscale systems (e.g., Zhang et al. 2016, 2022b). The large-scale environments are also improved when all-sky infrared radiances are assimilated, verified against available, independent rawinsonde observations that the RMSEs are generally reduced throughout the entire 48-hour forecast period. Furthermore, the ensemble forecasts from EnKF analyses assimilating all-sky infrared radiances contain larger ensemble spreads than when these observations are not assimilated, leading to a better balance between ensemble variance and error and a better representation of the uncertainty of the atmospheric states. The increased ensemble spread extends throughout the entire troposphere and lasts till the end of the 48-hour forecast period. SKEB, SPPT, and multiplicative covariance inflation also help to effectively increase and persistently maintain a more reasonable ensemble spread.

One notable difference when comparing the performance of the PSU WRF-EnKF system during the two years is the smaller and shorter-lasting improvements of assimilating all-sky IR BTs in CONUS (the "PRE"-CIP2021 experiments) when compared with the East Asia and West Pacific (the PRECIP2020 experiments) forecasts. The "PRE"-CIP2021 experiments captured primarily diurnally driven local convective storms as revealed by the apparent diurnal cycle in their percentile thresholds. These convective storms intrinsically have shorter predictability than convection embedded within large-scale Meiyu/Baiu fronts that were captured by the PRECIP2020 experiments. The dense surface observations in CONUS might also have contributed to the less impact of assimilating all-sky IR BTs in the "PRE"-CIP2021 experiments by preconditioning the environment better than the PRECIP2020. Additionally, we didn't assimilate other space-borne observations in our system's real-time forecasts, such as infrared sounder radiances, all-sky microwave radiances, and derived atmospheric motion vectors (AMVs).

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Although none of these observations – except for recently developed high-temporal-resolution AMVs (e.g., Zhao et al. 2021) – have the adequate temporal resolution to observe the fast-evolving convective phenomena, the impact of assimilating all-sky IR BTs from geostationary satellites on QPF might be diluted when other space-borne observations are also assimilated. Similar dilutions are also expected when ground-based Doppler weather radar observations are assimilated, although Zhang et al. (2019) suggests that all-sky IR BTs from geostationary satellites – owing to their earlier detection of convection initiation – can still bring in additional benefits compared with ground-based radars based on a case study. Observing system experiments (OSEs) are needed to isolate the influence of conventional observations and other types of remotely sensed observations when they are simultaneously assimilated with all-sky IR BTs from geostationary satellites, which is beyond the scope of this current study.

Lastly, there are still many remaining issues associated with all-sky IR BT assimilations, such as adequate observation error modeling, treatment of the correlated observation errors, proper covariance localization and simultaneous assimilation of multiple channels, multi-scale constraints of the environmental conditions, and the non-Gaussianity of the all-sky observations (e.g., Chan et al. 2020a). Additional studies are still warranted to make better, more efficient, and more effective use of all-sky IR BTs from geostationary satellites. However, considering the overall best spatiotemporal resolution and coverage for the monitoring and prediction of convectively driven rainfall systems compared with other remote sensing platforms, such as infrared and microwave radiances from low-Earth-orbiting satellites and ground-based and airborne Doppler weather radars, the multi-year multi-region evaluation of the real-time ensemble forecasts of the PSU WRF-EnKF system with the assimilation of all-sky IR BTs from the geostationary Himawari-8 and GOES-16 satellites presented here demonstrated the tremendous value of these observations in improving short-term QPF. Detailed evaluation of the PRECIP 2022 forecasts using the special field observations collected in East Asia will be reported in a subsequent study.

Acknowledgments

We would like to thank Drs. Rosimar Rios-Berrios (NCAR), James H. Ruppert, Jr. (University of Oklahoma), and many PRECIP participants for their helpful discussions and feedback on the configuration of the PSU WRF-EnKF system and the planning of the real-time modeling for the field campaign, and Dr. Kristen Rasmussen (Colorado State University) for securing the

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

computational resources for the 2022 real-time forecasts. We would also like to thank the three anonymous reviewers, whose comments prominently improved this manuscript. This work is supported by NSF grants AGS-1712290, AGS-1854607, and AGS-1854559, ONR grant N000141812517, NOAA NGGPS grant through University of Michigan Subcontract 3004628721, NOAA grant NA18OAR4590369, and NASA grant 80NSSC19K0728. The real-time and retrospective data assimilations and forecasts are performed on the Cheyenne supercomputer (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory (CISL) sponsored by the NSF, and the Stampede 2 supercomputer of the Texas Advanced Computing Center (TACC) through the Extreme Science and Engineering Discovery Environment (XSEDE) program (now the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support program, or ACCESS) supported by the National Science Foundation (NSF).

Data Availability Statement

The model output used in this study is subject to the PRECIP data policy, with a restricted release of one year after data generation and public release thereafter (upon request).

References

Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts, *Monthly Weather Review*, **138**, 1273–1292.

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, **127**, 2741–2758.

Benjamin, S. G., G. A. Grell, J. M. Brown, and T. G. Smirnova, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Monthly Weather Review*, **132**, 473–494.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **125**, 2887–2908.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Chan, M.-Y., J. L. Anderson, and X. Chen, 2020a: An efficient bi-Gaussian ensemble Kalman filter for satellite infrared radiance data assimilation. *Monthly Weather Review*, **148**, 5087–5104.

Chan, M.-Y., F. Zhang, X. Chen, and L. R. Leung, 2020b: Potential impacts of assimilating all-sky satellite infrared radiances on convection-permitting analysis and prediction of tropical convection, *Monthly Weather Review*, **148**, 3203–3224.

Chen, G., W. Sha, and T. Iwasaki, 2009: Diurnal variation of precipitation over southeastern China: Spatial distribution and its seasonality. *Journal of Geophysical Research*, **114**, D13103.

Chen, X., and F. Zhang, 2019: Relative roles of preconditioning moistening and global circumnavigating mode on the MJO convective initiation during DYNAMO. *Geophysical Research Letters*, **46**, 1079–1087.

Chen, X., O. M. Pauluis, and F. Zhang, 2018: Regional simulation of Indian summer monsoon intraseasonal oscillations at gray-zone resolution. *Atmospheric Chemistry and Physics*, **18**, 1003–1022.

Chen, X., F. Zhang, and K. Zhao, 2016: Diurnal variations of the land-sea breeze and its related precipitation over South China. *Journal of the Atmospheric Sciences*, **73**, 4793–4815.

Chen, X., L. R. Leung, Z. Feng, and F. Song, 2022: Crucial Role of Mesoscale Convective Systems in the Vertical Mass, Water, and Energy Transports of the South Asian Summer Monsoon, *Journal of Climate*, **35**, 91–108.

Chen, X., L. R. Leung, Z. Feng, and Q. Yang, 2022: Precipitation-moisture coupling over tropical oceans: Sequential roles of shallow, deep, and mesoscale convective systems. *Geophysical Research Letters*, **49**, e2022GL097836.

Chen, X., L. R. Leung, Z. Feng, F. Song, and Q. Yang, 2021: Mesoscale convective systems dominate the energetics of the South Asian summer monsoon onset. *Geophysical Research Letters*, **48**, e2021GL094873.

Clark, A. J., and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Monthly Weather Review*, **139**, 1410–1418.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program Spring Experiment, *Bulletin of the American Meteorological Society*, **93**, 55–74.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Clark, A. J., and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bulletin of the American Meteorological Society*, **99**, 1433–1448.

Dey, S. R. A., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Monthly Weather Review*, **142**, 4091–4107.

Du, Y., and G. Chen, 2019: Climatology of low-level jets and their impact on rainfall over southern China during the early-summer rainy season. *Journal of Climate*, **32**, 8813–8833.

Du, Y., and R. Rotunno, 2018: Diurnal cycle of rainfall and winds near the south coast of China. *Journal of the Atmospheric Science*, **75**, 2065–2082.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Monthly Weather Review*, **129**, 2461–2480.

Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **137**, 2024–2037.

Geer, A. J., and Coauthors, 2018: All-sky satellite assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1191–1217.

Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2015: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. *Weather and Forecasting*, **33**, 739–765,

Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble, MOGREPS-UK. *Quarterly Journal of the Royal Meteorological* Society, **143**, 2846–2861.

Han, Y, P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber, 2006: JCSDA Community Radiative Transfer Model (CRTM)—version 1. NOAA Tech. Rep. NESDIS 122, 40 pp.

Hartman, C. M., X. Chen, E. E. Clothiaux, and M.-Y. Chan, 2021: Improving the analysis and forecast of Hurricane Dorian (2019) with simultaneous assimilation of GOES-16 all-sky infrared brightness temperatures and tail Doppler radar radial velocities. *Monthly Weather Review*, **149**, 2193–2212.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Hartman, C. M., X. Chen, M.-Y. Chan, 2022: Improving tropical cyclogenesis forecasts of Hurricane Irma (2017) through the assimilation of all-sky infrared brightness temperature. *Monthly Weather Review*, in review.

He, J., and Coauthors, 2019: Development and evaluation of an ensemble-based data assimilation system for regional reanalysis over the Tibetan Plateau and surrounding regions. *Journal of Advances in Modeling Earth Systems*, **11**, 2503–2522.

Honda, T., and Coauthors, 2018a: Assimilating all-sky Himawari-8 infrared radiances: A case of Typhoon Soudelor (2015). *Monthly Weather Review*, **146**, 213–229.

Honda, T., S. Kotsuki, G.-Y. Lien, Y. Maejima, K. Okamoto, and T. Miyoshi, 2018b: Assimilation of Himawari-8 all-sky radiances every 10 minutes: Impact on precipitation and flood risk prediction. *Journal of Geophysical Research: Atmospheres*, **123**, 965–976.

Hong, S., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, *Monthly Weather Review*, **134**, 2318–2341.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129**, 123–137.

Huang, L., Y. Luo, and L. Bai, 2022: An evaluation of convection-permitting ensemble simulations of coastal nocturnal rainfall over South China during the early-summer rainy season. *Journal of Geophysical Research: Atmospheres*, **127**, e2021JD035656.

Huffman, G. J., and Coauthors, 2019: NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). NASA Algorithm Theoretical Basis Doc., version 06, 38 pp.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, **113**, D13103.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012. A revised scheme for the WRF surface layer formulation, *Monthly Weather Review*, **140**, 898–918.

Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Monthly Weather Review*, **143**, 3087–3108.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Jones, T. A., P. Skinner, N. Yussouf, K. Knopfmeier, A. Reinhart, X. Wang, K. Bedka, W. Smith, Jr., and R. Palikonda, 2020: Assimilation of GOES-16 radiances and retrievals into the Warn-on-Forecast system, *Monthly Weather Review*, **148**, 1829–1859.

Lee, J., E.-H. Lee, and K.-H. Seol, 2019: Validation of Integrated MultisatellitE Retrievals for GPM (IMERG) by using gauge-based analysis products of daily precipitation of East Asia. *Theoretical and Applied Climatology*, **137**, 2497–2512.

Lin, Y., and Mitchell K. E., 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. Preprints. *19th Conference on Hydrology*, San Diego, CA, American Meteorological Society, 1.2. https://ams.confex.com/ams/pdfpapers/83847.pdf.

Lu, X., Y. Wang, and Y. Qiu, 2022: A convection-permitting numerical study of diurnal cycles of pre-summer rainfall over southern China. *Quarterly Journal of the Royal Meteorological Society*, **148**, 3677–3693.

Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Weather and Forecasting*, **19**, 1106–1114.

Mason, P., and D. Thompson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. *Journal of Fluid Mechanics*, **242**, 51–78.

Minamide, M., and D. J. Posselt, 2021: Hurricane-seasonal analysis on the performances of convection-permitting ensemble tropical cyclone initializations with all-sky satellite radiance assimilation. *101st AMS Annual Meeting*, New Orleans, LA, American Meteorological Society, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/384263.

Minamide, M., and F. Zhang, 2017: Adaptive Observation Error Inflation for Assimilating All-Sky Satellite Radiance, *Monthly Weather Review*, **145**, 1063–1081.

Minamide, M., and F. Zhang, 2018: Assimilation of all-sky infrared radiances from Himawari-8 and impacts of moisture and hydrometer initialization on convection-permitting tropical cyclone prediction. *Monthly Weather Review*, **146**, 3241–3258.

Minamide, M., and F. Zhang, 2019: An adaptive background error inflation method for assimilating all-sky radiances. *Quarterly Journal of the Royal Meteorological Society*, **145**, 805–823.

Minamide, M., F. Zhang, and E. E. Clothiaux, 2020: Nonlinear forecast error growth of rapidly intensifying Hurricane Harvey (2017) examined through convection-permitting ensemble

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

assimilation of GOES-16 all-sky radiances. *Journal of the Atmospheric Sciences*, **77**, 4277–4296.

Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Boundary-Layer Meteorology*, **112**, 1–31.

Okamoto, K., Y. Sawada, and M. Kunii, 2019: Comparison of assimilating all-sky and clear-sky infrared radiances from Himawari-8 in a mesoscale system. *Quarterly Journal of the Royal Meteorological Society*, **145**, 745–766.

Otkin, J. A., and R. Potthast, 2019: Assimilation of all-sky SEVIRI infrared brightness temperatures in a regional-scale ensemble data assimilation system. *Monthly Weather Review*, **147**, 4481–4509.

Ou, T., D. Chen, X. Chen, C. Lin, K. Yang, H.-W. Lai, and F. Zhang, 2020: Simulation of summer precipitation diurnal cycles over the Tibetan Plateau at the gray-zone grid spacing for cumulus parameterization. *Climate Dynamics*, **54**, 3525–3539.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, *Monthly Weather Review*, **136**, 78–97.

Ruppert, J. H., Jr., and X. Chen, 2020: Island rainfall enhancement in the Maritime continent. *Geophysical Research Letters*, **47**, e2019GL086545.

Sawada, Y., K. Okamoto, M. Kunii, and T. Miyoshi, 2019: Assimilating every- 10- minute Himawari- 8 infrared radiances to improve convective predictability. *Journal of Geophysical Research: Atmospheres*, **124**, 2546–2561.

Schumacher, R. S., and A. J. Clark, 2014: Evaluation of ensemble configurations for the analysis and prediction of heavy-rain-producing mesoscale convective system. *Monthly Weather Review*, **142**, 4108–4138.

Schwartz, C. S., 2019: Medium-range convection-allowing ensemble forecasts with a variable-resolution global model. *Monthly Weather Review*, **147**, 2997–3023.

Schwartz, C. S., R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Monthly Weather Review*, **145**, 3397–3418.

Schwartz, C. S., G. S. Romine, and D. C. Dowell, 2021: Toward unifying short-term and next-day convection-allowing ensemble forecast systems with a continuously cycling 3-km ensemble

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Kalman filter over the entire conterminous United States. *Weather and Forecasting*, **36**, 379–405.

Schwartz, C. S., G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Weather and Forecasting*, **29**, 1295–1318.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR's real-time convection-allowing ensemble project. *Bulletin of the American Meteorological Society*, **100**, 321–343.

Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C- grid staggering. *Monthly Weather Review*, **140**, 3090–3105.

Skamarock, W. C., and Coauthors: A description of the Advanced Research WRF version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp.

Surcel, M., I. Zawadzki, and M. K. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *Journal of the Atmospheric Sciences*, **72**, 216–235.

Takaya, Y., I. Ishikawa, C. Kobayashi, H. Endo, and T. Ose, 2020: Enhanced Meiyu-Baiu rainfall in early summer 2020: Aftermath of the 2019 super IOD event. *Geophysical Research Letters*, **47**, e2020GL090671.

Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca, 2004: Implementation and verification of the unified NOAH land surface model in the WRF model. *20th conference on weather analysis and forecasting/16th conference on numerical weather prediction*, pp. 11–15.

Thompson, G., and T. Eidhammer, 2014: A Study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *Journal of the Atmospheric Sciences*, **71**, 3636–3658.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, **136**, 5095–5115.

Wang, S., A. H. Sobel, F. Zhang, Y.-Q. Sun, Y. Yue, and L. Zhou, 2015: Regional simulation of the October and November MJO events observed during the CINDY/DYNAMO field campaign at gray zone resolution. *Journal of Climate*, **28**, 2097–2119.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Weng, Y., and F. Zhang, 2012: Assimilating airborne Doppler radar observations with and ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Monthly Weather Review*, **140**, 841–859.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences* (3rd ed.). Elsevier.

Xiao, Q., and J. Sun, 2007: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP_2002, *Monthly Weather Review*, **135**, 3381–3404.

Xu, W., and E. J. Zipser, 2011: Diurnal variations of precipitation, deep convection, and lightning over and east of the eastern Tibetan Plateau. *Journal of Climate*, **24**, 448–465.

Yussouf, N., J. S. Kain, A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Weather and Forecasting*, **31**, 957–983.

Zhang, F., M. Minamide, and E. E. Clothiaux, 2016: Potential impacts of assimilating all-sky infrared satellite radiance from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophysical Research Letters*, **43**, 2954–2963.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter, *Monthly Weather Review*, **132**, 1238–1253.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review*, **137**, 2105–2125.

Zhang, F., M. Minamide, R. G. Nystrom, X. Chen, S.-J. Lin, and L. M. Harris, 2019: Improving Harvey forecasts with next-generation weather satellites: Advanced hurricane analysis and prediction with assimilation of GOES-R all-sky radiances. *Bulletin of the American Meteorological Society*, **100**, 1217–1222.

Zhang, Y., F. Zhang, and D. J. Stensrud, 2018: Assimilating all-sky infrared radiances from GOES-16 ABI using an ensemble Kalman filter for convection-allowing severe thunderstorms prediction. *Monthly Weather Review*, **146**, 3363–3381.

Zhang, Y., D. J. Stensrud, and F. Zhang, 2019: Simultaneous assimilation of radar and all-sky satellite infrared radiance observations for convection-allowing ensemble analysis and prediction of severe thunderstorms, *Monthly Weather Review*, **147**, 4389–4409.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.

Zhang, Y., D. J. Stensrud, and E. E. Clothiaux, 2021a: Benefits of the Advanced Baseline Imager (ABI) for ensemble-based analysis and prediction of severe thunderstorms. *Monthly Weather Review*, **149**, 313–332.

Zhang, Y., E. E. Clothiaux, and D. J. Stensrud, 2022a: Correlation structures between satellite all-sky infrared brightness temperatures and the atmospheric states at storm scales. *Advances in Atmospheric Sciences*, **39**, 714–732.

Zhang, Y., X. Chen, and Y. Lu, 2021b: Structure and dynamics of ensemble correlations for satellite all-sky observations in an FV3-based global-to-regional nested convection-permitting ensemble forecast of Hurricane Harvey. *Monthly Weather Review*, **149**, 2409–2430.

Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2016: Intrinsic predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma at storm scales. *Monthly Weather Review*, **144**, 1273–1298.

Zhang, Y., H. Yu, M. Zhang, Y. Yang, and Z. Meng, 2022b: Uncertainties and error growth in forecasting the record-breaking rainfall in Zhengzhou, Henan on 19–20 July 2021. *Science China Earth Science*, in press.

Zhang, Y., S. B. Sieron, Y. Lu, X. Chen, R. G. Nystrom, M. Minamide, M.-Y. Chan, C. M. Hartman, Z. Yao, J. H. Ruppert, Jr., A. Okazaki, S. J. Greybush, E. E. Clothiaux, and F. Zhang (2021c). Ensemble-based assimilation of satellite all-sky microwave radiances improves intensity and rainfall predictions of Hurricane Harvey (2017). *Geophysical Research Letters*, **48**, e2021GL096410.

Zhao, J., J. Gao, T. A. Jones, and J. Hu, 2021: Impact of assimilating high-resolution atmospheric motion vectors on convective scale short-term forecasts: 2. Assimilation experiments of GOES-16 satellite derived winds. *Journal of Advances in Modeling Earth Systems*, **13**, e2021MW002486.

Accepted for publication in *Weather and Forecasting*. DOI 10.1175/WAF-D-22-0156.1.