SPECIAL
RELAMPAGO-CACTI
COLLECTION

High-Resolution NWP Forecast Precipitation Comparison over Complex Terrain of the Sierras de Córdoba during RELAMPAGO-CACTI

GIMENA CASARETTO,[a,b] MARIA EUGENIA DILLON,[a] PAOLA SALIO,[b,c,d] YANINA GARCÍA SKABAR,[a,d] STEPHEN W. NESBITT,[e] RUSS S. SCHUMACHER,[f] CARLOS MARCELO GARCÍA,[g,h] AND CARLOS CATALINI[i,j]

[a] *Servicio Meteorológico Nacional and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina*
[b] *Departamento de Ciencias de la Atmósfera y los Océanos, FCEN, Universidad de Buenos Aires, Buenos Aires, Argentina*
[c] *Centro de Investigaciones del Mar y la Atmósfera (CIMA/CONICET-UBA), Buenos Aires, Argentina*
[d] *Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (UMI IFAECI/CNRS-CONICET-UBA), Buenos Aires, Argentina*
[e] *Department of Atmospheric Sciences, University of Illinois at Urbana–Champaign, Urbana, Illinois*
[f] *Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado*
[g] *Instituto de Estudios Avanzados en Ingeniería y Tecnología (IDIT CONICET/UNC), Córdoba, Argentina*
[h] *Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Católica de Córdoba, Córdoba, Argentina*
[i] *Instituto Nacional del Agua-Subgerencia Centro de la Región Semiárida (INA-CIRSA), Córdoba, Argentina*
[j] *Facultad de Ingeniería, Universidad Católica de Córdoba (UNC), Córdoba, Argentina*

ABSTRACT: Sierras de Córdoba (Argentina) is characterized by the occurrence of extreme precipitation events during the austral warm season. Heavy precipitation in the region has a large societal impact, causing flash floods. This motivates the forecast performance evaluation of 24-h accumulated precipitation and vertical profiles of atmospheric variables from different numerical weather prediction (NWP) models with the final aim of helping water management in the region. The NWP models evaluated include the Global Forecast System (GFS), which parameterizes convection, and convection-permitting simulations of the Weather Research and Forecasting (WRF) Model configured by three institutions: University of Illinois at Urbana–Champaign (UIUC), Colorado State University (CSU), and National Meteorological Service of Argentina (SMN). These models were verified with daily accumulated precipitation data from rain gauges and soundings during the RELAMPAGO-CACTI field campaign. Generally all configurations of the higher-resolution WRFs outperformed the lower-resolution GFS based on multiple metrics. Among the convection-permitting WRF Models, results varied with respect to rainfall threshold and forecast lead time, but the WRFUIUC mostly performed the best. However, elevation-dependent biases existed among the models that may impact the use of the data for different applications. There is a dry (moist) bias in lower (upper) pressure levels which is most pronounced in the GFS. For Córdoba an overestimation of the northern flow forecasted by the NWP configurations at lower levels was encountered. These results show the importance of convection-permitting forecasts in this region, which should be complementary to the coarser-resolution global model forecasts to help various users and decision-makers.

KEYWORDS: Forecast verification/skill; Forecasting; Numerical weather prediction/forecasting; Model comparison; Model evaluation/performance

1. Introduction

The Sierras de Córdoba (SDC; Fig. 1) is a complex terrain region in central Argentina characterized by the occurrence of extreme precipitation events (Zipser et al. 2006). Storms over this region are unique in terms of their vertical extent (Rasmussen et al. 2014) and known to produce high-impact weather events including large hail (Mezher et al. 2012; Kumjian et al. 2020) and heavy precipitation (Matsudo and Salio 2011; Rasmussen et al. 2014).

The prediction and diagnosis of accumulated precipitation by numerical weather prediction (NWP) models in this area

(SDC) is challenging due to its steep terrain (Yáñez-Morroni et al. 2018). The SDC are characterized by a transition from a mountainous semiarid region over mountain peaks to a large agricultural flat landscape eastward. Topography produces strong modifications of the synoptic-scale circulation, leading to high spatial variability in the low-level flow over mountains (Whiteman 2000). In addition, orography plays an important role in the distribution of precipitation over land and particularly, mountain shape impacts on precipitation location and amount (Robinson and Businger 2019). Therefore, accurately simulating the influence of complex terrain on the atmosphere is required to provide reliable NWP model forecasts. For example, Barthlott et al. (2006) showed that simulations with 2.8-km grid spacing represented terrain-forced secondary circulation systems better than simulations with 7-km grid

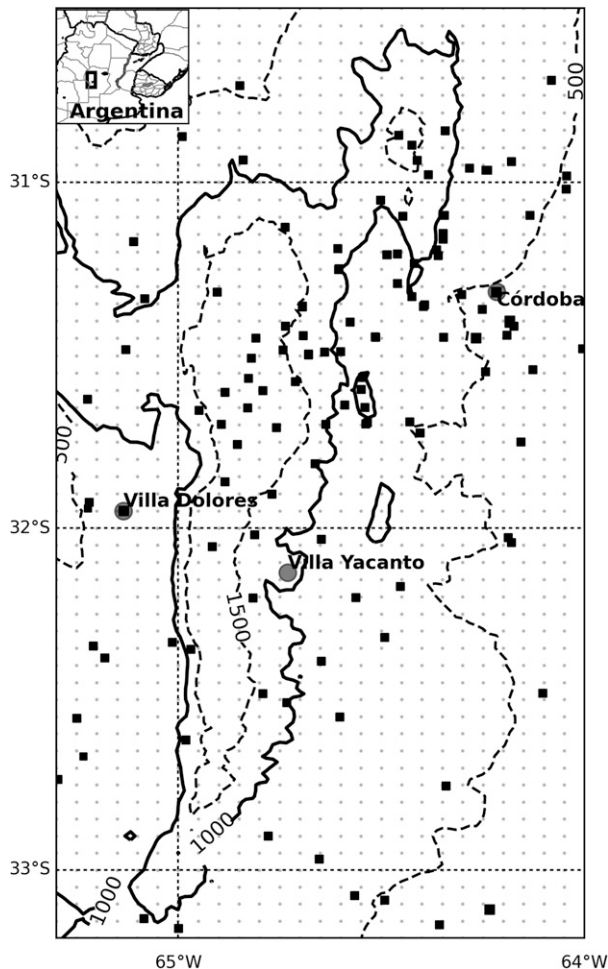*Corresponding author*: Gimena Casaretto, gcasaretto@smn.gob.ar

FIG. 1. Black boxes show rain gauge sites and gray circles identify the location of the sounding sites considered for evaluation (Córdoba, Villa Dolores, and Villa Yacanto). Contour lines represent the Sierras de Córdoba height MSL contours (dashed lines at 500 and 1500 m and solid line at 1000 m). The small gray dots represent the precipitation interpolation grid.

spacing, leading to horizontal mass convergence over the main mountain ridge prior to the formation of a thunderstorm over the Black Forest in Germany. They found that mesoscale circulation systems can be a factor controlling the initiation and development of convection above complex terrain under weak synoptic flows. Schwartz (2014) and Mahoney (2016) suggested that heavy rainfall forecasts may benefit from grid spacings finer than 4 km in complex terrain by analyzing the 2013 Colorado Flood. Nevertheless, Liu et al. (2011) and referenced papers suggested that high horizontal resolutions alone are not enough to obtain skillful simulations of topographic precipitation. Other factors such as resolving cloud structure, adequate treatments of cloud microphysics and subgrid processes in NWP models are also crucial.

Although there is a general consensus that rainfall forecast skill is significantly improved by using convection-permitting models rather than coarser-resolution and parameterized convection in NWP models (Clark et al. 2016), some studies have reported on the benefits of decreasing horizontal grid spacing beyond 4 km depending on the variables analyzed and the type of verification considered (Loken et al. 2017). For example, Kain et al. (2008) showed that 2-km horizontal grid spacing forecasts produced more convective detail, such as finer scale structure, but with no significant added value relative to the 4-km forecasts. Additionally, Potvin and Flora (2015) found that employing a 4-km resolution was too coarse to reliably resolve key supercell processes in idealized simulations.

Several studies have examined different NWP model resolutions and schemes, and their influence on the rainfall forecast skill over distinct complex terrain areas in the world during the warm season (e.g., McBride and Ebert 2000; Weisman et al. 2008; Zhu et al. 2018; Takemi 2018; Yáñez-Morroni et al. 2018). The results suggest significant added value in high-resolution convection-permitting forecasts of precipitation events. In Argentina, progress in warm season precipitation forecasting capabilities has been achieved through the implementation of a high-resolution NWP model over southern South America (García Skabar et al. 2018). Nevertheless, the ability of the NWP model to represent convection initiation over the SDC remains relatively unquantified. Mulholland et al. (2019) and Alvarez Imaz et al. (2021) documented the importance of terrain representation in high-resolution NWP models in order to improve high-impact weather forecasts over this region. Piersante et al. (2021) showed that models with parameterized convection have persistent biases in midlevel moisture in both North and South America, which also degrade precipitation forecasts. Therefore, considering that the heavy precipitating systems influenced by the SDC are not well represented (Ward et al. 2011), it is troublesome to achieve a good water resources management in the province of Córdoba.

The Remote sensing of Electrification, Lightning, And Mesoscale/microscale Processes with Adaptive Ground Observations (Nesbitt et al. 2021)–Clouds, Aerosols, and Complex Terrain Interactions (Varble et al. 2021) (RELAMPAGO-CACTI) field campaigns were conducted during the 2018/19 austral warm season, with an intense observing period (IOP) during November and December, with the aim of observing convective storms leeward of the Andes in the center of Argentina. Forecasting the location and timing of convective initiation, the convective mode and its further possible propagation was critical to the success of RELAMPAGO-CACTI. Mobile asset deployment was decided largely based on forecast briefings given twice a day at the operational center. The location of radiosonde launches, mobile radars, surface stations, as well as the radar scan strategies of fixed radars, *GOES-16* Advanced Baseline Imager 1-min mesoscale domain sectors, DOE Gulfstream-1 flights, as well as many other decisions during the field campaign were implemented considering the forecast team products. To support forecasting activities, different NWP models were run and used regularly by the forecast team at the operational center. Particularly, there were three different configurations of the convection-permitting Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008; Powers et al.

2017) runs carried out during the RELAMPAGO-CACTI project by the following institutions: National Meteorological Service of Argentina, University of Illinois at Urbana–Champaign, and Colorado State University. These high-resolution NWP model forecasts are the primary dataset analyzed in this study.

The motivation of this work lies in the existing need to evaluate high-resolution NWP models particularly over complex terrain. Precipitation prediction in the SDC region plays an important role in the decision making of the diverse water resources needs in the region. Water reservoir systems in the area originate on SDC, and provide drinking and irrigation water, flood control, the generation of electricity and recreational activities. Different basins over the SDC present torrential characteristics and a rapid response to rainfall events. They are also widely used for recreational purposes, which carries a risk associated with flash floods. The optimal management of their water resources is an important issue considering that this region suffers from a long dry period from May to September (Penalba and Vargas 2004). Note that the extensive plain area eastward of SDC is one of the most important agricultural and livestock regions in the country. There are significant floods in agricultural and urban areas, mainly caused by water accumulation due to heavy rainfall events (Matsudo and Salio 2011; Rasmussen et al. 2014) and little soil infiltration capacity. An adequate short-term rainfall forecast is a fundamental tool for a successful water management and to mitigate hydrological risk over both the mountainous regions and the plains.

The objective of the present paper is to evaluate the forecast performance of 24-h accumulated precipitation and atmospheric profiles of the different high-resolution convection-permitting (3 and 4 km) regional WRF simulations implemented during RELAMPAGO-CACTI IOPs, including a comparison with the coarser resolution Global Forecast System (GFS) forecasts, which has a horizontal grid spacing of approximately 13 km. The accuracy of the forecasts over the complex terrain of the SDC is a particular focus. The forecasts were evaluated using the available surface rain gauge stations interpolated to 0.05° and relevant RELAMPAGO-CACTI atmospheric soundings as a reference.

This article is structured as follows: section 2 describes the main characteristics of the available surface station network data and the configuration of the different numerical weather prediction models, as well as the methodology carried out to evaluate their 24-h accumulated precipitation forecasts and their atmospheric profiles. Section 3 presents the validation results for the entire region, as well as an analysis with respect to terrain type, i.e., differentiating between orography and plain areas. This section also discusses the validation of the different atmospheric profiles. Finally, in section 4, the conclusions and final considerations are presented.

2. Data and methodology

a. Surface stations and sounding sites

A network of 125 surface rainfall stations from different institutions within the RELAMPAGO-CACTI area of study

is considered (Fig. 1). The study period is established based on the RELAMPAGO-CACTI IOP and the common period among the forecasts evaluated: from 3 November to 16 December 2018 (43 days). The surface stations belong to seven different institutions: the National Hydrological Service of Córdoba (50 stations), the National Institute of Agricultural Technology of Argentina (2 stations), the National Meteorological Service of Argentina (SMN, 6 stations), the National System of Meteorological Radars of Argentina (5 stations), the Meteorological Stations Network of San Luis (11 stations), and Córdoba Board of Trade and the Ministry of Production of Córdoba (51 stations). They are all automatic stations with tipping-bucket rain gauges, except the ones from SMN that are conventional meteorological pluviometers. The 24-h accumulated precipitation is considered because only a small portion of the available rain gauges stores the data hourly. Due to the existence of different measuring strategies among the institutions considered, we define 24-h accumulated precipitation as the accumulated precipitation at 1200 UTC of each day, considering the previous 24 h.

Given that the databases are provided by different institutions, the following quality control was done. First, only stations with 24-h precipitation data were considered. Periods of missing data (i.e., sensor downtime) and no-precipitation were identified separately. All stations with 24-h precipitation equal to zero due to a possible transmission error were evaluated initially by checking cloud cover at each site using satellite information. Only clear skies situations were included in the sample while cloudy skies were considered as a transmission error. Although some stations did not have the complete data series, missing data represents only 5% of each station during the RELAMPAGO-CACTI period. With the aim of evaluating the precipitation forecasts with the largest number of observations possible, none of the stations were rejected. To unify the irregularly spatially distributed precipitation data and achieve a high-resolution grid of accumulated 24-h precipitation observations, linear interpolation is used to take the surface station data to a regular grid with a 0.05° spatial resolution, covering the area enclosed by 33.2°–30.5°S and 65.3°–64°W. Due to the distribution of the available observations, when the smallest convex set that contains the stations is considered, a missing value is taken for the points that are outside of the convex hull of the input points but inside the area of study (see white regions in Fig. 4; Virtanen et al. 2020). The 0.05° spatial resolution was chosen for being lower than the WRF Model configurations. Nevertheless, it still provides the advantages of high resolution. Note, however, that this resolution is finer than the GFS output grid. It is worth mentioning that nearest neighbor and cubic interpolations were tested for the regular grid construction, but the one that best matched the available surface observations and that best represented the topographic features, i.e., heavy rain, was the linear interpolation.

During the RELAMPAGO-CACTI IOP there were multiple mobile and fixed sounding sites (Schumacher et al. 2021) within 33.2°–30.5°S and 65.3°–64°W with launches at 1200 UTC throughout the campaign. The locations sites with topographic
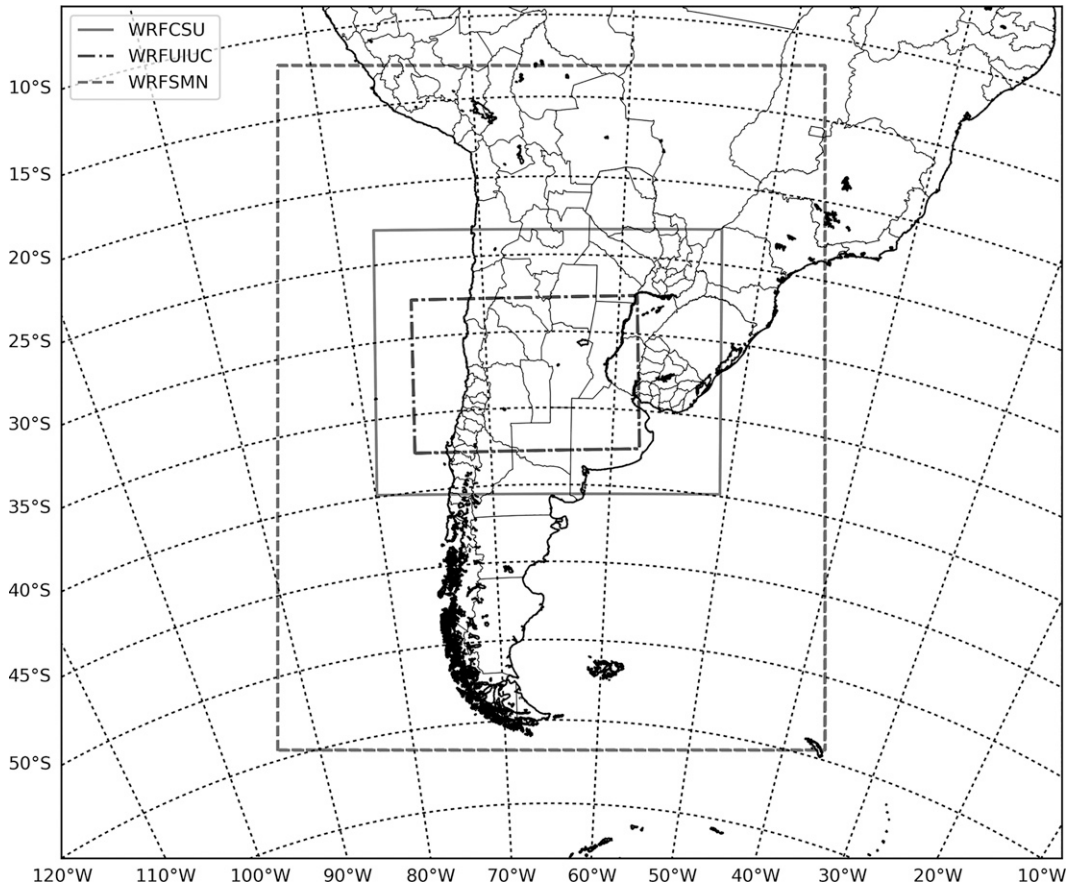
FIG. 2. Domain boundaries of the different WRF Model configurations.

features used for the atmospheric profile evaluation are displayed in Fig. 1, and include Villa Dolores (located westward of the SDC, 584 m MSL), Villa Yacanto (located eastward of highest peak of SDC, 1161 m MSL), and Córdoba (located eastward of the SDC, 490 m MSL). We focus our analysis on temperature (T), dewpoint temperature (Td), and wind components (u and v) at various pressure levels and forecast times to evaluate possible forecast biases.

b. Numerical weather prediction

1) GLOBAL FORECAST SYSTEM

The GFS is a global NWP model developed by the National Oceanographic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP). During the period of interest, the GFS version 14 with parameterized convection was available with a horizontal resolution of 0.25°. The initial conditions are obtained through the Global Data Assimilation System, which uses a hybrid four-dimensional ensemble variational formulation (Buehner et al. 2013). The GFS forecasts are run for 384 h in four different cycles each day, initialized at 0000, 0600, 1200, and 1800 UTC, with hourly data outputs. For a complete description of this version of the model, readers are

referred to https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php.

The forecasts from the 0000 and 1200 UTC cycles are considered in the analysis presented below.

2) WEATHER RESEARCH AND FORECASTING MODEL

The WRF Model is an open source state-of-the-art atmospheric modeling system, designed for both research and operational NWP (Skamarock et al. 2008; Powers et al. 2017). WRF offers a large variety of parameterizations for atmospheric processes of diverse temporal and spatial scales. This study uses the Advanced Research version of WRF (ARW) dynamic core developed and maintained by the National Center of Atmosphere Research (NCAR) Mesoscale and Microscale Meteorology Laboratory.

As discussed in the introduction, three different WRF-ARW configurations were used for forecast operations during the RELAMPAGO-CACTI IOP field campaign. These versions were implemented by the National Meteorological Service of Argentina (WRFSMN), the University of Illinois at Urbana–Champaign (WRFUIUC), and the Colorado State University (WRFCSU).

These WRF configurations have different simulation domains (see Fig. 2), but their boundary conditions were all obtained

TABLE 1. Characteristics and parameterization of each WRF Model used. Detailed explanation in text (section 2).

Characteristic	WRF UIUC	WRF CSU	WRF SMN
WRF version	4	4	3.8
Horizontal resolution (km)	3	4	4
Vertical levels	51 (defined by user)	51 (defined by user)	38 (by default)
Model top	20 hPa	50 hPa	50 hPa
Geographic data resolution	MODIS 30 s	USGS lakes 30 s	30 s
Boundary and initial condition	GFS 0.25°, 3-h frequency	GFS 0.25°, 3-h frequency	GFS 0.25°, 1-h frequency
Projection	Lambert (centered in 32.79°S, 67°W)	Lambert (centered in 32°S, 65.25°W)	Lambert (centered in 35°S, 60°W)
Parameterizations			
Microphysics	Thompson (two-moment)	Morrison (two-moment)	WSM6 (one-moment)
Boundary layer	YSU	MYJS	MYJS
Radiation	RRTMG	RRTMG	RRTM
Surface	Noah	Noah	Noah

from GFS. The main characteristics and parameterizations of these NWP WRF configurations are shown in Table 1.

The orography represented by each regional NWP model is slightly different (see Fig. 3). Although the three model configurations use the same geographic data resolution of 30 arc s (approximately 0.927 km), different Lambert map projection parameters are used which, in addition with the horizontal resolution of each model configuration (3 or 4 km), lead to different grid points. It is also worth mentioning that the WRFUIUC and WRFCSU have a hybrid coordinate system option that was added in WRF v.4 with levels defined by user, which is not included in the WRFSMN configuration. A stretched vertical grid was used in WRFUIUC and WRFCSU with higher resolution in the atmospheric boundary layer to better represent terrain flows and convective cold pools in the lower atmosphere. The default sigma coordinates are used in the WRFSMN configuration.

The different parameterizations employed by each WRF-ARW configuration (Table 1) are briefly described here. The three WRF configurations were executed in a convection-permitting mode (i.e., no convective parameterization). Note that as described above, GFS uses a convective parameterization.
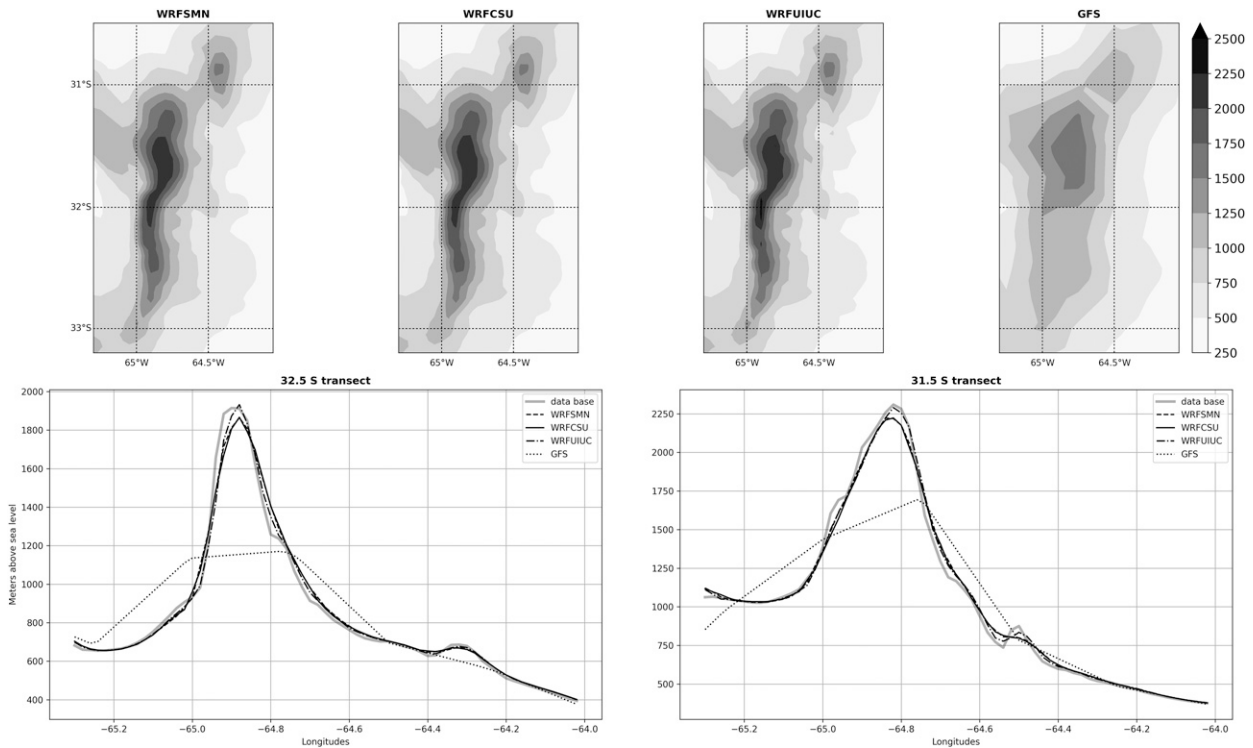

FIG. 3. (top) Contours of terrain elevation represented from 100 m MSL for each model used and (bottom) latitudinal transects at 32.5° and 31.5°S together with the database considered as the reference topography.

The microphysics schemes used in the different WRF configurations are Thompson (Thompson et al. 2008), WRF single-moment 6-class (WSM6; Hong et al. 2006), and Morrison double-moment scheme (MDMS; Morrison et al. 2009). Thompson and MDMS are double-moment schemes for ice, snow, rain, and graupel. WRFSMN uses WSM6, which employs a single moment representation for all species.

Regarding the radiation parameterizations used, the Rapid Radiative Transfer Model (RRTM) (Mlawer et al. 1997) scheme has calculations for multiple bands and microphysics species, while the RRTMG (Iacono et al. 2008) is the new version of RRTM that includes updates considering the overlap of clouds.

The Noah land surface model (Niu et al. 2011) is used in all three model configurations. It is a development by NCEP/NCAR/Air Force Weather Agency with a scheme of temperature and humidity of four soil layers.

Last, the boundary layer parameterizations used include the Mellor–Yamada–Janjić scheme (MYJS; Janjić 1994), containing a one-dimensional prognostic turbulent kinetic energy scheme with local vertical mixing, and the Yonsei University Scheme (YSU; Hong et al. 2006), which considers counter-gradient fluxes and generally represents deep planetary boundary layer circulation more accurately than local schemes.

These model configurations and schemes were determined by the different institutions (i.e., SMN, UIUC, CSU) considering their high performance computer capabilities to run operationally during RELAMPAGO-CACTI IOPs.

c. Period and area of study

NWP model verification over the complex terrain of the SDC is carried out over a limited study area. Taking into account the domains of all the NWP models and the availability of the surface network observations, the study area is defined as a box that contains the most important river basins of the Córdoba province. This is crucial to aid decision making regarding the diverse water resources needs in the region (including the river basins of Anizacate, Cruz del Eje, La Cañada, La Viña, Los Molinos, Piedras Moras, Río Cuarto, Saldan, and San Roque). The latitudinal and longitudinal limits of this bounding box are 33.2°, 30.5°S and 65.3°, 64°W, respectively (see Fig. 1).

As mentioned above, 24-h accumulated precipitation forecasts at 1200 UTC are evaluated for a common validation period of 43 days (from 3 November to 16 December). Accordingly, for the 1200 UTC NWP model initializations, the accumulated precipitation forecasts were obtained from both the integration of the first 24 h and between 24- and 48-h lead times, hereafter referred to as "F 24-h" and "F 48-h," respectively. Regarding the evaluation for the 0000 UTC cycle initialization, the accumulated precipitation forecasts were obtained from the integration between 12- and 36-h lead times, hereafter referred to as "F 36-h." Note that in order to maintain the same validation period among the forecasts, the initializations from 2 (1) November to 15 (14) December were taken for F 24-h (F 36-h and F 48-h). The NWP model forecasts were interpolated linearly to a regular grid with a 0.05° spatial resolution like the observations, as described in section 2a, to facilitate their comparison.

TABLE 2. Contingency table for calculating categorical statistics.

Precipitation (mm day^{-1})	Observed $\geq$ threshold	Observed $<$ threshold
Forecast $\geq$ threshold	H	F
Forecast $<$ threshold	M	CN

The same forecast lead times are considered for atmospheric profile evaluation at different sounding sites (F 24-h and F 48-h for 1200 UTC initialization; F 36-h for 0000 UTC initialization). The nearest grid point of each NWP model to the location of the sounding site was used for the validation. The sounding levels, of high vertical resolution (5 hPa), were interpolated vertically to the NWP model vertical resolutions.

d. Statistical methods for evaluation

The 24-h accumulated precipitation forecasts were evaluated considering the interpolated observation fields as a reference. In this study, different strategies are used for verification and are described below.

The 24-h accumulated precipitation values greater than 0.1 mm day^{-1} are defined as a precipitation event. The percentiles for each grid point and categorical statistics that measure the correspondence of the occurrence (or lack of occurrence) of the precipitation event were calculated. A contingency table was constructed taking into account different precipitation thresholds (0.1, 1, 3, 5, 10, 15, 20, 25, 30, 40, and 50 mm day^{-1}). For each precipitation event when the forecast is correct, either a hit (H) or a correct rejection (CN) is obtained, while when the forecast is incorrect, either a false alarm (F) or a miss (M) is obtained (Table 2). A hit represents the number of cases (grid points) in which the accumulated precipitation values that were observed and predicted are above or equal to the threshold value of interest, while a miss represents the cases in which the observed value was above or equal to the threshold, but the predicted value was below it. False alarms are recorded when the accumulated precipitation predicted was above or equal to the threshold of interest but observed below it and, finally, correct rejection corresponds to when both the accumulated precipitation observed and predicted were below the threshold of interest. From this contingency table, different indices are calculated to evaluate the performance of the models.

Different categorical statistics are used for forecast evaluation.

- The bias score (BIASS) is the ratio of the relative frequency of forecasted events and observed events. Its value can vary between zero and infinity; its optimal value is equal to one. BIASS values above one (below one) indicate that the forecast system exhibits overestimation (underestimation) of the frequency of events [Eq. (1)]:

$$\text{BIASS} = \frac{H + F}{H + M}. \tag{1}$$

- The equitable threat score (ETS) is the correctly predicted precipitation fraction considering a number of random hits

(H_{random}). Its value can vary between -0.33 and 1, where 1 indicates a perfect forecast [Eq. (2)]:

$$\text{ETS} = \frac{H - H_{\text{random}}}{H + M + F - H_{\text{random}}};$$
$$H_{\text{random}} = \frac{(H + M)(H + F)}{\text{total}}. \quad (2)$$

- The probability of detection (POD) measures the proportion of the observed events that were correctly forecasted over the total population, where a perfect score equals one [Eq. (3)]:

$$\text{POD} = \frac{H}{H + M}. \quad (3)$$

- The false alarm ratio (FAR) provides information on the relationship between the number of false alarm cases and the number of predicted events. It ranges between zero and one, where the optimal value is equal to zero [Eq. (4)]:

$$\text{FAR} = \frac{F}{H + F}. \quad (4)$$

The fractions skill score (FSS) was also calculated. The FSS is a spatial verification measurement used for assessing the performance of precipitation forecasts from NWP models (Roberts and Lean 2008). Previous studies have shown that the FSS is able to give a direct measure of the error in the placement of precipitation as it considers different domain sizes for the verification (Sobash et al. 2011; Baldauf et al. 2011, among others). To evaluate the FSS the thresholds used are the same as those described above for the categorical statistics. A binary grid was created in which all the grid squares that exceed the threshold of interest are assigned a value of 1 and the others a value of 0. For each square neighborhood (1×1, 5×5, 9×9), the corresponding fraction in Eq. (5) is calculated for each grid point. FSS values are obtained for each threshold and each neighborhood gridpoint size; the perfect score is one. Following (Skok and Roberts 2016), the FSS minimum value for a "useful" forecast is 0.5 (i.e., better than a random chance). In the calculations the observations regular grid mask of missing values (Fig. 4) is used for the forecast fields, and missing values are considered null (not considered), so both observational data and forecast fields are treated the same way. Additionally, zero accumulated precipitation is assumed outside the domain and only the neighborhoods with centers located inside the domain are used to calculate the FSS value (Jones 2014).

The FSS is divided by the sum of the mean square forecast and observed fraction [Eq. (5)] where N is the number of windows in the domain, P_{fcst} the forecast fraction, and P_{obs} is the observed fraction of the sliding window:

$$\text{FSS} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(P_{\text{fcst}} - P_{\text{obs}})^2}{\frac{1}{N}\sum_{i=1}^{N}(P_{\text{fcst}})^2 + \frac{1}{N}\sum_{i=1}^{N}(P_{\text{obs}})^2}. \quad (5)$$

In addition, a comparison among volumetric probability distribution functions (PDFs) is presented (Amitai et al. 2012). Such PDFs represent the distribution of rain volume (R), as they are constructed according to the relative contribution made by each precipitation interval to the total precipitation volume. The volumetric PDF is defined as the sum of the rain rates for a given $1 - \text{dB}R$ interval divided by the total sum of the rain rates [$\text{dB}R = 10\log(R/1 \text{ mm day}^{-1})$; units of $R = \text{mm day}^{-1}$]:

$$\text{PDF}(R_i) = \frac{\int_{R_i - 0.5}^{R_i + 0.5} R \times P(R)dR}{\int_{0}^{\infty} R \times P(R)dR}. \quad (6)$$

In the PDF(R) definition shown in Eq. (6), $P(R)$ represents the probability of rain-rate occurrence and i is the $1 - \text{dB}R$ interval bin number. The area under the pdf curve from $R = 0$ to $R = R_i$ represents the cumulative distribution function (CDF) of R_i, or the fraction of rain volume contributed by intensities below Ri. For example, half of the rain volume is contributed from $R < R_i$ for CDF(R_i) = 0.5. However, if the CDF represents the CDF of rain-rate occurrence (as opposed to volumetric CDF), then half of the rainy area (as opposed to rain volume) is contributed from $R < R_i$. Thresholds used are the same as for the categorical statistics adding extreme events 24-h accumulated precipitation bins 75 and 100 mm day^{-1}. This kind of measure represents a general comparison among the different precipitation distributions and, therefore, is complementary to the evaluation given by ETS, POD, FAR, and BIASS.

Finally, bias mean error was considered for atmospheric profile validation, where a perfect score is equal to zero [Eq. (7)]:

$$\text{bias} = \text{forecast} - \text{observation}. \quad (7)$$

3. Results

a. Precipitation fields predicted by the NWP models

Figure 4 (left panel) shows the observed total accumulated precipitation at each grid point for the entire period of study, i.e., 3 November 2018–16 December 2018. It has a spatially and temporally averaged 24-h precipitation of 3.3 mm (note that the missing values were not considered in the calculation, but null values yes). This figure further shows that the areas of maximum accumulated precipitation are observed over the SDC, indicating that complex terrain enhances precipitation. Finally, in the lower right and upper left quadrant of Fig. 4 (left panel), i.e., 33°S, 64.5°W, in the plains, we identify the areas with the minimum rainfall for the period of study, which indicates that the plain regions are drier, consistent with the typical summer precipitation regime in the area (Penalba and Vargas 2004).

The corresponding forecasted accumulated precipitation and the spatially and temporally averaged 24-h accumulated precipitation is also computed for WRFSMN, WRFCSU,
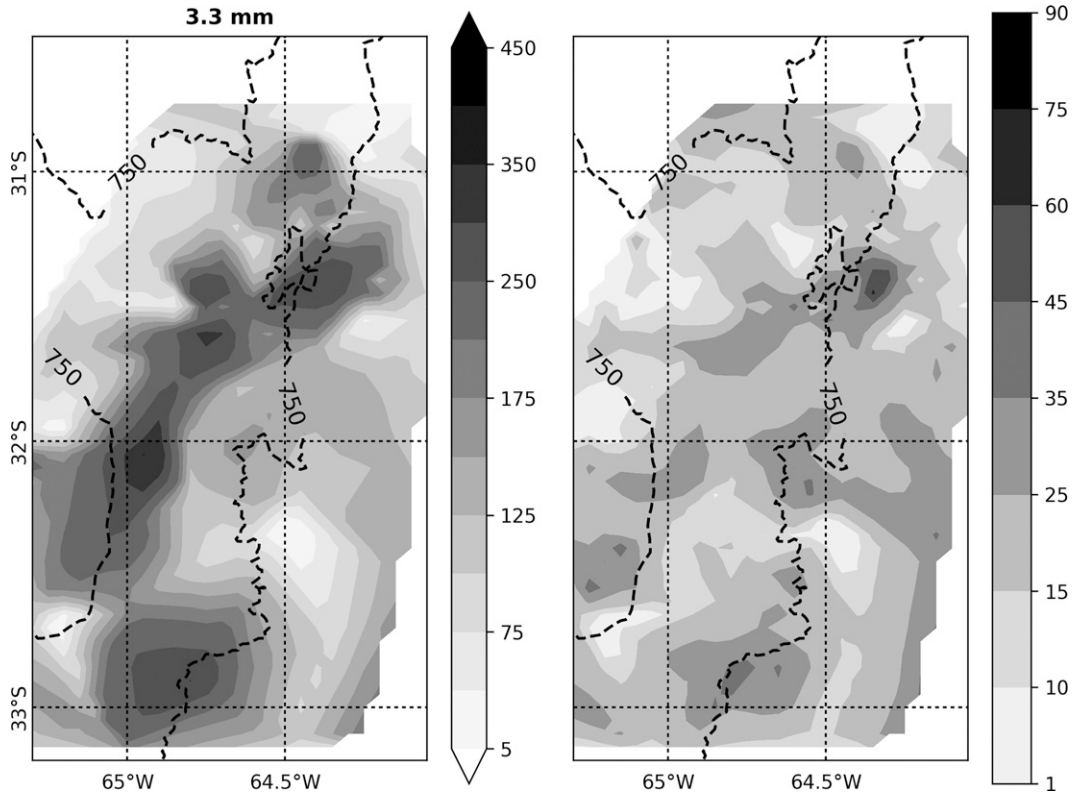
Fig. 4. Analyzed accumulated precipitation (mm) for the period of study, (left) with the spatially and temporally averaged precipitation in 24 h in the title and (right) 90th percentile of 24-h accumulated precipitation (mm every 24 h). The dashed line is the contour of 750 m MSL. Regions in white denote no observational data to interpolate.

WRFUIUC, and GFS, for each forecast time (see Fig. 5). In addition, Fig. 6 shows the difference between the accumulated precipitation forecasted by each NWP model and the observed accumulated precipitation (NWP models field minus observations field) to aid the analysis.

In general, all NWP models have a similar behavior. They show a shift in the location of maximum precipitation, and they fail to detect the observed precipitation maximum located near the southern edge of the verification domain. The differences between the forecasted and observed 24-h accumulated precipitation in Fig. 6 reveal that all the NWP configurations studied show the largest overestimations over the orography and underestimations to the west of the SDC. The spatially and temporally averaged 24-h forecasted precipitation (listed in the different titles of Figs. 5) shows a general overestimation over the domain of interest.

The WRFSMN F 36-h particularly, shows larger differences in the orographic precipitation enhancement zone than the forecast runs initialized at 1200 UTC, i.e., F 24-h and F 48-h (Fig. 6), due to the different initialization times of the NWP models. In the region located west of the SDC, all the forecasts show underestimation, negative biases, being the F 24-h the forecast with the largest differences. Positive biases are shown eastward of SDC.

WRFCSU shows for F 36-h and F 48-h forecasts that the maximum differences are extended to the north-northeast of the study region. The largest positive differences with respect to the observation field are observed to the north of the complex terrain where the topography elevation and slope are less pronounced than the central part of the SDC. The WRFCSU configuration has the largest averaged 24-h accumulated precipitation differences with respect to the observed field for the F 24-h and F 36-h forecasts (1.9 mm day^{-1} in both forecast lead times, respectively); whereas the WRFSMN shows the largest differences for the F 48-h forecast (1.8 mm day^{-1}).

Regarding WRFUIUC, it overestimates the precipitation in some regions and for all forecast times, but in a smaller magnitude compared to the previously mentioned WRF Model configurations. Furthermore, the low bias of WRFUIUC compared with the observations show a larger areal extent than the other WRF Model configurations. In addition, the maximum precipitation areas for F 24-h and F 36-h are correctly placed. The spatially and temporally averaged precipitation shows that the WRFUIUC has the closest values to the observed field.

The lower-resolution GFS forecasts show several precipitation maxima both over the orography and the plains east and northeast of the SDC. The behavior of the three forecast times analyzed is similar: overestimations over the SDC and in smaller magnitude, to the east of them, and underestimations
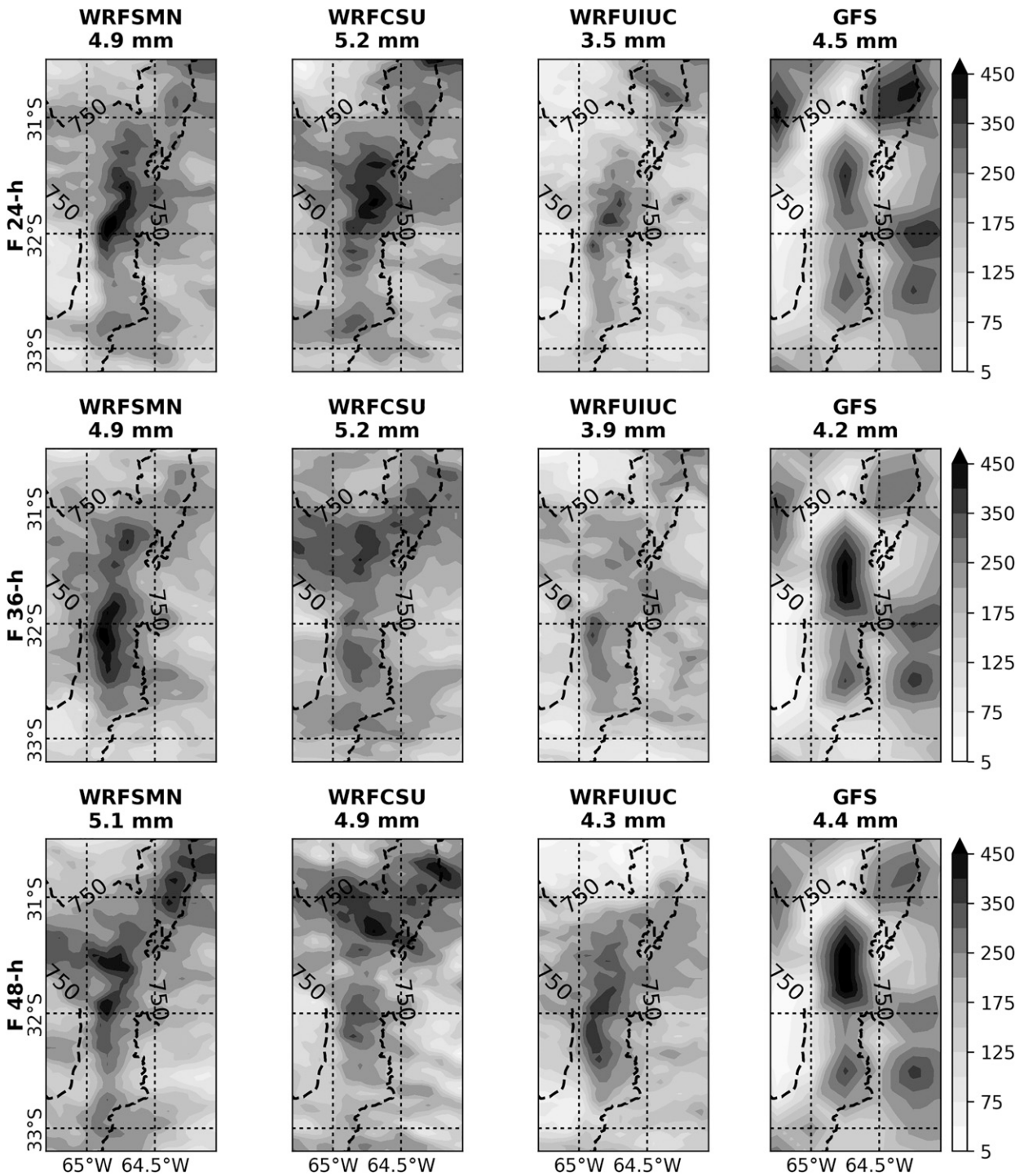
FIG. 5. Accumulated precipitation fields (mm) for the period of study of all the NWP models (columns) and forecast times (rows). Rows show the precipitation in the 24-h period ending at forecast hours 24, 36, and 48, respectively. (first column) WRFSMN, (second column) WRFCSU, (third column) WRFUIUC, and (fourth column) GFS. The spatially and temporally averaged precipitation in 24 h (considering the same mask of Fig. 4) is indicated in the title of each panel. The dashed line is the contour of 750 m MSL.

to the west of the orography. Gridpoint convection has been a major issue in the GFS model prior to the implementation of the Finite Volume Cubed-Sphere dynamical core during 2019

(FV3, https://www.weather.gov/news/fv3). GFS microphysics rains out much faster than the available WRF microphysics schemes.
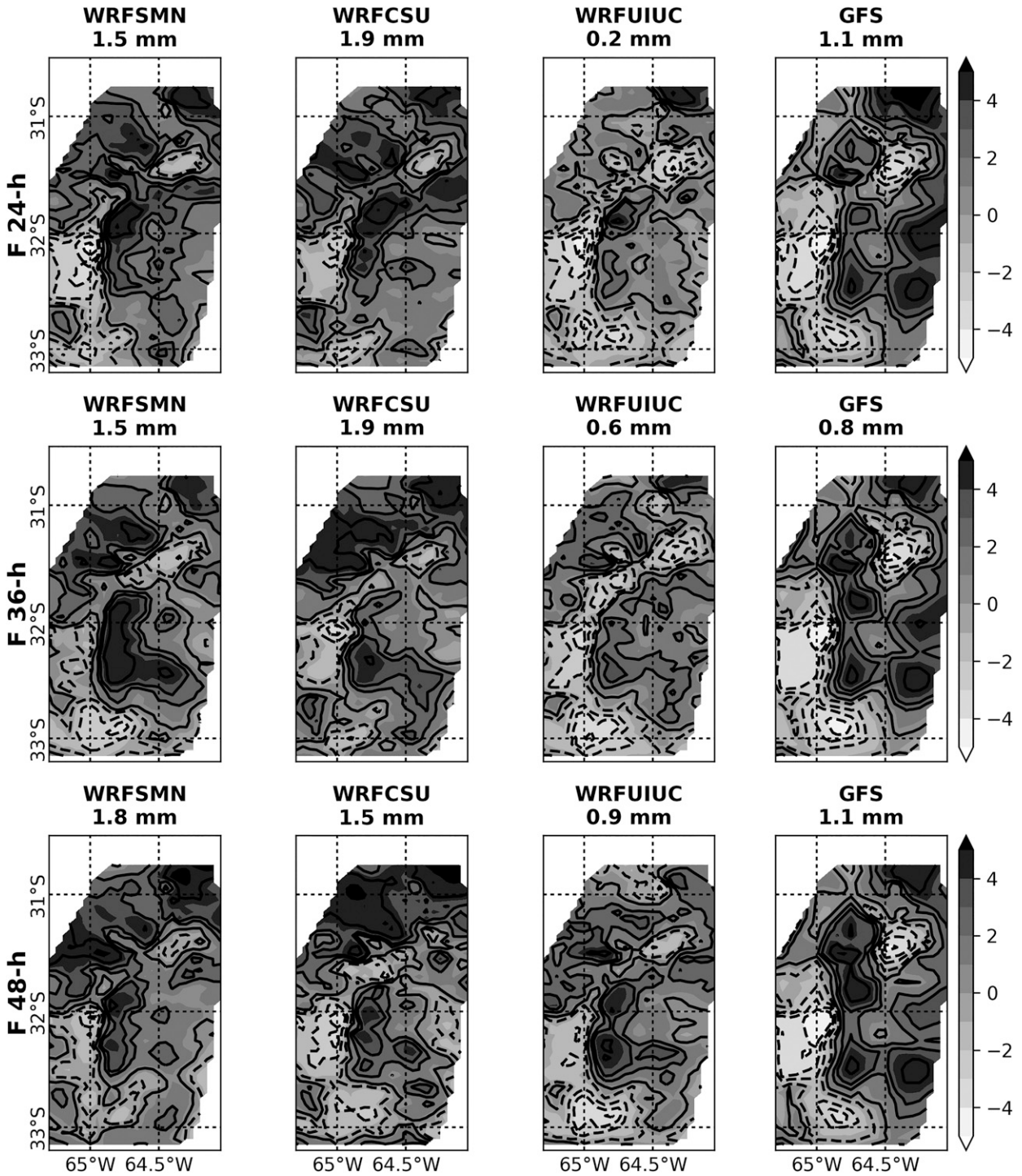
FIG. 6. Mean differences between accumulated precipitation fields (mm day^{-1}) for the period of study between NWP models and observation fields (NWP models field $-$ observations field) for each forecast time divided by the study period (43 days). Solid contour lines show positive values, and the dashed contour lines show negative values. (top) F 24-h, (middle) F 36-h, and (bottom) F 48-h forecasts. (first column) WRFSMN, (second column) WRFCSU, (third column) WRFUIUC, and (fourth column) GFS. In addition, the differences between the spatially and temporally averaged precipitation in 24 h forecasted and observed are indicated in the title of each figure (using one decimal place rounding).
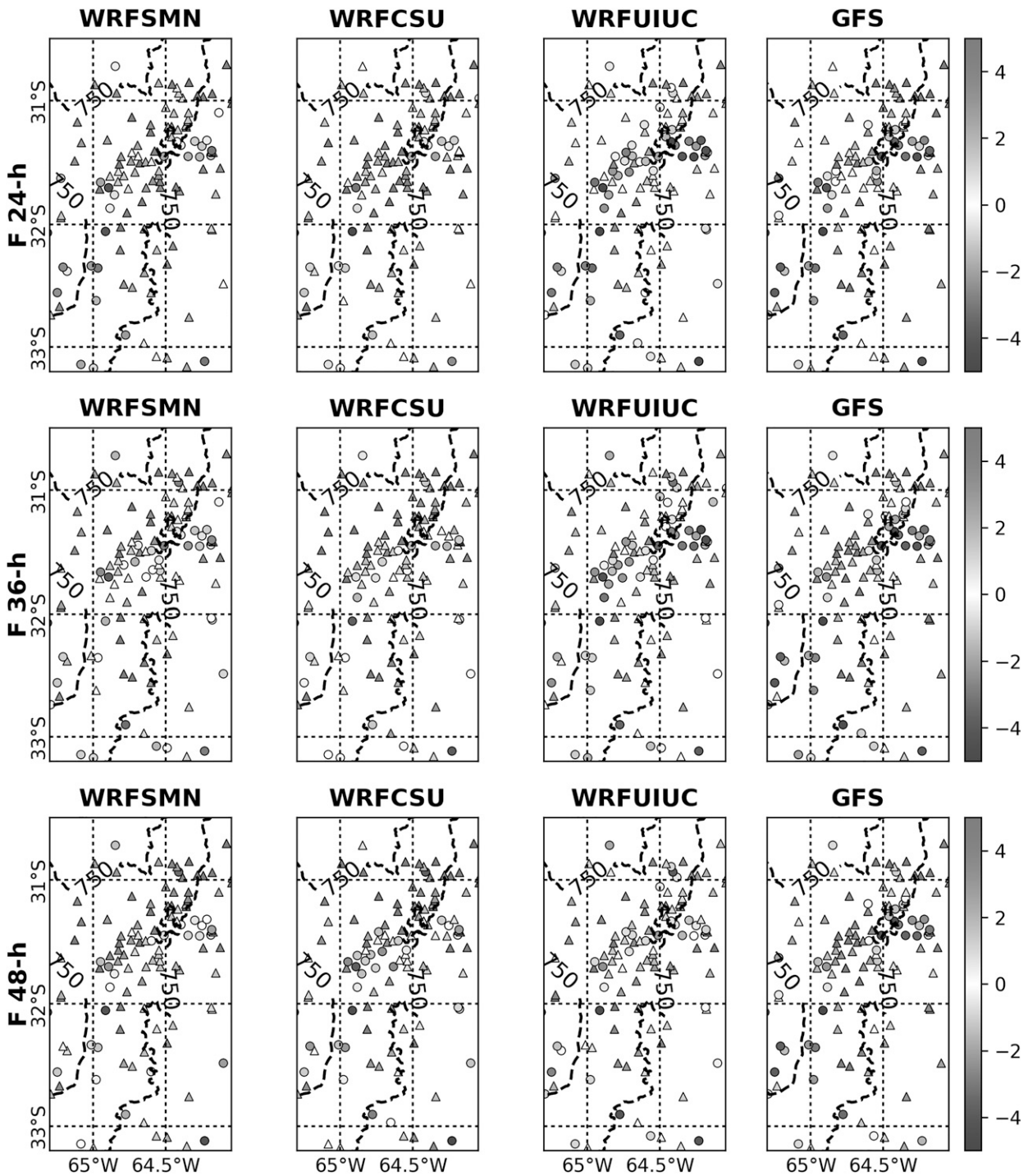
FIG. 7. Surface stations forecast mean bias of all the NWP models (columns) and forecast times (rows). (top) F 24-h, (middle) F 36-h, and (bottom) F 48-h forecasts. (first column) WRFSMN, (second column) WRFCSU, (third column) WRFUIUC, and (fourth column) GFS. Circles denote negative values and triangles positive. The dashed line is the contour of 750 m MSL.

Additionally, the mean bias is calculated for all the forecast lead times using Eq. (7) and a nearest neighbor interpolation of the NWP to the surface station locations (Fig. 7). All the NWP configurations studied show the largest overestimations over the orography and underestimations to the west of the SDC and around 31.5°S and 64.25°W. The WRFUIUC shows lower values of wet and dry biases for all the forecast times, but for F 24-h and F 36-h has a larger number of stations with

negative biases (underestimation) compared with the other NWP models. In general, the GFS is the NWP model with the largest bias magnitudes, except in the F 36-h where the WRFCSU has wider overestimations extended to the north-northeast of the study region. Last the WRFSMN F 36-h shows minor values of negative biases in the SDC and to the west than the forecast runs initialized at 1200 UTC, i.e., F 24-h and F 48-h. Note that these results are consistent with the discussion presented above regarding Fig. 6. Authors are aware that the verification of the NWP models is sensitive to the method of validation chosen, nearest neighbor interpolation or regular grid and which method of interpolation is used, which implies an additional difficulty to evaluate the different forecasts products. Using a regular grid instead of nearest neighbor interpolation produces smoothing effects in the calculated scores; but also, the chosen grid determines which statistical methods can be computed. Therefore, hereafter we decided to show all the statistics calculated using the 0.05° regular grid linearly interpolated.

The 90th percentile of accumulated daily precipitation for each model configuration is shown in Fig. 8 for the period of interest. Similarly, Fig. 4 (right panel) shows the 90th percentile of the observed accumulated daily precipitation. The Fig. 8 shows that the high-resolution NWP models (i.e., the different WRF configurations) have their extremes in the north region of the study area and over the SDC; which is consistent with the previous analysis of accumulated daily precipitation and the observations. Specifically, comparing the NWP model fields with the observations, we notice a particular similarity in the detection of a relative maximum near 33°S and 65°W in the WRF Model configurations F 24-h. Although this relative maximum is overestimated, the WRFUIUC has the values closest to the percentile of the observation, but the maximum is misplaced. Most of the NWP model configurations overestimate the 90th percentile; however, the comparison of the 90th percentile fields with the accumulated precipitation shows that the statistics have a more homogenous distribution and a smaller difference between SDC and the plains areas.

Figure 9 shows the 24-h accumulated precipitation averaged in latitude between 33.2° and 30.5°S in Hovmöller diagrams (time–longitude plots) for the whole study period and for each NWP model configuration and the available observations. We plot the diagram with respect to longitude since this captures the principal eastward propagation of precipitation systems over the midlatitudes in South America (Vila et al. 2008; Salio et al. 2007), which is seen in the events of 4 November and 13 December although neither of the NWP models manage to represent the observed propagation. Differently, the 11 and 26 November events were convection systems with initiation in the SDC (http://catalog.eol.ucar.edu/relampago) with a south–north propagation, which is not reflected in the figure due to the small domain considered. Overall, the figure reveals a relatively accurate forecast of the events and their trajectory. Besides the general detection of events, for 11 November we observe that the high-resolution NWP models overestimate the amount of precipitation more than the GFS, particularly for F 24-h forecasts. While for

27 November, the precipitation location and magnitudes are better represented by WRF than the GFS, all the models fail to detect the 1 December event for F 24-h, but they correctly represent it at F 36-h, particularly WRFSMN and WRFCSU. Overall, the NWP models displace temporarily the maximum of precipitation i.e., precipitation is observed later than the forecast, for the three forecast lead times (F 48-h results are similar to F 24-h ones, not shown).

The categorical statistics for the entire area and period of interest were obtained for each NWP model, using Eqs. (1) and (2) (see Fig. 10).

In terms of the BIASS, all the convective NWP models overestimate high precipitation thresholds for the three forecast lead times. Focusing on the F 24-h forecasts, the WRFSMN and WRFCSU overestimate the 24-h accumulated precipitation (BIASS greater than one) for all the thresholds; whereas the WRFUIUC underestimates it (BIASS less than one) for thresholds smaller than 15 mm day^{-1} and overestimates it for thresholds above that value, showing an outstanding performance. It is remarkable that the GFS performs better than WRFUIUC for thresholds bigger than 20 mm day^{-1}, but with opposite sign.

For F 36-h and F 48-h forecasts all the WRF Model configurations have a positive BIASS at all thresholds, supporting the above analysis of total accumulated precipitation (Figs. 5). The GFS has the closest BIASS to the perfect score for the F 36-h and F 48-h for thresholds bigger than 15 mm day^{-1}, whereas the best ratio between forecasted and observed events for the smaller thresholds corresponds to WRFUIUC.

The bottom row in Fig. 10 shows the F 24-h, F 36-h, and F 48-h 24-h accumulated precipitation ETS. All the NWP models exhibit a similar behavior: as the threshold value increases the ETS gets closer to zero, meaning that the correctly predicted precipitation is almost equal to the considered number of random hits. For F 24-h the WRFSMN and WRFCSU stand out in their performance, they have better correspondence between observed and forecast events considering random chance. At the same time, the GFS F 24-h shows the lowest ETS values. For the F 36-h forecasts, the WRFUIUC configuration has the values of ETS closest to 1 between the 10- and 30-mm day^{-1} thresholds. For the F 48-h forecasts, the GFS has a better performance than F 24-h and F 36-h forecasts, obtaining similar ETS values to the NWP WRF Model configurations except the WRFCSU that has the smallest values for the middle thresholds (between approximately 8 and 30 mm day^{-1}). The calculated score values in the present study are within the range of previously reported statistics calculated computed over the region of interest with WRF forecasts at 4 and 40 km resolutions, despite the fact that the previously examined analyses focused on hourly accumulated precipitation (Matsudo et al. 2015; Dillon et al. 2016). In addition, the decreasing behavior of the ETS score can be explained by the fact that it has a dependency on the climatology, the sample size, and on the locations where precipitation was more frequent (Hamill and Juras 2006).

Considering these results, there is a general agreement in the behavior of all the WRF Model configurations at all forecast times which shows an overestimation in terms of a wet
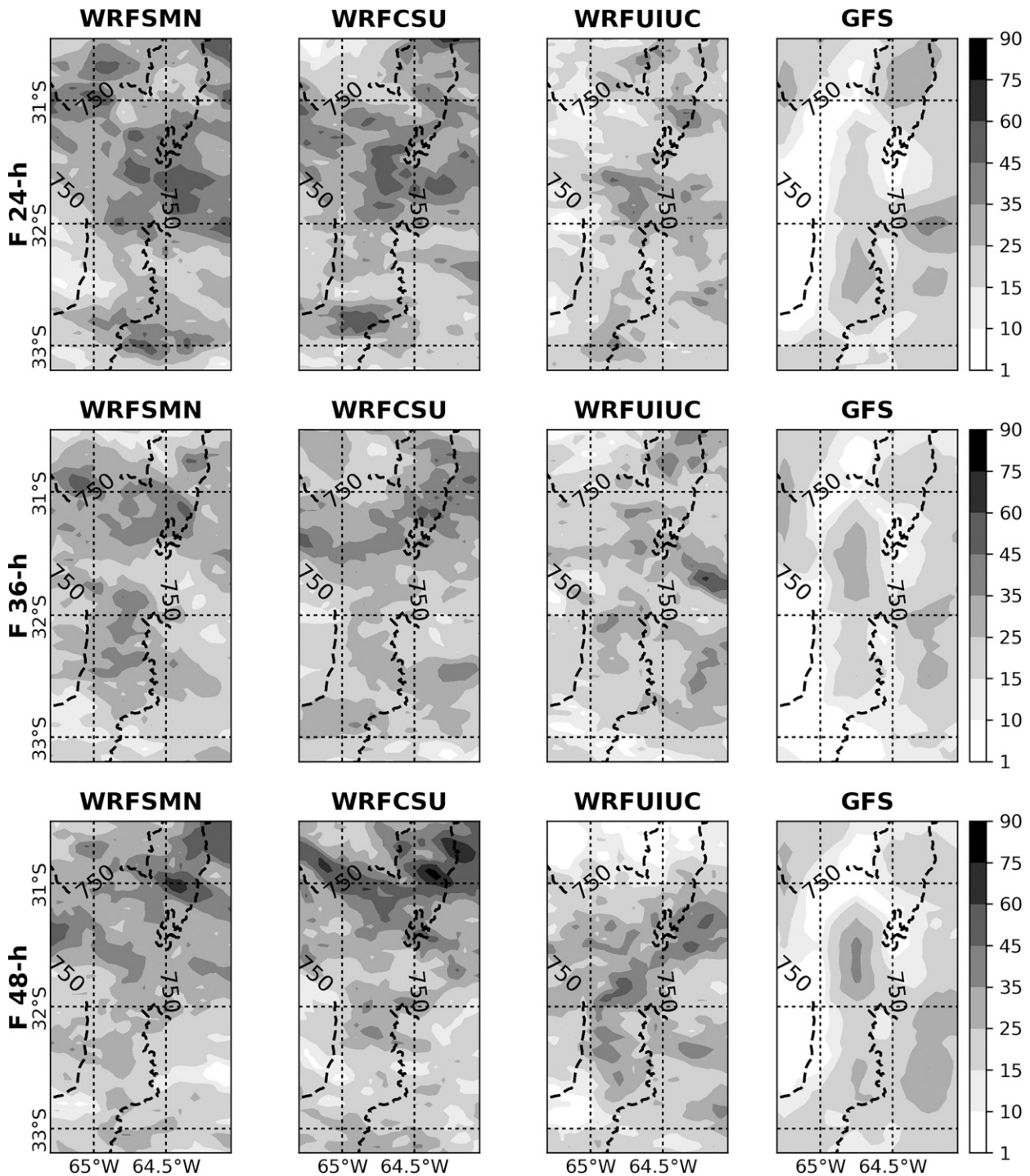
FIG. 8. The 90th percentile of 24-h accumulated precipitation fields (mm) for the period of study of all the NWP models (columns) and forecast times (rows). (top) F 24-h, (middle) F 36-h, and (bottom) F 48-h forecasts. (first column) WRFSMN, (second column) WRFCSU, (third column) WRFUIUC, and (fourth column) GFS. The dashed line is the contour of 750 m MSL.

BIASS at thresholds superior to 5 mm day^{-1}, with the WRFUIUC being the high-resolution model closest to the optimal value. However, the GFS stands out for its BIASS score for all the forecast lead times at thresholds bigger than

15 mm day^{-1}, but with a dry BIASS in F 24-h. For the ETS, although the results do not show a consensus for a best NWP model forecast, the F 24-h and F 36-h show that all three WRF Model configurations perform mostly better than the
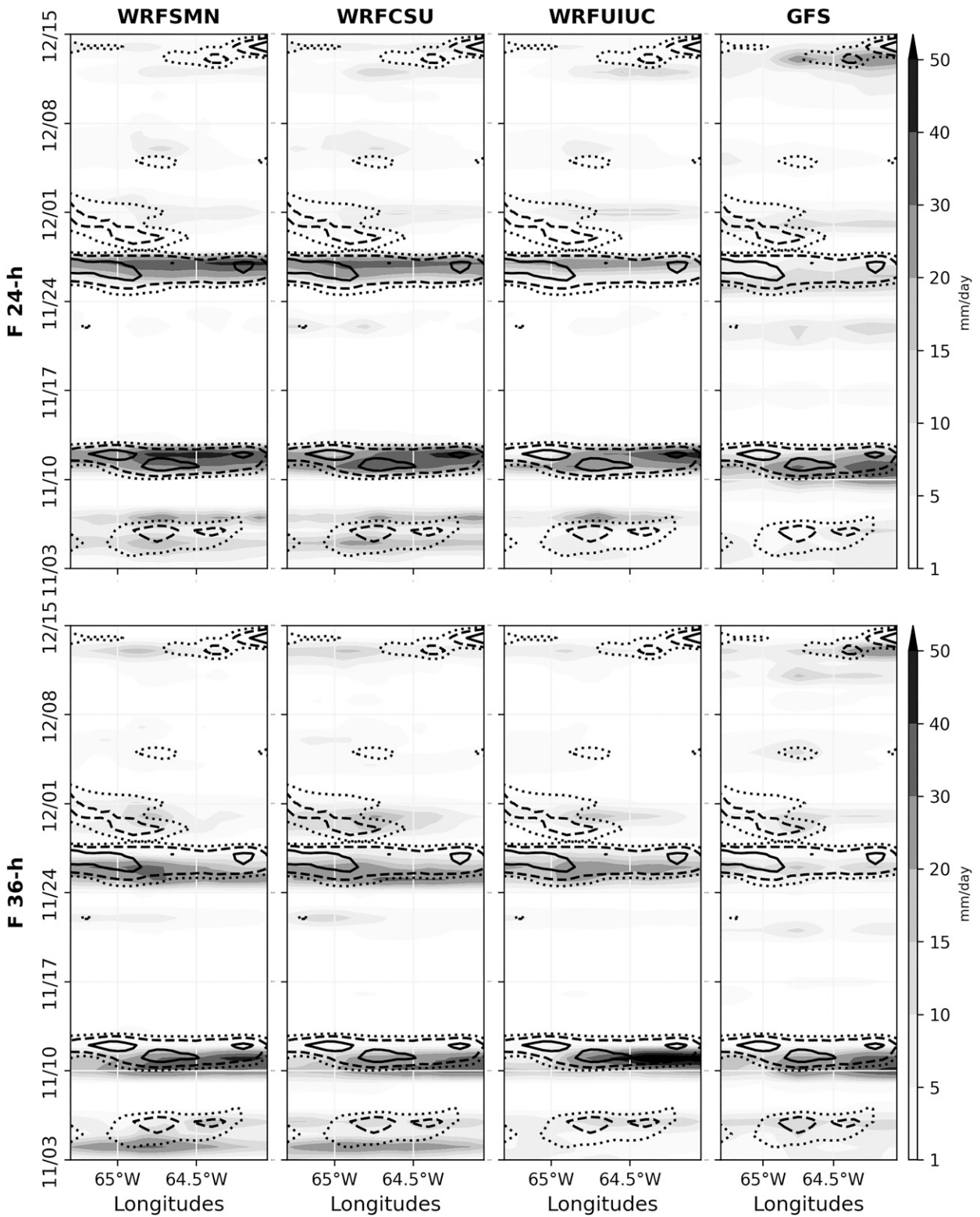
FIG. 9. Hovmöller (time–longitude) of 24-h accumulated precipitation fields for the period of study of all the NWP models (columns) and forecast times (rows). (top) F 24-h and (bottom) F 36-h. (first column) WRFSMN, (second column) WRFCSU, (third column) WRFUIUC, and (fourth column) GFS. The dotted, dashed, and solid lines are the contours equal to 5, 10, and 20 mm in 24 h, respectively, of the observation field.
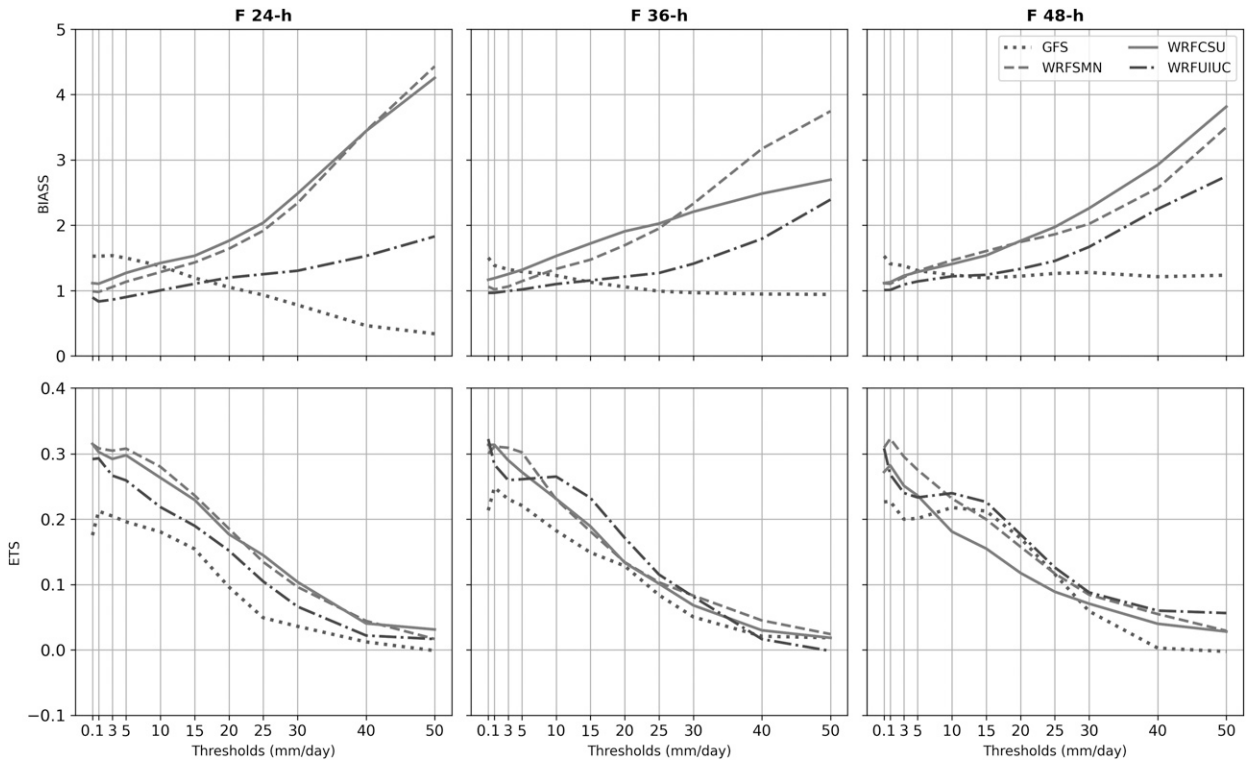
FIG. 10. BIASS and ETS for each NWP model (rows) for (left) F 24-h, (center) F 36-h, and (right) F 48-h forecast times.

GFS forecasts. Whereas for F 48-h the GFS outstands some WRF Models for some thresholds, in terms of ETS.

Figure 11 shows the results of the calculated FSS, which measures the 24-h accumulated precipitation forecast success according to the square neighborhood size for a given 24-h accumulated precipitation threshold. The expected behavior of the score is an increase as either the size of the square neighborhood (in this case the gridpoint scales) is higher or the threshold is lower, which is observed for all the NWP models configurations.

The performance of all the NWP models seems to worsen (lower FSS values) as the 24-h accumulated precipitation threshold increases, evidencing the difficulty of forecasting extreme events. This behavior of FSS scores for large thresholds in high-resolution models was also observed by (Ebert 2009). Overall, the WRFUIUC is shown to obtain the best performance in all the forecast lead times for all the different gridpoint scales, although the WRFSMN shows the best performance in the F 36-h for all the neighborhood grid scale in the extreme threshold. Either the WRFCSU or the GFS show the worst performance for nearly all neighborhood grid scales and thresholds. For this NWP WRF Model configuration, as the thresholds of 24-h accumulated precipitation increase the FSS values decrease, getting farther away from the results of the remaining high-resolution NWP models, showing a better detection for the smaller thresholds in the F 24-h and F 36-h. For higher thresholds, the GFS has the worst performance in terms of FSS values in all the forecast times, supporting previous statistical results (ETS close to zero). The WRFUIUC

configuration stands out from the others as it obtains the best FSS for almost all the thresholds and neighborhood grid scales. The highest values of event detection are found in the F 24-h (up to 0.99).

b. Verification by terrain elevation

The prediction of accumulated precipitation by the NWP models in this area is challenging due to the steep terrain and orography which plays an important role in the distribution of precipitation over land (Robinson and Businger 2019). As we noted in the previous section, there is a tendency in the models studied to overestimate precipitation over the mountainous areas. To further analyze the performance of the forecasts as a function of the terrain, the study area is subdivided according to the height MSL (orography).

Two regions have been defined:

- Orography: altitude greater or equal to 1000 m.
- Plains: altitude less than 1000 m.

Since each NWP model considered has a different representation of the terrain, if the terrain from each model is used, the locations of these regions would differ among them (Fig. 3). Therefore, it was decided to divide the regions of each sample using an independent dataset of terrain elevation (Fig. 1).

The categorical statistics are obtained for each subregion separately for each model, considering the entire period of study. The F 24-h results for both subregions are shown in
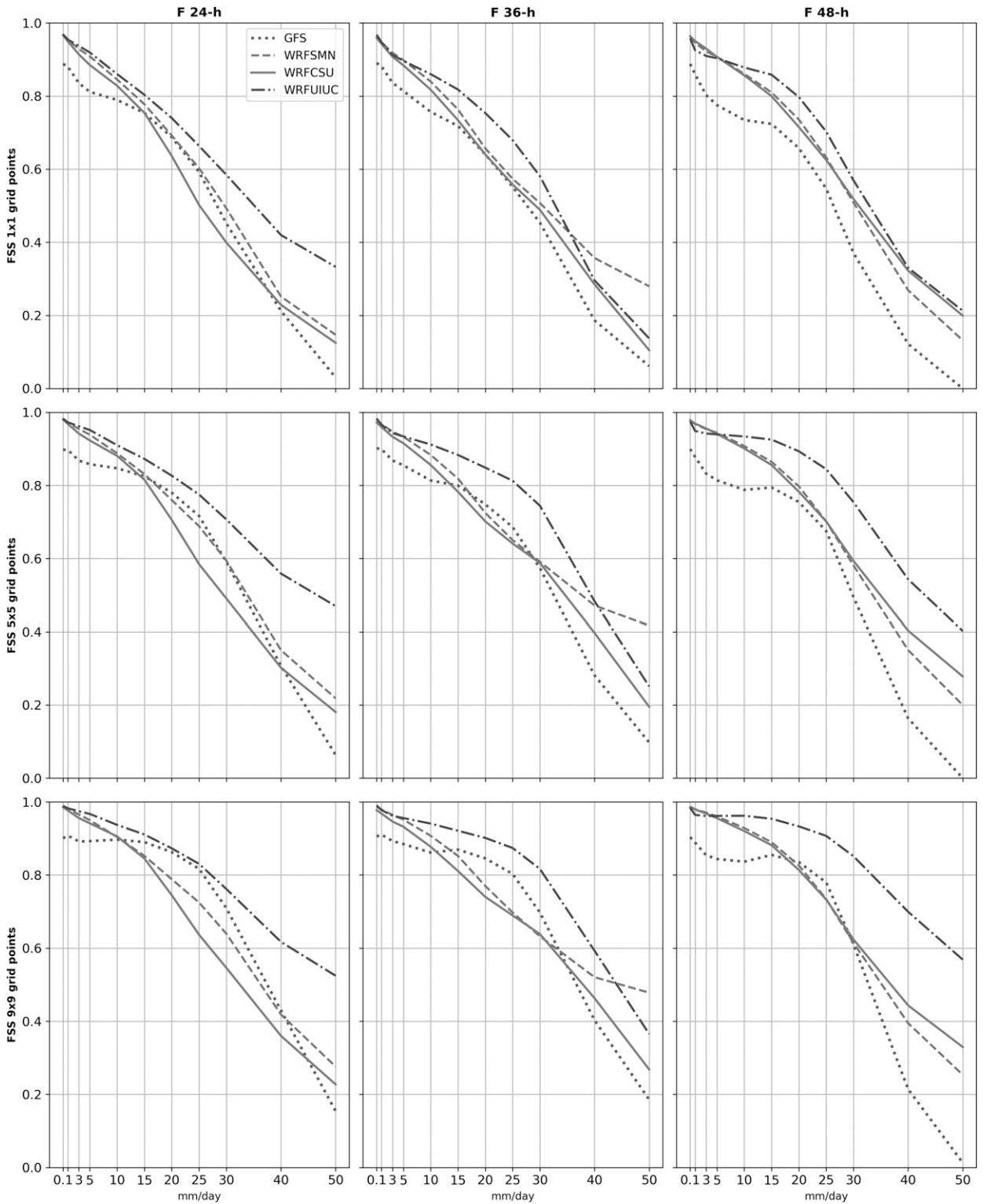
FIG. 11. FSS for each NWP model for different gridpoint scales of interpolation (rows) for (left) F 24-h, (center) F 36-h, and (right) F 48-h forecast times. (top)1×1, (middle) 5×5, and (bottom) 9×9 grid points. All data were interpolated to a 0.05° resolution.
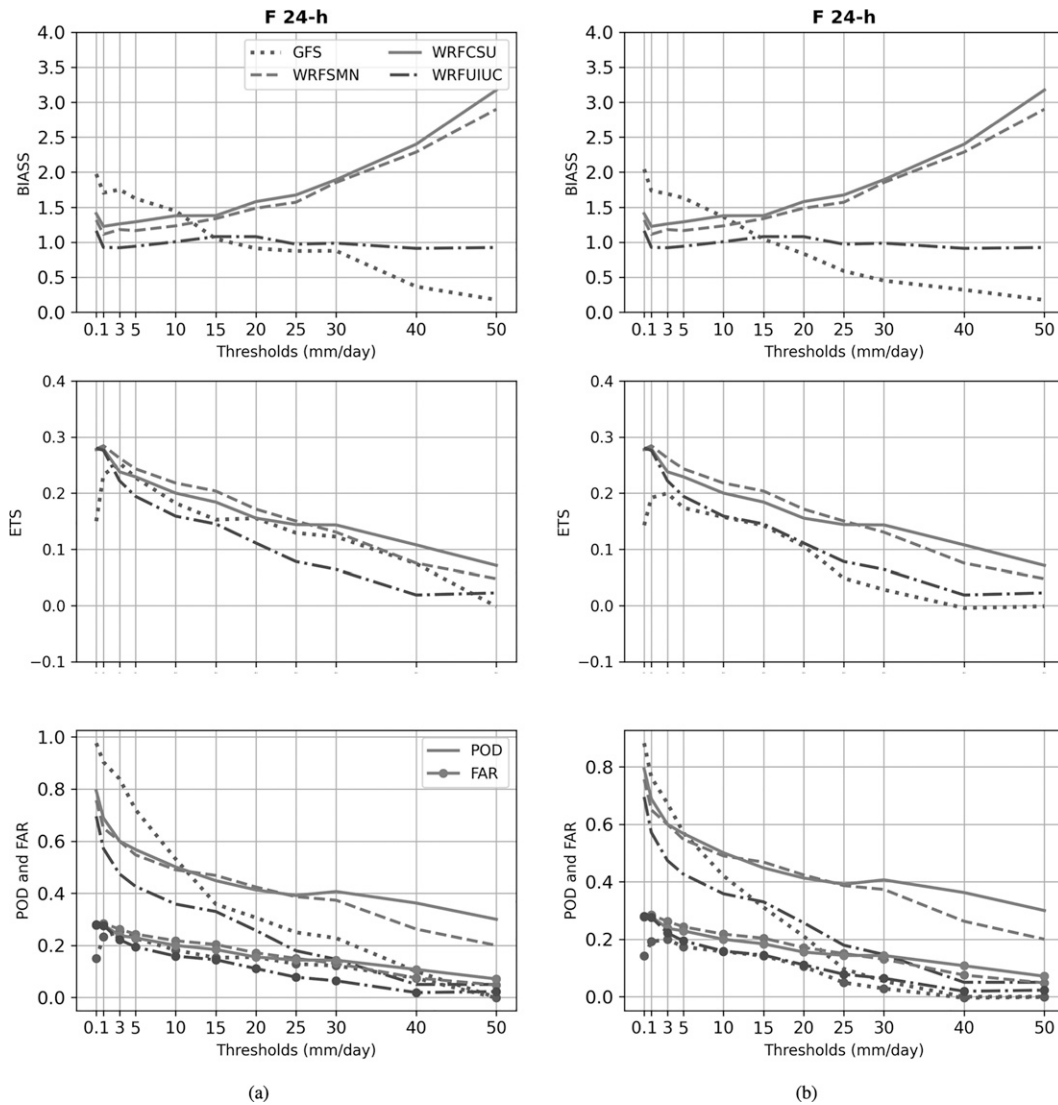
FIG. 12. Categorical statistics BIASS, ETS, POD, and FAR for each model for the forecast time F 24-h (a) orography class and (b) plains class.

Fig. 12. Generally, the orography BIASS values are further from one than those calculated in the plains for all the NWP models. For thresholds up to 15 mm day^{-1}, WRFUIUC obtains the most unbiased performance in both subregions, while for the thresholds above 15 mm day^{-1}, the GFS model shows a better performance. Among the WRF configurations, the WRFUIUC configuration has the closest values to 1 for all the thresholds and is never greater than 2. In the plains (Fig. 12b), all the NWP models have a similar behavior for thresholds under 10 mm in 24-h, whereas the WRFSMN and WRFUIUC are equally apart from the perfect score, with WRFSMN overestimating and WRFUIUC underestimating. For higher thresholds, the GFS shows the least overestimation followed by WRFUIUC. The GFS in the orography for the largest thresholds (>25 mm day^{-1}) has a dry BIASS while in the plains the opposite is shown. For both subregions the

WRF Model configurations generally stand out in terms of improved ETS. As the thresholds increase, all the NWP models have a similar behavior. The POD has analogous outcomes for both regions, with GFS scores near one for the thresholds below 5 mm day^{-1}, indicating that it has the greatest number of correct forecasts. In the orography class, the WRFCSU configuration has the values closest to one between 5 and 25 mm day^{-1}, while the WRFSMN configuration scores better for the largest thresholds. The behavior of the WRF configurations in the plains is similar, while the WRFSMN once again in the largest thresholds has a greater number of correct forecasts. The WRFUIUC shows the least number of false alarms (FAR) in the plains (Fig. 12b) in comparison with the orography area, and also it is the NWP model that reaches the closest value to zero in both subregions for the thresholds below 30 mm day^{-1}.
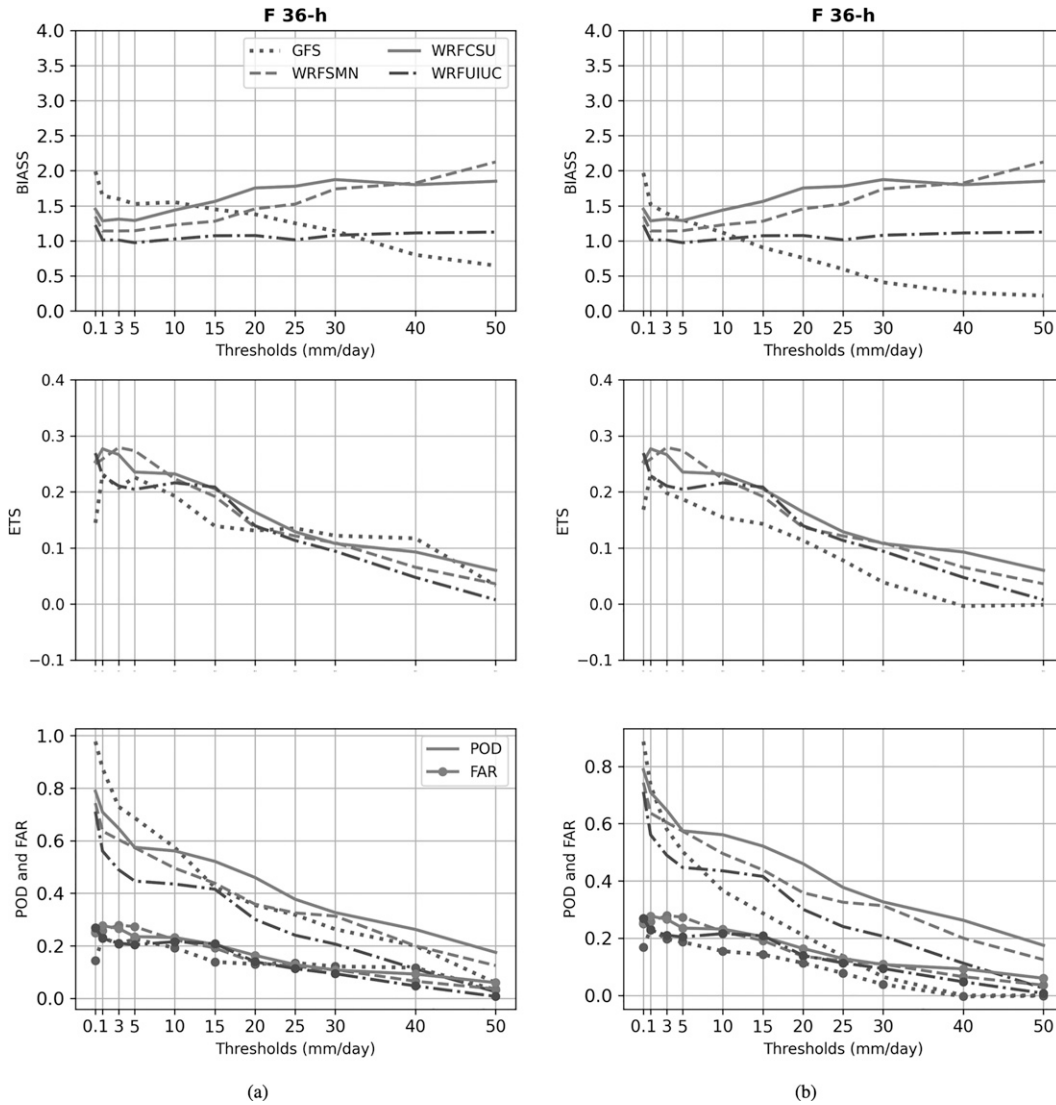
FIG. 13. Categorical statistics BIASS, ETS, POD and FAR for each NWP model for the forecast time F 36-h
(a) orography class and (b) plains class.

For F 48-h forecasts similar results to F 24-h are encountered for the BIASS and FAR (not shown). Regarding the BIASS, the WRFCSU has the worst performance for both subregions—being the NWP model with the largest wet BIASS. Regarding the ETS the WRFUIUC obtains the best performance for both subregions being the NWP model with the largest number of events that were both observed and forecasted. Last, the highest POD values correspond to the WRFCSU.

For the F 36-h forecasts (Fig. 13), similar differences between the subregions as in F 24-h forecasts are observed: the values of BIASS for the plains are lower in module than in orographic terrain for thresholds below 40 mm day^{-1}. The WRFUIUC BIASS is the closest to one in both regions, considering the high-resolution models, although in the plains the WRFSMN seems better for thresholds under 10 mm day^{-1}.

For all the forecast times and subregions analyzed, the BIASS corresponding to WRFSMN and WRFCSU shows smaller values for the smallest thresholds, indicating a larger overestimation in cases of greater intensity of precipitation. This was also documented by (Weisman et al. 2008) for WRF-ARW precipitation forecasts for the spring-summer season of 2005 over the United States. Regarding the GFS, for thresholds greater than 15 mm day^{-1} it shows the closest values to one in both subregions.

For the ETS, the behavior for the orography is similar to that for the F 24-h results (Fig. 13a); but it changes for the plains, where in general the WRFCSU (GFS) obtains the closest (furthest) scores to one (Fig. 13b). Regarding the POD, in the plains and orography classes the WRFCSU generally obtains a larger number of correct forecasts than the other NWP models. Once again, the WRFUIUC shows the lowest
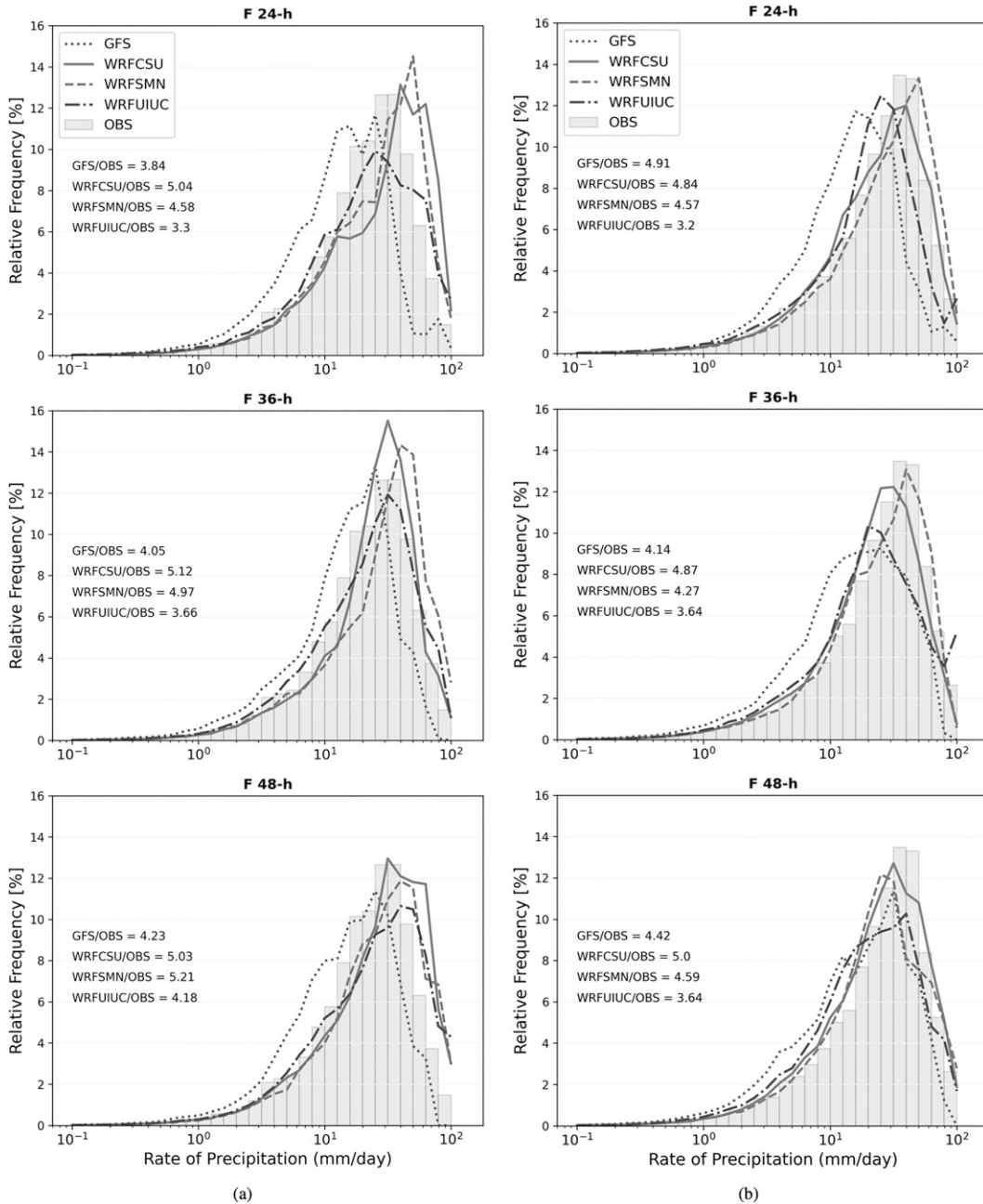
FIG. 14. Volumetric PDF for all the forecast times and with their corresponding rain volume ratio (a) orography class and (b) plains class.

number of false alarms (FAR) for the plains class and presents a similar behavior as in F 24-h forecasts. For the orography class the WRFSMN and WRFCSU have the best performance for the different thresholds.

Using the methodology introduced in section 2d and the Eq. (6), we calculated the values of the volumetric PDF for each subregion of the area of study with the aim of finding differences between the performance of each NWP model for each scenario (Fig. 14). In addition, to support the analysis,

the difference between the volumetric PDF curves of the observations and each NWP model is shown (Fig. 15).

First, for the F 24-h forecasts the WRFUIUC shows the best performance in both subregions since its volumetric PDF curves adjust quite well to the observations, showing the smallest differences for almost all the precipitation rates in Fig. 15. The WRFCSU and WRFSMN configurations provide a better performance over the plains in comparison with the orography for the largest values of 24-h accumulated
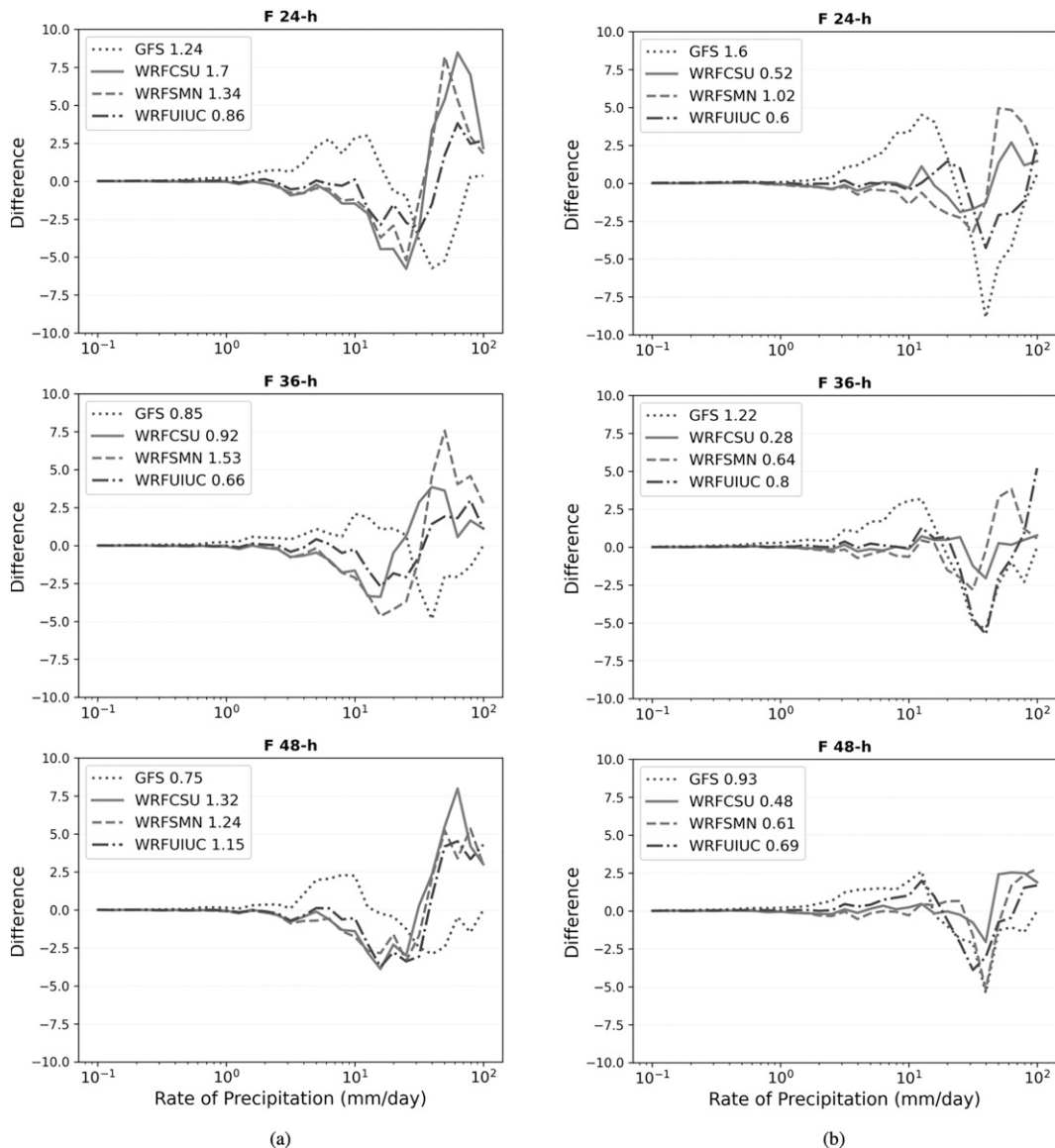
Fig. 15. Difference between the observed and forecasted volumetric PDF for each NWP model, for all the forecast times; next to their label the corresponding mean of the absolute values are included for (a) orography class and (b) plains class.

precipitation. Regarding the detection of the most frequent 24-h precipitation observed record, in the subregion of complex terrain the WRFCSU succeeds in achieving a similar value (13% for 40 mm day^{-1}), while in the subregion of the plains the WRFSMN shows the best values (14% for 50 mm day^{-1}).

Second, for the F 36-h forecast, the PDF curve of WRFUIUC and WRFCSU seems equidistant to the PDF curve of observations over the orography, whereas the WRFCSU stands out with the smallest differences with respect to the observations in the plains. In both subregions the WRFSMN is the model which has the best detection of the PDF maximum compared with the observations, meaning that it has the best performance

forecasting the most frequent 24-h accumulated precipitation observed (40 and 50 mm day^{-1} for orography and plains, respectively).

Last, for the F 48-h forecasts, for both subregions the WRF configurations indicate similar behavior to the observation PDF, although for the orography (plains) the WRFSMN (WRFCSU) stands out as the best. The detection of the most frequent observations by the models occurs more accurately in the orography than in the plains, being the WRFSMN and WRFCSU the ones which achieve the best performance.

The GFS has the largest differences with the observations: over forecasting (under forecasting) thresholds below (above)

the observation maximum of relative frequencies, demonstrating a less satisfactory performance in comparison with the other NWP model. This result agrees with (García Skabar et al. 2018), who showed that a global model overestimates (underestimates) thresholds below (above) the mode of relative frequencies over Argentina.

Among the different WRF configurations, there is no optimal configuration with respect to fitting the observation PDF curve. For each forecast time, a different WRF configuration stands out: for F 24-h the WRFUIUC in both subregions; for F 36-h the WRFCSU over the plains terrain and the WRFUIUC in the orography; for F 48-h the WRFSMN in the orography.

Finally, the rain volume ratio is calculated simply by integrating all the rain intensities in the PDFs of each NWP model and dividing it by the integration obtained from observations. All the values of the rain volume ratio are above one, showing a generalized overestimation of the total 24-h accumulated precipitation seen in previous analysis.

c. Atmospheric sounding evaluation

The mean bias profiles of temperature (T), dewpoint (Td), and u and v wind components at 1200 UTC at the Villa Yacanto de Calamuchita, Córdoba, and the Villa Dolores sounding sites (Fig. 1) are shown for F 24-h forecast in Fig. 16. Note that Villa Yacanto is closest to the highest peak in the SDC and is located in a mountainous region.

In Villa Dolores, which is located upwind from the SDC (considering the mid and upper-level flow), there is evidence of drier low level conditions in all the NWP models (from 940 to 700 hPa) while the upper levels present a positive humidity bias. Also, as the forecast lead time increases (only F 24-h shown), the models continue showing dry conditions at lower levels. In the midlevels, the biases are closer to zero. Comparing among all the models analyzed, the GFS at the midlevels (700 hPa) has an almost unbiased result while it shows the largest wet bias in upper levels (above 500 hPa). All the WRF Model configurations behave similarly but WRFSMN has the values closest to a null bias at the lower levels; for the remaining mid and higher levels there is no clear best model performance as it varies depending on the forecast time and the pressure level. Regarding the temperature profiles, at low levels all the models show a warm bias surpassing 2°C in the case of GFS, and not reaching 1°C in the case of WRFSMN, while in general the four NWP models stay nearly unbiased above 800 hPa approximately. At this location (upwind the SDC), the 1200 UTC mean circulation is zonal wind toward the SDC and meridional poleward near the surface (940–900 hPa) (not shown). Zonal wind (u component) bias evidence underestimation by all the NWP models near the surface (920–900 hPa), being the WRFSMN the NWP configuration closest to have null bias. On the contrary in upper levels (650–400 hPa) the NWP models overestimate the zonal wind, behavior which is observed for all the sounding sites. Considering the meridional wind component, there is a difference between the higher resolution WRF Models and the GFS: overestimation and underestimation of the flow magnitude at lower levels, respectively.

For Villa Yacanto the dry bias near the surface of all the NWP models is not as dry as in Villa Dolores. The midlevel values are closer to be unbiased for all the NWP WRF configurations. For the levels above 600 hPa there is a wet bias with the GFS having the largest differences. The WRFSMN is nearly unbiased from the surface up to 700 hPa for all the forecast times (only F 24-h shown). The temperature representation is more accurate than the humidity, particularly for the GFS which is unbiased along all the atmosphere, while the different WRF configurations show a little cold bias at midlevels. The results evidence zonal wind bias overestimation of all the NWP models for the upper and lower levels. The GFS evidences that the zonal (meridional) wind is overestimated (underestimated) from the surface level to 300 hPa, except at 600 hPa where the meridional wind is unbiased. The WRFCSU (WRFUIUC) is the closest to be unbiased for lower (upper) levels for the zonal wind. Last, the WRFUIUC underestimates the meridional wind component in the lower levels while the remaining WRF configurations overestimate it.

At the Córdoba sounding site, located eastward of the SDC, forecasts present a small upper level layer with a wet bias. The WRFCSU and WRFUIUC evidence a dry upper level layer (550–350 hPa) with a negative Td bias up to −2°C, while the WRFSMN and GFS have a positive bias. The lower levels behavior is similar to Villa Yacanto. For the WRFCSU we observe that as the forecast lead time increases, the midlevel (700–500 hPa) positive bias is smaller (only F 24-h shown). Regarding the temperature, below 800 hPa the bias is either null or warm, whereas above this level the values indicate a general cold bias for all the models. In this case the zonal wind bias is near zero for all the models in the lower levels, where the observed mean circulation is westward (not shown). Between 900 and 800 hPa there is a negative meridional wind bias for all the NWP models, but around 940 hPa it is positive for the WRFCSU and WRFSMN. For midlevels (800–650 hPa) the behavior differs from the previous stations analyzed; the WRFCSU, WRFUIUC and GFS evidence overestimation of the meridional wind up to 1.5 m s^{-1} (underestimating poleward intensity) while for lower levels (910–800 hPa) they show negative values, indicating an overestimation of northern flow.

In summary, at the three sounding sites the NWP major deficiencies appear in the humidity representation. Particularly, the GFS has a persistent moist bias above 500 hPa, surpassing a T bias of 8°C over SDC. For lower levels the WRF NWP models tend to be drier than the GFS. In addition, the overestimation of the northern flow forecasted by the NWP configurations at lower levels at Córdoba can partially explain the overestimation of the 24-h accumulated precipitation in the SDC, because of the role of poleward flows in carrying moisture into the region (Mulholland et al. 2018).

4. Conclusions

In this work, the 24-h accumulated precipitation forecasts at 1200 UTC and the corresponding atmospheric
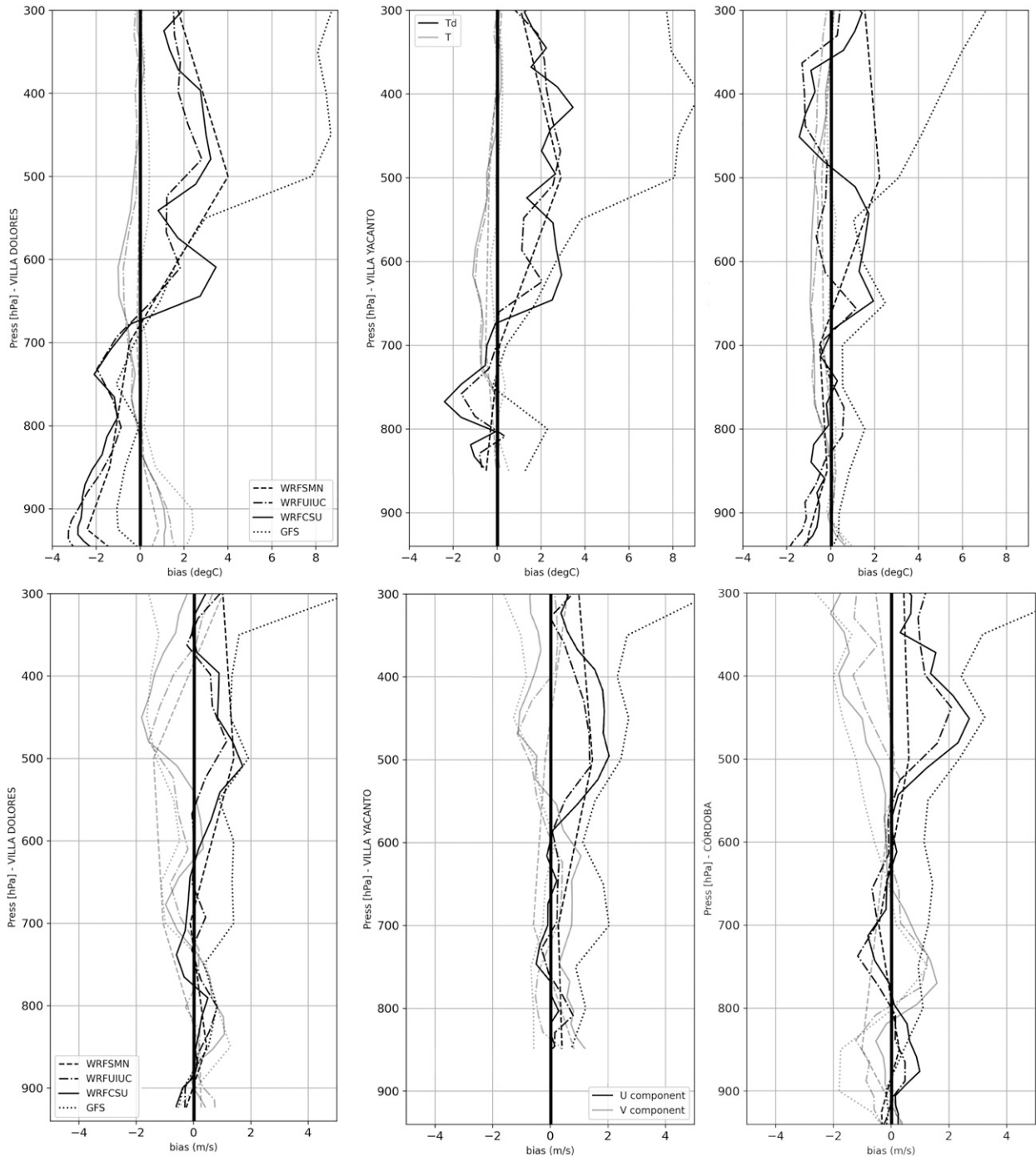
FIG. 16. (top) Mean *T* (gray) and Td (black) forecast bias at various pressure levels (vertical axis) for each NWP model different sounding sites (column) for forecast time F 24-h. (bottom) Mean *u* wind (black) and *v* wind (gray) components forecast bias at various pressure levels (vertical axis) for each NWP model different sounding sites (column) for forecast time F 24-h. (left) Villa Dolores, (center) Villa Yacanto, and (right) Córdoba.

profiles of various NWP model configurations that were used during the RELAMPAGO-CACTI campaign are evaluated, in a domain centered in the SDC using data from 125 surface rain gauge stations and 3 sounding sites.

The evaluation over the entire study domain and in the 43-day common period involves three convection-permitting high-resolution NWP models (4-km WRFSMN, 3-km WRFUIUC, and 4-km WRFCSU) and one global low-resolution NWP model

(0.25° GFS). It is the first time that such a comparison is carried out focusing on the complex terrain of SDC, which represents an important task particularly for this period of the year where storms are frequent and intense.

The results analyzed throughout the study domain suggest that NWP models tend to overestimate precipitation over the SDC orography and underestimate it on the area identified as plain terrain. From the analysis presented, the WRFUIUC model shows the best performance having the smallest wet BIASS for the WRF Model configurations, the ETS score closest to the perfect score for F 36-h and F 48-h forecasts for thresholds greater than 10 mm day^{-1}, and the greatest FSS values. Although the GFS has outstanding values for the BIASS, its performance in the ETS is close to zero for the F 24-h and F 36-h in most thresholds; this NWP model has a much coarser resolution, thereby spreading precipitation out over a larger area and being more accurate according to the metric used. A key finding is that the NWP WRF configurations struggle with the analyzed higher precipitation thresholds, having an increasing wet BIASS and values of FSS beneath 0.5 at large thresholds. It should be noted that the results vary depending on the forecast lead time and statistical score.

Considering the terrain height analysis, for each categorical statistic a different NWP model has the best performance. The WRF configurations evidence a general wet BIASS in both terrain subdivisions, with greater values in the orography class (altitude greater or equal to 1000 m). The GFS generally overestimates accumulated precipitation for thresholds below 20 mm day^{-1} and underestimates it for the largest thresholds in both classes. It also shows the least amount of correctly predicted precipitation fractions considering a random number of hits (ETS). Both the WRFCSU or the WRFSMN obtain the largest number of correct forecasts (POD), while the WRFUIUC and GFS obtain the lowest number of false alarms for F 24-h and F 36-h forecasts, respectively. Last, from the analysis of the volumetric PDFs, we can conclude that the global NWP model underestimates the values above the mode of precipitation rate, and particularly this underestimation is larger in the complex terrain, while the values beneath the mode are overestimated. The PDFs of precipitation rate from the WRF configurations generally show a better adjustment to the observed PDF compared with GFS. The NWP models that stand out for their performance are the WRFCSU and WRFUIUC, although it is worth noting that the latter fails in detecting correctly the most frequently observed threshold.

With the discussed results in mind, the last part of the study evaluates the forecasted atmospheric soundings against the recorded observations at selected RELAMPAGO-CACTI sounding sites. All the NWP WRF configurations show a dry bias at the low levels up to 800 hPa and a wet bias in the upper levels throughout the RELAMPAGO-CACTI field campaign. The GFS generally has the greatest wet bias westward of the SDC, having more available moisture for convective systems than the WRF Model configurations (Piersante et al. 2021). These results influence MCSs and large-scale precipitation events because a dry bias near the surface discourages

convection and a wet bias in upper levels leads to favorable conditions for the development of precipitation events. There is a clear moist bias with respect to the observations in the lower levels of the GFS model which is not seen in the NWP WRF configurations. Regarding the wind bias there is not a general consensus of an outstanding performance of one NWP model. Results evidence that all the NWP models overestimate the zonal wind at the upper levels (650–400 hPa) and underestimate the meridional component, behavior which is observed for all the sounding sites. For Villa Dolores and Villa Yacanto sounding sites the GFS generally has the greatest zonal wind bias, while in midlevels the WRFSMN has the furthest values from null meridional wind bias. For Córdoba an overestimation of the northern flow forecasted by the NWP configurations at lower levels was encountered.

Summarizing, the analysis shows that on the one hand, convection-permitting high-resolution NWP forecasts in the SDC complex terrain area are advantageous over the coarser resolution GFS with parameterized convection, as demonstrated by the 24-h accumulated precipitation fields, PDFs and FSS score. On the other hand, even though the GFS has a lower resolution, which results in having poorer spatial distributions of precipitation, it does have good results similar to some WRF configuration in some of the thresholds of the ETS analyses (except F 24-h).

The overall better performance of the WRFUIUC configuration can be attributed to several characteristics, including different vertical coordinates, the YSU PBL scheme (vs MYJS for the other models), its smaller grid spacing (3 vs 4 km), and the use of a dual-moment microphysical parameterization (Thompson). For example, (Casanovas et al. 2021) and (Rivelli Zea et al. 2021) showed that the SDC has DSDs associated with small drops and smaller sizes, while single moment schemes cannot reproduce this type of DSD configuration at small sizes. Since all of these models were driven by the same analysis, biases in the upper level flow appear to arise from biases in the GFS used as initial and boundary conditions, which can impact forecasts of mesoscale phenomena (Durran and Weyn 2016). At the low levels, one primary difference among these model configurations is the PBL scheme, which may be considered as an important factor in simulating the low level flow and moisture correctly in this region of complex terrain (Liu et al. 2011; Kacan and Levo 2019). For example (Alvarez Imaz et al. 2021) found larger differences in forecasts of both precipitation and atmospheric profiles for PBL schemes than for microphysics parameterizations in convective initiation over SDC. These errors suggest that both improved data assimilation (Dillon et al. 2021) and model physics are necessary for the improvements of precipitation forecasts in this region, and will be the topic of future study. It is worth noting that the implementation of an NWP ensemble would contribute to the improvement of precipitation forecasts, allowing probabilistic approaches (Stensrud et al. 2009). Verification of specific case studies would also be useful, in order to analyze short-term forecasts, for example 6 or 3 h where model spinup is a great challenge.

These results show the importance of convection-permitting high-resolution forecasts in this region, which should be

complementary to the coarser-resolution global model forecasts to help various users and decision makers.

The main outcomes of this work are being transferred to the Córdoba provincial authorities and will allow an improved management of the drained water volumes (operating valves, gates, etc.) in the different reservoirs within the study area, preserving valuable water resources, lives, and property in the basin. One of the authors of this work has generated an optimized meteorological information transfer tool that allows technicians from provincial water resources management agencies to access high-resolution (WRFSMN and WRFUIUC) and low-resolution (GFS) forecasts of mean precipitation values every 24 h by basin. This information is considered, in a first instance, as input data to the hydrological models, allowing to predict the evolution of reservoir levels and discharges.

Data availability statement. Global Forecast System data are available at https://rda.ucar.edu/datasets/ds084.1/#!description. WRF outputs are stored on the different institutions' servers (CSU, UIUC, and Argentinian National Meteorological Service) and are available upon request. Surface stations data are stored in the Argentinian National Meteorological Service server and are available upon request; and can also be accessed through the RELAMPAGO Data Archive (https://data.eol.ucar.edu/dataset/553.040).

REFERENCES

Alvarez Imaz, M., P. Salio, M. E. Dillon, and L. Fita, 2021: The role of atmospheric forcings and WRF physical set-up on convective initiation over Córdoba, *Argentina. Atmos. Res.*, **250**, 105335, https://doi.org/10.1016/j.atmosres.2020.105335.

Amitai, E., W. Petersen, X. Llort, and S. Vasiloff, 2012: Multiplatform comparisons of rain intensity for extreme precipitation events. *IEEE Trans. Geosci. Remote Sens.*, **50**, 675–686, https://doi.org/10.1109/TGRS.2011.2162737

Baldauf, M., A. Seifert, J. Forstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and

sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1

Barthlott, C., U. Corsmeier, C. Meißner, F. Braun, and C. Kottmeier, 2006: The influence of mesoscale circulation systems on triggering convective cells over complex terrain. *Atmos. Res.*, **81**, 150–175, https://doi.org/10.1016/j.atmosres.2005.11.010.

Buehner, M., J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes Geophys.*, **20**, 669–682, https://doi.org/10.5194/npg-20-669-2013.

Casanovas, C., P. Salio, V. Galligani, B. Dolan, and S. W. Nesbitt, 2021: Drop size distribution variability in Central Argentina during RELAMPAGO-CACTI. *Remote Sensing*, **13**, 2026, https://doi.org/10.3390/rs13112026.

Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, https://doi.org/10.1002/met.1538.

Dillon, M. E., and Coauthors, 2016: Application of the WRF-LETKF data assimilation system over southern South America: Sensitivity to model physics. *Wea. Forecasting*, **31**, 217–236, https://doi.org/10.1175/WAF-D-14-00157.1.

——, and Coauthors, 2021: A Rapid Refresh ensemble based data assimilation and forecast system for the RELAMPAGO field campaign. *Atmos. Res.*, **264**, 105858, https://doi.org/10.1016/j.atmosres.2021.105858.

Durran, D. R., and J. A. Weyn, 2016: Thunderstorms do not get butterflies. *Bull. Amer. Meteor. Soc.*, **97**, 237–243, https://doi.org/10.1175/BAMS-D-15-00070.1.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24**, 1498–1510, https://doi.org/10.1175/2009WAF2222251.1.

García Skabar, Y., C. Matsudo, M. P. Hobouchian, M. Sacco, J. Ruiz, and S. Righetti, 2018: Implementación del modelo WRF en alta resolución en el Servicio Meteorológico Nacional. *CONGREMET XIII*, Rosario, Argentina, CAM, http://cenamet.org.ar/congremet/wp-content/uploads/2018/11/T0135_GARCÍASKABAR.pdf.

Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, https://doi.org/10.1256/qj.06.25.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

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. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122≤0927:TSMECM≥2.0.CO;2.

Jones, B., 2014: How does the skill of global model precipitation forecasts over Europe depend on spatial scale? University of Reading, 88 pp., http://www.met.reading.ac.uk/~sws00rsp/teaching/postgrad/jones.pdf.

Kacan, K. G., and Z. J. Levo, 2019: Microphysical and dynamical effects of mixed-phase hydrometeors in convective storms using a bin microphysics model: Melting. *Mon. Wea. Rev.*, **147**, 4437–4460, https://doi.org/10.1175/MWR-D-18-0032.1.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, https://doi.org/10.1175/WAF2007106.1.

Kumjian, M. R., and Coauthors, 2020: Gargantuan hail in Argentina. *Bull. Amer. Meteor. Soc.*, **101**, E1241–E1258, https://doi.org/10.1175/BAMS-D-19-0012.1.

Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado headwaters region: Sensitivity to physics parameterizations. *Mon. Wea. Rev.*, **139**, 3533–3553, https://doi.org/10.1175/MWR-D-11-00009.1.

Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2017: Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. *Wea. Forecasting*, **32**, 1403–1421, https://doi.org/10.1175/WAF-D-16-0200.1.

Mahoney, K. M., 2016: The representation of cumulus convection in high-resolution simulations of the 2013 Colorado Front Range flood. *Mon. Wea. Rev.*, **144**, 4265–4278, https://doi.org/10.1175/MWR-D-16-0211.1.

Matsudo, C., and P. Salio, 2011: Severe weather reports and proximity to deep convection over Northern Argentina. *Atmos. Res.*, **100**, 523–537, https://doi.org/10.1016/j.atmosres.2010.11.004.

——, Y. García Skabar, J. Ruiz, L. Vidal, and P. Salio, 2015: Verification of WRF-ARW convective-resolving forecasts over southern South America. *Mausam*, **66**, 445–4567.

McBride, J. L., and E. E. Ebert, 2000: Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia. *Wea. Forecasting*, **15**, 103–121, https://doi.org/10.1175/1520-0434(2000)015≤0103:VOQPFF≥2.0.CO;2.

Mezher, R. N., M. Doyle, and V. Barros, 2012: Climatology of hail in Argentina. *Atmos. Res.*, **114–115**, 70–82, https://doi.org/10.1016/j.atmosres.2012.05.020.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, https://doi.org/10.1175/2008MWR2556.1.

Mulholland, J. P., S. W. Nesbitt, R. J. Trapp, K. L. Rasmussen, and P. V. Salio, 2018: Convective storm life cycle and environments near the Sierras de Córdoba, Argentina. *Mon. Wea. Rev.*, **146**, 2541–2557, https://doi.org/10.1175/MWR-D-18-0081.1.

——, ——, and ——, 2019: A case study of terrain influences on upscale convective growth of a supercell. *Mon. Wea. Rev.*, **147**, 4305–4324, https://doi.org/10.1175/MWR-D-19-0099.1.

Nesbitt, S. W., and Coauthors, 2021: A storm safari in subtropical South America: Proyecto RELAMPAGO. *Bull. Amer. Meteor. Soc.*, **102**, E1621–E1644, https://doi.org/10.1175/BAMS-D-20-0029.1.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Penalba, O. C., and W. M. Vargas, 2004: Interdecadal and interannual variations of annual and extreme precipitation over central-northeastern Argentina. *Int. J. Climatol.*, **24**, 1565–1580, https://doi.org/10.1002/joc.1069.

Piersante, O. J., R. S. Schumacher, and K. L. Rasmussen, 2021: Comparison of biases in warm-season WRF forecasts in North and South America. *Wea. Forecasting*, **36**, 979–1001, https://doi.org/10.1175/WAF-D-20-0062.1.

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for warn-on-forecast. *Mon. Wea. Rev.*, **143**, 2998–3024, https://doi.org/10.1175/MWR-D-14-00416.1.

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Rasmussen, K. L., M. D. Zuluaga, and R. A. Houze Jr., 2014: Severe convection and lightning in subtropical South America. *Geophys. Res. Lett.*, **41**, 7359–7366, https://doi.org/10.1002/2014GL061767.

Rivelli Zea, L., S. W. Nesbitt, A. Ladino, J. C. Hardin, and A. Varble, 2021: Raindrop size spectrum in deep convective regions of the Americas. *Atmosphere*, **12**, 979, https://doi.org/10.3390/atmos12080979.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Robinson, T. E., and S. Businger, 2019: A novel method for modeling lowest-level vertical motion. *Wea. Forecasting*, **34**, 943–957, https://doi.org/10.1175/WAF-D-18-0064.1.

Salio, P., M. Nicolini, and E. J. Zipser, 2007: Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet. *Mon. Wea. Rev.*, **135**, 1290–1309, https://doi.org/10.1175/MWR3305.1.

Schumacher, R. S., and Coauthors, 2021: Convective-storm environments in subtropical South America from high-frequency soundings during RELAMPAGO-CACTI. *Mon. Wea. Rev.*, **149**, 1439–1458, https://doi.org/10.1175/MWR-D-20-0293.1.

Schwartz, C. S., 2014: Reproducing the September 2013 record-breaking rainfall over the Colorado Front Range with high-resolution WRF forecasts. *Wea. Forecasting*, **29**, 393–402, https://doi.org/10.1175/WAF-D-13-00136.1.

Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Skok, G., and N. Roberts, 2016: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts. *Quart. J. Roy. Meteor. Soc.*, **142**, 2599–2610, https://doi.org/10.1002/qj.2849.

Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, https://doi.org/10.1175/WAF-D-10-05046.1.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

Takemi, T., 2018: Importance of terrain representation in simulating a stationary convective system for the July 2017 northern Kyushu heavy rainfall case. *SOLA*, **14**, 153–158, https://doi.org/10.2151/sola.2018-027.

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. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Varble, A., and Coauthors, 2021: Utilizing a storm-generating hotspot to study convective cloud transitions: The CACTI experiment. *Bull. Amer. Meteor. Soc.*, **102**, E1597–E1620, https://doi.org/10.1175/BAMS-D-20-0030.1.

Vila, D. A., L. A. Toledo Machado, H. Laurent, and I. Velasco, 2008: Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) using satellite infrared imagery: Methodology and validation. *Wea. Forecasting*, **23**, 233–245, https://doi.org/10.1175/2007WAF2006121.1.

Virtanen, P., and Coauthors, 2020: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261–272, https://doi.org/10.1038/s41592-019-0686-2.

Ward, E., W. Buytaert, L. Peaver, and H. Wheater, 2011: Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Adv. Water Resour.*, **34**, 1222–1231, https://doi.org/10.1016/j.advwatres.2011.05.007.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437, https://doi.org/10.1175/2007WAF2007005.1.

Whiteman, C., 2000: *Mountain Meteorology: Fundamentals and Applications.* Oxford University Press, 376 pp.

Yáñez-Morroni, G., J. Gironás, M. Caneo, R. Delgado, and R. Garreaud, 2018: Using the Weather Research and Forecasting (WRF) model for precipitation forecasting in an Andean region with complex topography. *Atmosphere*, **9**, 304, https://doi.org/10.3390/atmos9080304.

Zhu, K., and Coauthors, 2018: Evaluation of real-time convection-permitting precipitation forecasts in China during the 2013–2014 summer season. *J. Geophys. Res. Atmos.*, **123**, 1037–1064, https://doi.org/10.1002/2017JD027445.

Zipser, E. J., D. J. Cecil, C. Liu, S. W. Nesbitt, and D. P. Yorty, 2006: Where are the most intense thunderstorms on Earth? *Bull. Amer. Meteor. Soc.*, **87**, 1057–1072, https://doi.org/10.1175/BAMS-87-8-1057.