Evaluation of WRF Model Forecasts and Their Use for Hydroclimate Monitoring over Southern South America

OMAR V. MÜLLER AND MIGUEL A. LOVINO

Centro de Estudios de Variabilidad y Cambio Climático, Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral and CONICET, Santa Fe, Argentina

ERNESTO H. BERBERY

Earth System Science Interdisciplinary Center, Cooperative Institute for Climate and Satellites, University of Maryland, College Park, College Park, Maryland

ABSTRACT

Weather forecasting and monitoring systems based on regional models are becoming increasingly relevant for decision support in agriculture and water management. This work evaluates the predictive and monitoring capabilities of a system based on WRF Model simulations at 15-km grid spacing over the La Plata basin (LPB) in southern South America, where agriculture and water resources are essential. The model's skill up to a lead time of 7 days is evaluated with daily precipitation and 2-m temperature in situ observations for the 2-yr period from 1 August 2012 to 31 July 2014. Results show high prediction performance with 7-day lead time throughout the domain and particularly over LPB, where about 70% of rain and no-rain days are correctly predicted. Also, the probability of detection of rain days is above 80% in humid regions. Temperature observations and forecasts are highly correlated ($r > 0.80$) while mean absolute errors, even at the maximum lead time, remain below 2.7°C for minimum and mean temperatures and below 3.7°C for maximum temperatures. The usefulness of WRF products for hydroclimate monitoring was tested for an unprecedented drought in southern Brazil and for a slightly above normal precipitation season in northeastern Argentina. In both cases the model products reproduce the observed precipitation conditions with consistent impacts on soil moisture, evapotranspiration, and runoff. This evaluation validates the model's usefulness for forecasting weather up to 1 week in advance and for monitoring climate conditions in real time. The scores suggest that the forecast lead time can be extended into a second week, while bias correction methods can reduce some of the systematic errors.

1. Introduction

The Global Framework for Climate Services (GFCS) established by the WMO provides guidance for decision-making for agriculture and food security, water, health, and disaster risk reduction (WMO 2014). Within this context, six countries in southeastern South America have established a regional climate center (RCC-SSA) that involves a strong collaboration between weather services and academic institutions. The RCC-SSA's priorities are on climate monitoring and weather forecasting, two areas that benefit from the use of regional models. Regional modeling is relevant at local scales, where global models may not capture essential features and where the choice of alternative parameterizations may help the model perform better for the region. Specific real-time output variables and diagnostics needed by stakeholders are easily obtained from regional models when they are not provided by the global forecast centers. Forecasting and monitoring systems linked with appropriate decision support tools can substantially improve real-time choices needed for agricultural and water management (Stone and Meinke 2005). Because of their relative simplicity of use, regional models are useful for performing real-time forecasts; hence, verification of the results is an indispensable component for identifying a system's strengths and weaknesses. Continuous verification is also desired to keep track of changes in

---

the forecasts' quality because of updates either to the regional model or the global model that provides the initial conditions.

The region of interest for this research is the La Plata basin (LPB) in southeastern South America, a region that, because of its socioeconomic relevance, has been the subject of many hydroclimate studies in recent years (Berbery and Barros 2002; Sanchez et al. 2015 and references therein). The LPB is a large expanse over diverse climate regimes, from a monsoon climate toward its northern region to dominating mesoscale convective systems (MCSs) in the subtropical central and southern regions, and to midlatitude weather systems farther south. There are three important challenges for weather and climate modeling in the LPB. First is the ability to represent the intense and frequent MCSs and convective storms that develop in the region (Zipser et al. 2006). There is a general consensus that because of the difficulty in simulating these convective scale events the models underestimate the mean precipitation and yet overestimate the number of precipitation events (see, e.g., Rauscher et al. 2007; Menéndez et al. 2010). Second, the presence of the steep Andes Mountains to the west is poorly handled in the models, resulting in temperature and precipitation biases not only over and near the mountains but also downstream (Seluchi et al. 1998; Saulo et al. 2001; Chou et al. 2005; Müller et al. 2014). Finally, an accurate representation of the South American low-level jet and its associated moisture flux convergence is required to reproduce the precipitation spatial structure and diurnal cycle (see, e.g., Berbery and Collini 2000; Collini et al. 2008).

The current regional models used in weather services in southern South America are the Eta Model, the Brazilian Regional Atmospheric Modeling System (BRAMS), and the Weather Research and Forecasting (WRF) Model. The Eta Model centered over Brazil is used for hourly to seasonal predictions (Seluchi and Chou 2001; Chou et al. 2005; de Goncalves et al. 2006). According to de Moura et al. (2010), the short-term rainfall forecasts achieve good performance for up to 120 h of simulation, although a tendency to overestimate 24-h forecasts in northern Brazil was found. Short-term forecasts with the Eta Model at 0.25° grid spacing are also produced over Argentina, where the model tends to have too many false alarms, that is, to overforecast the frequency of rainfall events (Suaya 2004). High-resolution versions of the BRAMS and WRF Model are also used at the weather services on an experimental basis. García Skabar et al. (2011) and Dillon et al. (2013) report that the WRF shows greater skill, although both models tend to underestimate the average precipitation.

Still, they better represent the intense precipitation that fails to be detected by the lower-resolution (0.25° grid spacing) Eta Model. Other short-term forecasts are systematically performed at academic institutions serving the needs of the local community. As with the models run in operational environments, these simulations present systematic errors like the overestimation of the frequency of low-intensity precipitation events and the underestimation of heavy rainfall events (see, .e.g., Saulo et al. 2008; García Skabar et al. 2012).

Evaluations of temperature forecasts have shown that models exaggerate high temperatures in the subtropics during the warm season, while negative biases develop during the cold months throughout the region (Menéndez et al. 2010), leading to a larger-amplitude annual cycle. Near-surface temperatures in climate simulations with the WRF Model have a high resemblance with observations away from the Andes Mountains, with biases smaller than about 2°C in magnitude (Müller et al. 2014). Warm biases are found toward the middle of the continent, and cold biases are noticed toward the east and near the coastlines. The systematic and nonsystematic temperature errors are sensitive to different parameterizations in the corresponding land surface models (Ruiz et al. 2010; Müller et al. 2014; Pei et al. 2014).

The first objective of this work is to evaluate the predictive and monitoring capabilities of a real-time forecast system based on numerical simulations with the WRF Model in a region of high agricultural productivity and water needs. Precipitation and 2-m temperature forecast skill levels are evaluated during a period of 2 yr as a preliminary stage before continuous evaluation. The emphasis on precipitation and temperature follows the approach taken at the RCC-SSA, whose priority is to develop datasets of those two variables. Other variables, like wind and humidity, were not included at this stage of the evaluation. The second objective of this research is to explore the performance of the WRF Model products when monitoring the recent hydroclimate. This is tested for an intense drought with high socioeconomic impacts in southern Brazil and a slightly above normal precipitation season in northeastern Argentina. This research focuses on applications that are being transitioned to the Regional Climate Center for Southern South America to complement their own products. Section 2 presents the model configuration, output, and observational datasets. Sections 3 and 4 describe the verification methods and discuss their application to precipitation and temperature forecasts. Section 5 describes the model's capabilities for hydroclimate monitoring. A summary and our conclusions are offered in section 6.
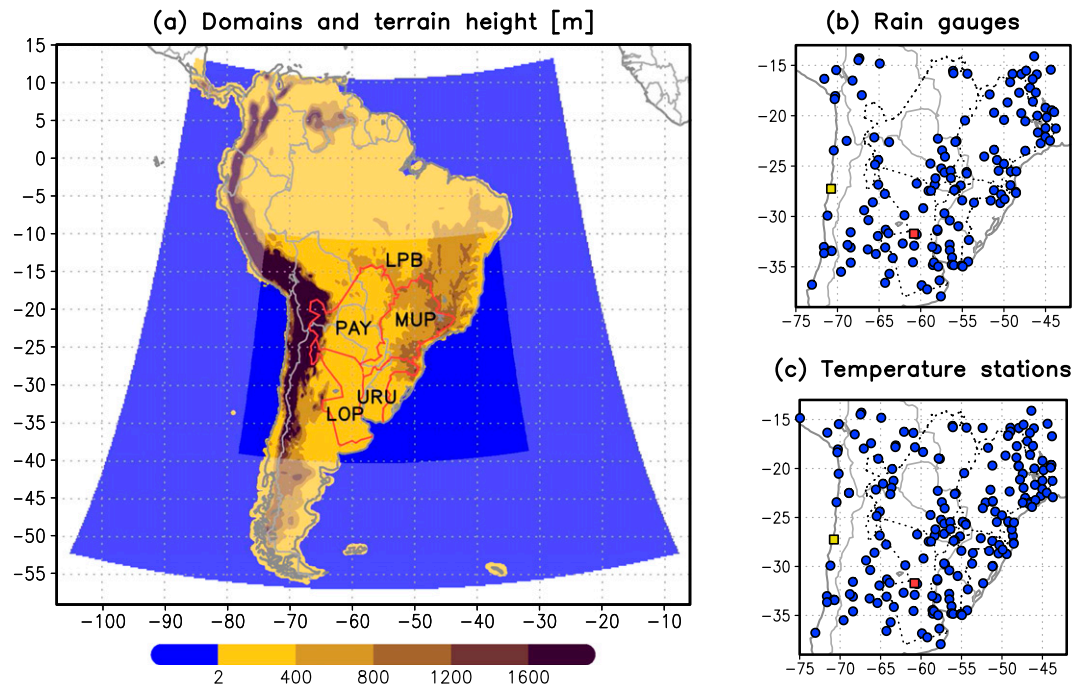
FIG. 1. (a) Model domains and terrain height. The largest domain is shown with lighter colors while the nested domain is highlighted with stronger colors. Red contours denote the LPB and its subbasins: Paraguay (PAY), mid- to upper Paraná (MUP), lower Paraná (LOP), and Uruguay (URU). (b) Rain gauges and (c) temperature stations used to assess WRF Model forecasts. The gauge stations Sauce Viejo and Desierto de Atacama are highlighted with red and yellow squares, respectively.

2. Model and datasets

This study evaluates the WRF Model forecast skill in southern South America focusing on the LPB and the subbasins of the main rivers of the region: the Paraná, Paraguay, and Uruguay (see Fig. 1a). The Paraná River basin is divided into the mid- to upper and the lower Paraná, as each is impacted by different climate regimes. The forecast evaluation is done on precipitation and temperature at 2 m, as these variables are most commonly used for observations. The period of evaluation covers 2 yr, from 1 August 2012 to 31 July 2014.

The system is based on daily runs of the WRF-ARW model, version 3.1, and its evaluation is a crucial aspect in the reliability of the system. The model's parent domain covers South America with a grid spacing of 45 km, and the LPB with a nested domain at 15-km grid spacing (see Fig. 1a). The initial and 6-h boundary conditions are obtained from the Global Forecast System (GFS) (Environmental Modeling Center 2003; Campana and Caplan 2006; Werth and Garrett 2011; see also Saha et al. 2014). The GFS dataset, obtained daily from the NOAA/NCEP servers, includes gridded variables needed to run WRF, such as sea surface temperature and soil moisture, among others. The lateral boundary conditions are updated at 6-h intervals. The use of 3-h

intervals was not considered since this time span would have resulted in few, if any, advantages (see Dimitrijevic and Laprise 2005 and references therein). The model physics configuration is based on the selection of schemes and options tested in Lee (2010) and Lee and Berbery (2012), who evaluated an optimal combination that represents South America's climate with smaller biases. Further evaluation was carried out by Müller et al. (2014) for long-term simulations. A summary of the model configuration and parameterizations is presented in Table 1.

The evaluations discussed in this article are based on in situ observations. The observation datasets include information obtained from the former NCDC [now part of the National Centers for Environmental Information (NCEI)], which collects daily rain gauge data and temperature observations from over 9000 worldwide stations. Rain gauge data from Brazil's Instituto Nacional de Meteorologia (INMET) were also included to improve the data coverage. To ensure statistical significance, the selected gauges are those that have measured on 90% or more of the total days (657 days out of 730 possible days). A total of 142 rain gauges and 193 temperature stations meet the minimum threshold in the model's inner domain. Of those, 80 and 100, respectively, are within the LPB. The in situ networks of

TABLE 1. Summary of WRF configuration.

Parameter	Parent domain	Nested domain
Region	South America	LPB
Grid resolution	45 km	15 km
Grid size	502 × 367 grid points	709 × 475 grid points
No. of vertical levels	28	
Period	168 h (7 days)	
Integration time step	240 s	
Dynamic solver	ARW	
Boundary conditions	GFS	Parent domain
Microphysics	Eta Model microphysics (Rogers et al. 2001)	
Cumulus convection	Betts–Miller–Janjić scheme (Janjić 1994, 2000)	
Surface layer	Janjić Eta Model scheme (Janjić 1996)	
Land surface model	Noah LSM (Chen and Dudhia 2001)	
Land cover classification	USGS	
Planet boundary layer	Mellor–Yamada–Janjić scheme (Janjić 1994)	
Shortwave radiation	Dudhia scheme (Dudhia 1989)	
Longwave radiation	Rapid Radiative Transfer Model (Mlawer et al. 1997)	

precipitation and temperature are shown in Figs. 1b and 1c, respectively. Note the areas with scarce observations particularly toward the northwestern sector. The use of known gridded observational datasets was avoided in the evaluation because they include many but not all of the observations we used. Moreover, the interpolation involves an amount of smoothing that can make gridded data deviate from the in situ observations. The evaluation based on in situ observations is quantitative, although it comes at the cost of losing spatially continuous representation of the patterns.

As shown in Table 1, the simulations predict the weather for 168 h (7-day lead time). Each lead day is evaluated against the observations of the corresponding day. In all cases, station data are compared to the forecast for the nearest grid point. More complex approaches were tested, but no obvious advantage was found over the nearest-gridpoint approach, despite its simplicity. These results are consistent with the findings of Pappenberger et al. (2009).

3. Evaluation of precipitation forecasts

a. Statistical approach

The precipitation is evaluated in terms of occurrence/ no occurrence of daily events. When observed precipitation for a given day exceeds a preset rainfall threshold, it is considered a rainy day. Once the days are classified as rain or no rain for both the observations and forecasts, contingency tables are prepared (Wilks 2011). A contingency table identifies the four possible combinations of forecasts (rain/no rain) and observations (rain/no rain) and is called the joint distribution (see Table 2). The combinations are hit (h), where an event is

forecast to occur and does occur; miss (m), where an event is forecast not to occur but does occur; false alarm (fa), where an event is forecast to occur but does not occur; and correct negative (cn), where an event is forecast not to occur and does not occur.

The model's skill in forecast precipitation is measured from statistics (usually called skill scores) that are computed from the elements in the contingency table (Hogan and Mason 2012, and references therein). Five skill scores or categorical statistics are defined and summarized in Table 3: accuracy (ACC), probability of detection (POD), false alarm ratio (FAR), Heidke skill score (HSS), and frequency bias (BIAS). The accuracy measures the forecast's skill in identifying a day as rain or no rain. The probability of detection measures the ability to predict a rain day, ignoring false alarms. POD is complemented with the FAR, which counts no-rain days that are forecast as rain days. The bias score (or frequency bias) is the ratio between the number of correct forecasts and the number of cases of observed precipitation. A bias score that is less than one indicates that the forecast system underforecasts events, and if it is larger than one, it overforecasts them. Finally, the Heidke skill score measures the fraction of correct forecasts that did not occur by random chance. If the HSS score becomes negative, then the random forecast is better than the actual forecast, which is said to be

TABLE 2. Contingency table.

	Observed yes	Observed no	Total
Forecast yes	Hits (h)	False alarms (fa)	Forecast yes
Forecast no	Misses (m)	Correct negatives (cn)	Forecast no
Total	Observed yes	Observed no	Total

TABLE 3. Skill scores used to evaluate daily precipitation events.

Statistics	Formula	Definition	Range
Precipitation accuracy	$\mathrm{ACC} = \dfrac{h + \mathrm{cn}}{\mathrm{total}}$	The fraction of correct forecasts	0–1, with 1 being a perfect score
Probability of detection	$\mathrm{POD} = \dfrac{h}{h + m}$	The fraction of the observed "yes" events that were correctly forecast	0–1, with 1 being a perfect score
False alarm ratio	$\mathrm{FAR} = \dfrac{\mathrm{fa}}{\mathrm{fa} + h}$	The fraction of predicted positive events that actually did not occur	0–1, with 0 being a perfect score
Heidke skill score	$\mathrm{HSS} = \dfrac{h + \mathrm{cn} - \mathrm{hr}}{\mathrm{total} - \mathrm{hr}}, \quad$ where $\mathrm{hr} = \dfrac{(h + m)(h + \mathrm{fa}) + (\mathrm{cn} + m)(\mathrm{cn} + \mathrm{fa})}{\mathrm{total}}$	No. of correct forecasts that did not occur by chance	From $-\infty$ to 1, with 1 being a perfect score
Bias score	$\mathrm{BIAS} = \dfrac{h + \mathrm{fa}}{h + m}$	The ratio of the frequency of forecast events to the frequency of observed events.	From $-\infty$ to ∞; 1 is a perfect score; BIAS < 1 underforecast; BIAS > 1 overforecast

unskilled. HSS is an alternative to the equitable threat score (ETS; Hamill 1999). As Hogan et al. (2010) argue, ETS does not satisfy the requirements to be considered "equitable" (Gandin and Murphy 1992).

These scores are convenient for station data because of their ease of interpretation (Tartaglione 2010). One caveat is that when a forecast storm is shifted in space or time with respect to the observed event, the point-to-point evaluation may generate misses and false alarms in the nonoverlapping sectors, resulting in a double penalty (Kok et al. 2008; Rossa et al. 2008).

b. Results

Contingency tables for daily precipitation events (not shown) were computed for all stations in the domain. The skill scores were then computed for each station and each forecast lead time. The rainfall threshold used to define "rain days" was set to $0.25\,\mathrm{mm\,day}^{-1}$ to agree with the minimum measurable precipitation of most bucket-type rain gauges. The methodology is presented first for one sample station near the center of the domain (Sauce Viejo, 31.7°S, 60.82°W; see Fig. 1b). The results for all stations are then summarized in graphical form using maps and skill score graphics.

Figure 2 summarizes the skill scores changes with the forecast lead time for the sample station. It shows that the precipitation forecasts have an accuracy score of 0.80 for day 1, and then it decays slightly to 0.72 at the maximum lead time (day 7). The probability of detection also exhibits a slight decrease as lead time increases, varying from 0.73 on day 1 to 0.59 on day 7. In other words, observed rain days are predicted (hits) in about 73% of the cases on day 1, slightly decreasing to roughly 60% on day 7. The false alarm ratio is smaller

than 0.5 for days 1–3, meaning that false alarms are fewer than the number of hits. From day 4 onward, there are more false alarms than hits. At the same time, the bias score ranges from 1.39 to 1.60, indicating that the model tends to overforecast events (BIAS > 1), consistent with the high FAR values. The bias score (see equation in Table 3) could increase or decrease with lead time as both false alarms and misses tend to rise at different rates. For the sample station, the bias score is larger for days 5 and 6 than it is for day 7. Information from the set of contingency tables (not shown) suggests that this is due to an increase in false alarms on days 5 and 6, while the values of misses remain in the same range as on days 1–4. The Heidke skill score ranges from 0.48 on day 1 to 0.30 on day 7, indicating that at all times there is a higher rate of success of the model forecasts over the random forecasts. In general, the successes (ACC, POD, and HSS) tend to decrease and failures (FAR) tend to rise as the forecast time increases, as expected.

The results presented so far are for one sample station. A similar approach was followed with all 142 rain gauge stations in the domain, and the results are summarized in Fig. 3. The observed precipitation frequency (Fig. 3a) is defined as the ratio between the number of rain days and the total number of days in the record. It has a west–east gradient, with the highest values over the eastern sector of the LPB, where it rains on one out of three days and has a minimum frequency over the Andes Mountains and the Atacama Desert. The latter is the driest region in South America, where it rains just a few days per year.

Figures 3b–f summarize the skill scores averaged over the 7-day lead time. In other words, $S_m = (1/7)\sum_{i=1}^{7} S_i$, where S_i is the skill score for lead time i and S_m is the
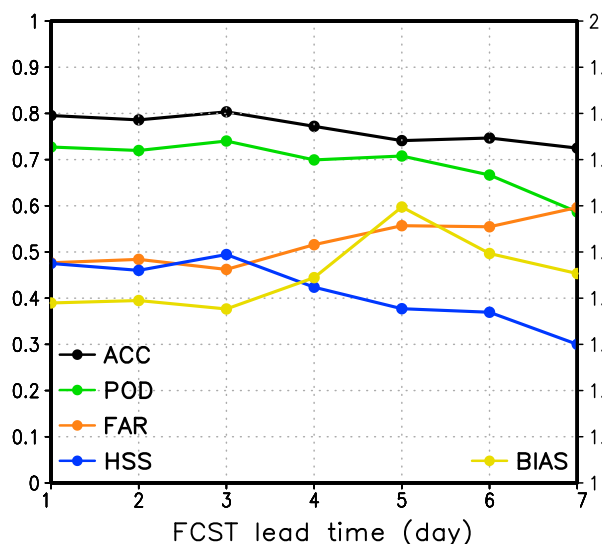
FIG. 2. Evolution of the precipitation skill scores for the 7-day lead times for the sample station Sauce Viejo. The skill scores are ACC, POD, FAR, HSS, and BIAS. The values along the left axis are for the first four scores. BIAS values are indicated along the right axis.

average of those seven scores. The accuracy (Fig. 3b) is almost homogeneous over the LPB with values ranging from 0.6 to 0.8. The best scores are found around the southern subbasins. The probability of detection scores (Fig. 3c) are in a range from 0.6 toward the southwest to 0.9 in the mid- to upper Paraná subbasin. All false alarm ratio scores (Fig. 2d) are above 0.3 and may exceed 0.6 in mountain-dominated areas. Similar WRF high false alarm ratios have been reported in other regions of the world (see, e.g., Weisman et al. 2008; Dravitzki and McGregor 2011). Two issues could contribute to the large values. First, the model tends to overforecast rain days, as seen in Fig. 3e for the bias scores larger than one; the second issue is the already-discussed double counting of false alarms and misses that result when a storm is shifted in space or time. The Heidke skill score (Fig. 3f) is in the range 0.2–0.5, indicating that the forecast is better than a simple random forecast over all of the domain. The equitable threat score (not shown) ranges from 0.15 to 0.4 in the LPB for days 1–7. These values compare favorably with other model evaluations for the same region, which yielded similar scores but in shorter lead times (García Skabar et al. 2012; Dillon et al. 2013).

The evolution with forecast lead time of the skill scores averaged for the 80 stations within the LPB is presented in Fig. 4a. The precipitation accuracy shows almost no changes with lead time, with values in general above 0.70. That is, the model succeeds in forecasting at

least 70% of the occurrence of rain or no-rain days for the LPB. The probability of detection shows a slight decrease with time, from 81% on day 1 to about 68% on day 7. The false alarm ratio score indicates that about half of the forecast rain days did not occur. This is also reflected in Fig. 2 for the sample station. The HSS averages 0.42 for the first day and 0.30 for day 7 of the forecasts. The values of HSS are always notably higher than for a random forecast (HSS below 0). The shaded bands for each curve in Fig. 4a represent the standard deviation as a measure of the spatial spread of the scores.

Dry and wet regions have distinct features in most statistics. Dry regions (low frequency of precipitation in Fig. 3a) tend to have lower probabilities of detection and false alarm ratios than other regions. The probability of detection of the few rain days in the dry region is below 0.7 (Fig. 3c), and the false alarm ratio is above 0.6 (Fig. 3d). Both lower scores are linked to the small number of rain days in the region. For instance, the Desierto de Atacama rain gauge station (27.25°S, 70.76°W; see Fig. 1b) reported 31 rain days (out of 707 days), while the model forecasted 39 rain days at 1-day lead time. While these seem to be similar values, the forecasts matched the rain days in only four cases. Interestingly, the high accuracy in the dry region, with scores above 0.8 (Fig. 3b), is a direct consequence of the high number of correct negatives in an arid region and is therefore of limited value. In brief, while the model forecasted a similar small number of rain days as in the observations (leading to high accuracy), seldom did they match in time, leading to a low probability of detection and a high false alarm ratio. Over wet regions like the upper Paraná basin in the northeastern sector (high frequency of precipitation in Fig. 3a), the probability of detection of rain days exceeds 0.8 (Fig. 3c). The false alarm ratio is lower than 0.4 (Fig. 3d), while the accuracy values are at or below 0.7 (Fig. 3b). This behavior is expected in a wet region where it rains at least one out of three days, and thus correct negatives are fewer than in dry regions. Despite the slight reduction in the accuracy in the wet region, the other scores exhibit more skill than in the dry region.

The time evolution of the area-averaged scores (Fig. 4b) indicates that the model had good performance for all scores, with small but distinct differences between the wet and dry regions. The evolution of skill scores reveals that the model performs better over wet than over dry regions, as estimated by the probability of detection (16% higher on average), the false alarm ratio (14% smaller on average), and Heidke skill score (7% higher on average). The lower scores in the dry region are due to the difficulties in forecasting the few
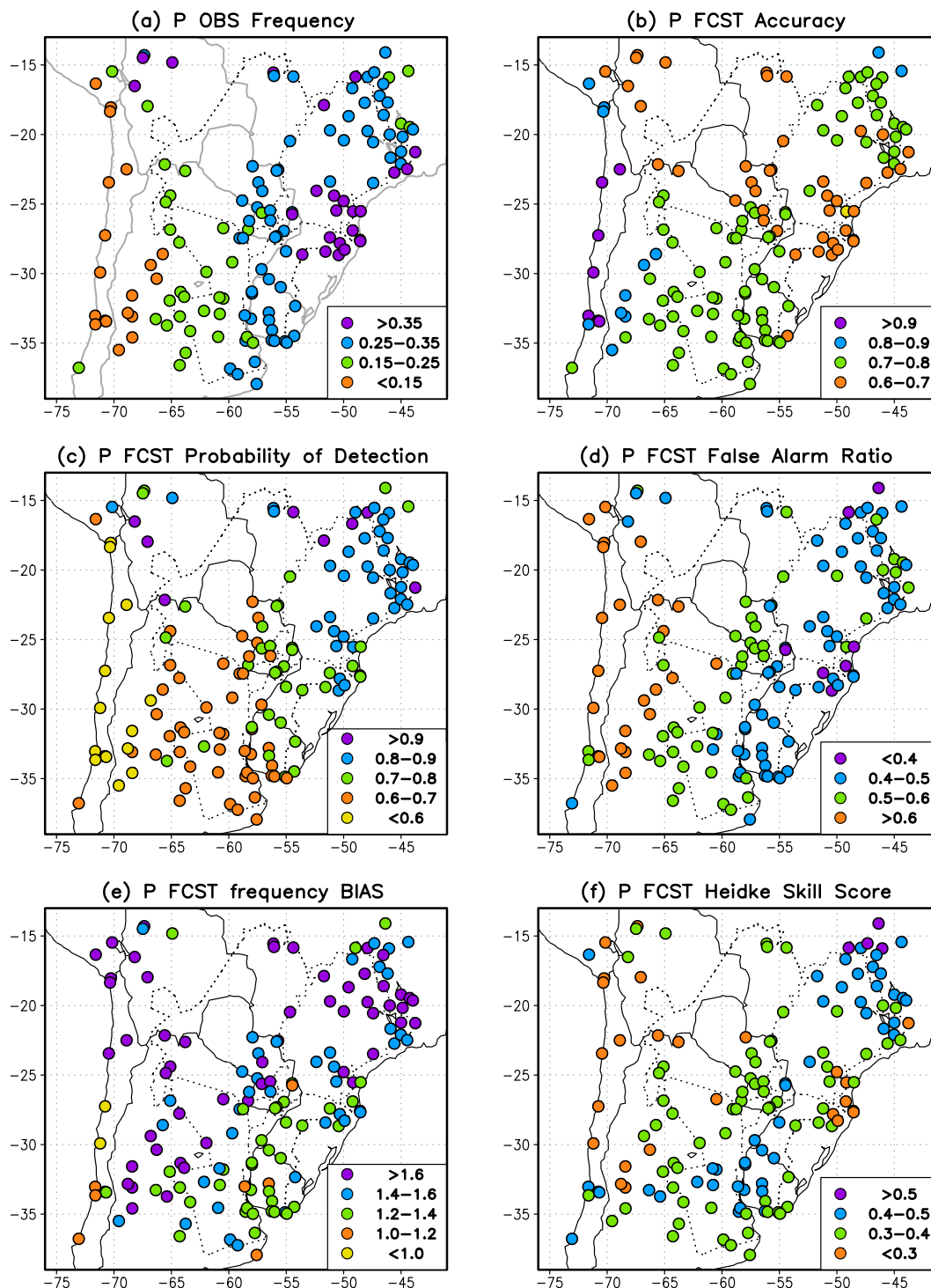
FIG. 3. (a) Geographical distribution of the observed precipitation frequency, followed by the forecast precipitation skill scores averaged over the seven lead times: (b) ACC, (c) POD, (d) FAR, (e) BIAS, and (f) HSS.

rain days that occur. The only score that is better over the dry region is the accuracy (on average 7% higher than in the wet region), largely because of the rise in correct negatives.

Despite the diversity of model versions and parameterizations, the ranges of the skill scores are similar to those found over North America even for individual cases (e.g., Pennelly et al. 2014; Wolff et al. 2014). The

(a) P skill for LPB

(b) P skill for wet/dry climate
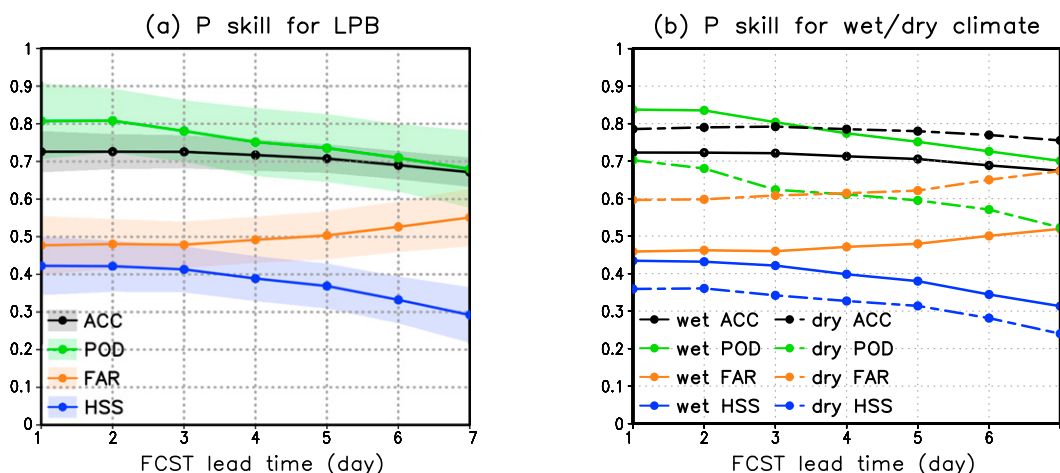
FIG. 4. Evolution with lead time of the area-averaged precipitation skill scores: (a) LPB where the shaded areas show the standard deviations for each skill score and (b) the dry and wet sectors separated by 60°W. Solid lines correspond to the wet sector and dashed lines to the dry sector.

results suggest that with the proper choice of parameterizations, the model performance over southern South America is similar to that in North America.

4. Evaluation of temperature forecasts

The forecasts of minimum, mean, and maximum temperature are evaluated for each of the seven lead times by comparing the observations at all stations and the model's nearest grid points. Minimum, mean, and maximum daily temperature (T_{mn}, T, and T_{mx} respectively) are evaluated using scatterplots and two objective score measures: the correlation coefficient r and the mean absolute error, $\mathrm{MAE} = (1/N)\sum_{i=1}^{N}|\mathrm{Tf}_i - \mathrm{To}_i|$, where N is the total number of days, and Tf_i and To_i are the forecast and observed temperatures (minimum, mean, or maximum), respectively, for the day i. Perfect forecasts would have a zero mean absolute error; in real cases, the larger the value, the poorer the forecast will be.

The scatterplots for the sample station (Sauce Viejo) at two lead times (1 and 7 days) are presented in Fig. 5. The mean temperature has correlation coefficients that range from 0.97 (day 1) to 0.89 (day 7). Both maximum and minimum temperatures have slightly lower correlations, yet always exceed 0.83. The slope m of the regression lines of the mean and maximum temperatures are notably close to 1 (1.02 and 0.97, respectively) for 1-day lead time and decay to 0.91 and 0.88 for 7-day lead times. The slope of the minimum temperature regression line varies between 0.87 and 0.74. In general, as the forecast lead time moves from day 1 to day 7, the dispersion becomes larger while the correlation coefficient becomes smaller and the slope m of the regression lines moves off from 1. Note that most values of T_{mn}, T, and

T_{mx} are below the 1:1 line (line not shown), suggesting that the forecast temperatures for Sauce Viejo tend to underestimate the observed temperatures.

Since the daily mean temperature is computed from the average of eight values per day, some errors are canceled out, resulting in higher correlation values. Moreover, the minimum and maximum values are obtained from the 3-h forecasts, carrying a 1.5-h uncertainty with respect to the precise timing of the maximum or minimum value. Temperature errors tend to be systematic, and if desired, bias correction methods can be applied for more accurate forecasts.

Figure 6 presents the spatial distributions of the point correlations and the mean absolute errors of T_{mn}, T, and T_{mx} averaged over each of the 7 lead times (as was done for precipitation skill scores in Fig. 3). The patterns of the correlation maps (Figs. 6a,c,e) are similar, with the highest values toward the south (values close to 0.9) and decreasing toward the north of the domain. The mean absolute error reveals a more heterogeneous distribution. The daily mean and minimum temperature MAE ranges from 1° to 2.5°C throughout the domain, with isolated areas larger than 3°C (Figs. 6b,d). The MAE of the maximum temperature (Fig. 6f) can also reach values exceeding 3°C. Larger values are found over arid and mountainous regions, both near the Andes and toward the northeast corner of the domain, likely related to the topographical features of the Brazilian Highlands. As for precipitation, these values are not different from those found over the United States (e.g., Cheng and Steenburgh 2005; Wyszogrodzki et al. 2013).

The spatial gradients suggest a better level of performance by the model toward the southern part of the domain where the temperature forecasts (either minimum,
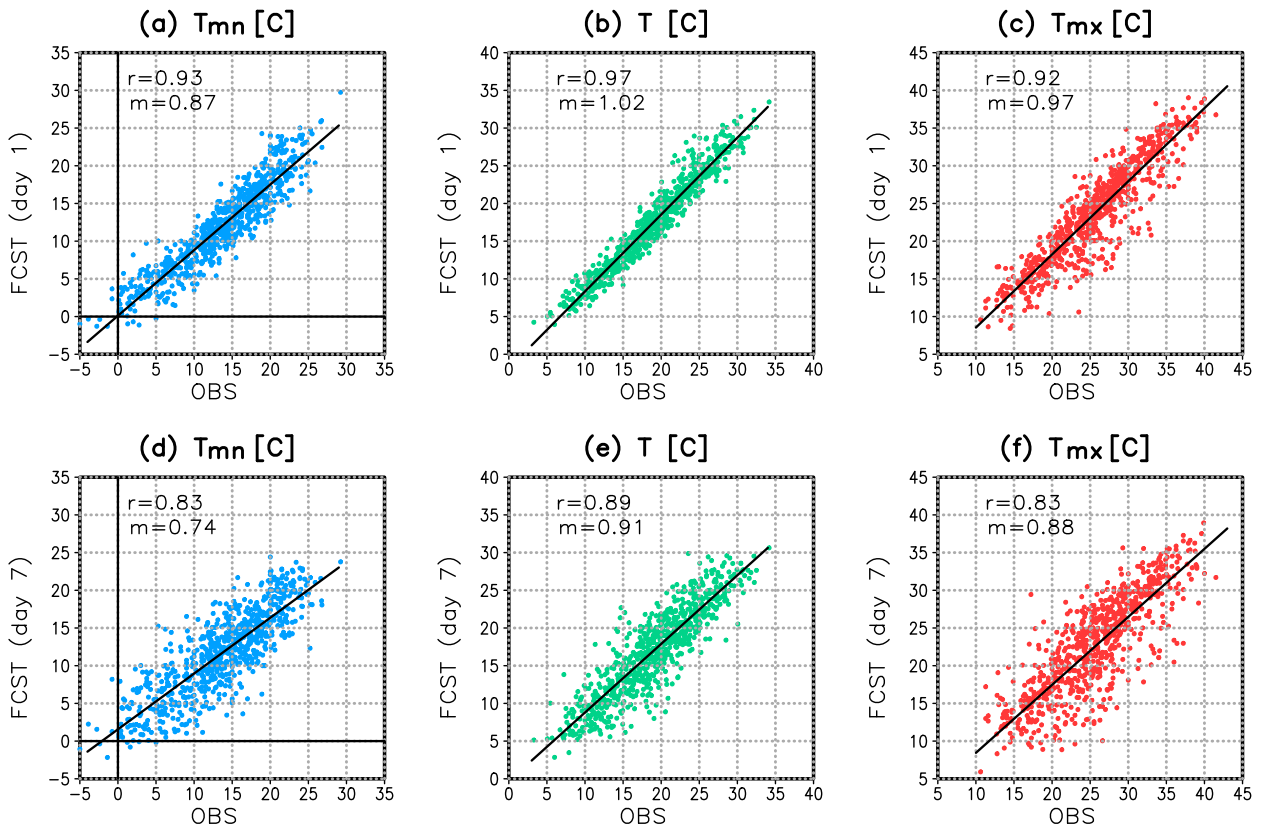
FIG. 5. Scatterplots of observed vs (top) day 1 and (bottom) day 7 forecast lead times. (a),(d) T_{mn}, (b),(e) T, and (c),(f) T_{mx}. The linear regression line is plotted, and its slope m is presented along with the correlation coefficient r in the top-left corner of each scatterplot.

mean, or maximum temperature) have the largest correlations and smallest mean absolute errors. At the subbasin level, the Uruguay and lower Paraná basins in the southern LPB exhibit almost homogeneously distributed temperature correlations with values above $r \sim 0.85$, while the temperature correlations across the Paraguay and the mid-to upper Paraná subbasins range from 0.75 to 0.9. As in the case of the sample station, the daily mean temperature correlations are slightly higher than for minimum and maximum temperatures.

The evolution with forecast lead time of the LPB-averaged temperature statistics (r and MAE) is presented in Fig. 7. As in Fig. 4a, the shaded bands in Fig. 7 represent the spread in the scores, that is, the scores plus/minus their standard deviation. These spreads are smallest for the minimum temperature and largest for the maximum temperature. Good correspondence is found between the observed and forecast temperatures with almost all correlation coefficients exceeding $r \sim 0.8$. Correlations for lead times of 1 to 7 days are in the range 0.95–0.85 for the mean temperature (Fig. 7b); values are slightly lower for the minimum temperature, whose range is 0.9–0.8 (Fig. 7a), and even slightly lower for the

maximum temperature, whose range is 0.88–0.78 (Fig. 7c). The mean absolute errors of the minimum and daily mean temperature forecasts are between about 2° and 2.7°C for days 1 and 7, respectively (Figs. 7a,b). Somewhat higher values are found for the maximum temperature with the MAE starting at 2.5°C on day 1 and reaching errors of 3.7°C on day 7 (Fig. 7c).

5. Application of the forecasts for monitoring

The model that was evaluated here is part of a daily updated system where the first 24 h out of the 168-h forecasts are used to create continuous time series employed to monitor recent hydroclimate conditions. The system thus consists of a forecasting module to characterize the anticipated weather up to 1 week in advance (as discussed in sections 3 and 4), and a monitoring module to depict the hydroclimate of recent months. All information is available online (www.atmos.umd.edu/~berbery/research/forecasts.html). Although the evaluation was done for the higher-resolution domain that covers the LPB, all diagnostics are also offered, untested, for the larger domain covering South America.
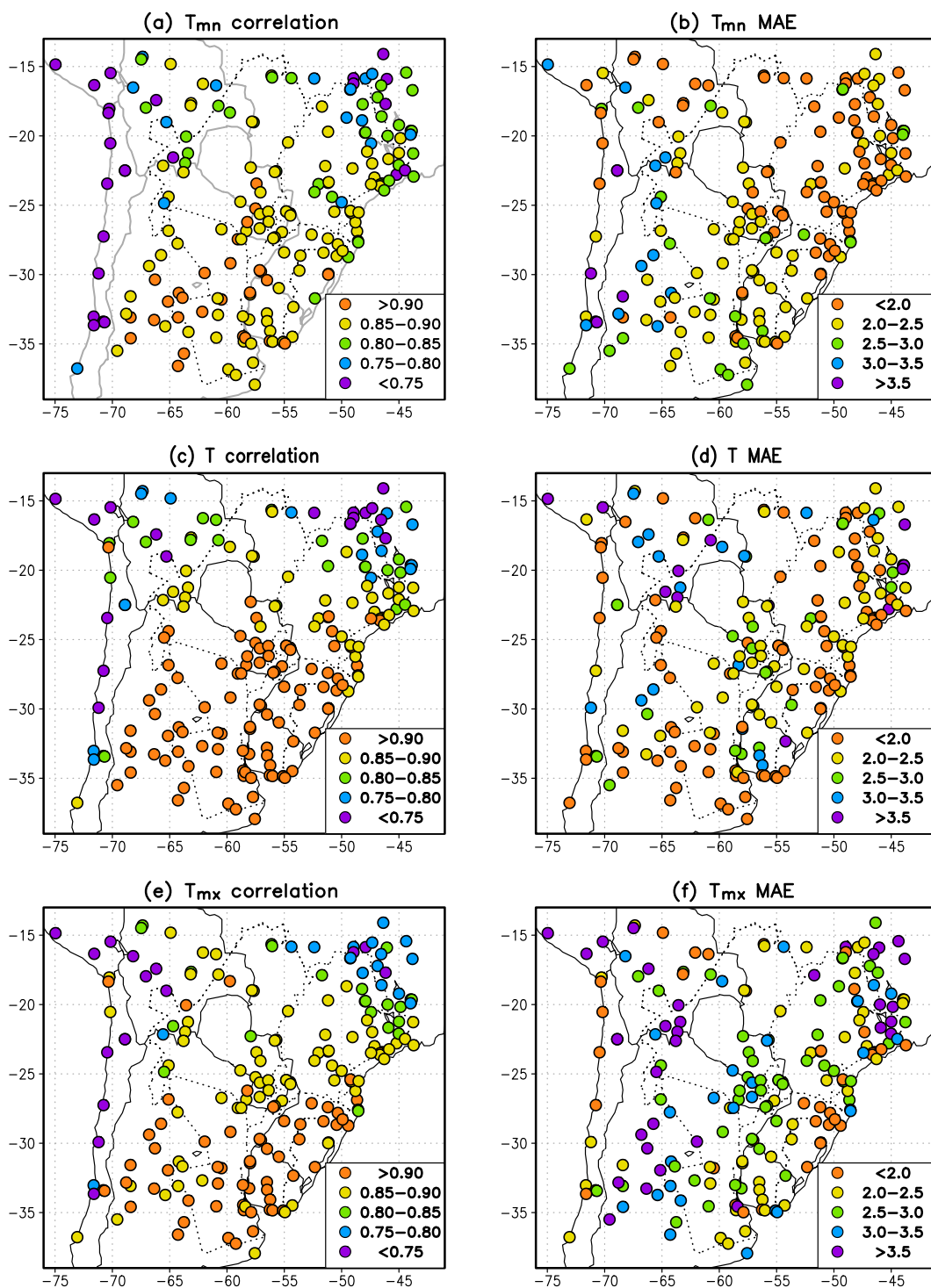
FIG. 6. Geographical distribution of temperature statistical scores averaged over the seven lead times: (left) correlation and (right) MAE of (a),(b) T_{mn}, (c),(d) T, and (e),(f) T_{mx}.

The variables that are displayed are precipitation, 2-m temperature, 10-m winds, sea level pressure, soil moisture, evapotranspiration, and runoff. When relevant, variables are presented as anomalies, that is, deviations from a climatology obtained from a 10-yr simulation (2001–10) carried out with the same model parameterizations and domains as used here. An evaluation of the climatology is discussed in Müller (2015).
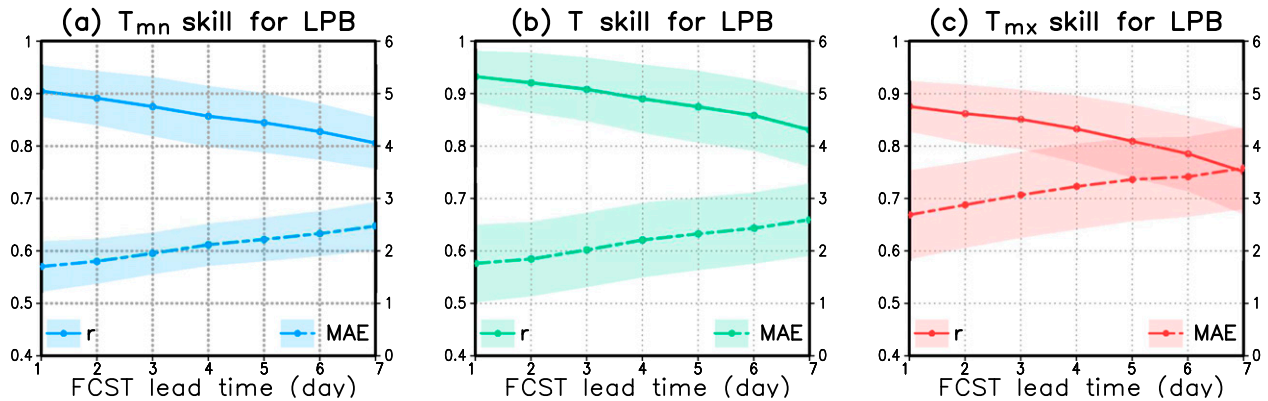
FIG. 7. Evolution with forecast lead time of the temperature statistics (correlation and MAE) area averaged over LPB: (a) T_{mn}, (b) T, and (c) T_{mx}. The scale along the left axis of each panel is for the correlation coefficient, and the scale along the right axis is for the MAE. The shaded bands indicate the corresponding standard deviations.

Table 4 summarizes the set of graphics for the forecast module in the system. The module displays the 3-hourly forecasts for up to 7 days for each variable, as well as mean maps and anomaly maps for the 7-day average. Time series at selected locations (meteograms) are also available. The monitoring module (Table 5) displays mean and anomaly maps of the hydrometeorological variables for the last 30, 60, and 90 days (updated daily). The maps are complemented with basin-averaged time series depicting deficits or excesses for the appropriate variables.

The use of WRF data to monitor the recent climate is exemplified for two main regions: the upper LPB (i.e., Paraguay and the mid- to upper Paraná subbasins), where a prolonged 2014/15 drought took place, and the lower LPB (Uruguay and lower Parana subbasins), where conditions were favorable for agricultural activities. This north–south structure reflects a well-documented warm season dipole pattern (see Casarin and Kousky 1986; Kousky and Cavalcanti 1988; Nogués-Paegle and Mo 1997, among others). The observational dataset used to assess the monitoring results consists of in situ observations interpolated to a regular grid at NCEP's Climate Prediction Center (Chen et al. 2008).

Southern Brazil experienced a dramatic water crisis with hydropower supply deficits leading to rolling power cuts, as well as constraints on agriculture and industrial production (Guardian 2015; Escobar 2015; Getirana 2016). Reports have also stated that the drought was the worst in the last 84 yr in southeastern Brazil, causing a large impact on the metropolitan area of São Paulo, where its 19 million inhabitants experienced never before seen water deficits (Escobar 2015; International Business Times 2015).

TABLE 4. Forecast module tools.

Tool	Description	Variables	Domains
3-h maps	Animations showing the evolution of the forecasts for the next 168 h with a time step of 3 h	Precipitation	SA, LPB
		Temp Winds at 10 m	
7-day avg map	Time avg map for the forecasts in next 168 h; it can be seen as the avg of all 3-h maps	Precipitation	SA, LPB
		Evapotranspiration Soil moisture at 0.4- and 2.0-m depths Max and min temp Winds at 10 m	
Meteogram	Time evolution of all variables for a specific location in one graphic (one below another)	Precipitation	Specific locations
		Evapotranspiration Soil moisture at 0.4- and 2.0-m depths Runoff Temp Winds at 10 m	

TABLE 5. Monitoring module tools.

Tool	Description	Variables	Domains
Maps	Time avg maps in the last 30, 60, and 90 days showed as total values and anomalies	Precipitation	SA, LPB
		Evapotranspiration	
		Runoff	
		Soil moisture at 0.4- and 2.0-m depths	
		Max and min temp	
Time series	Areal avg in the last 30, 60, and 90 days of hydrological variables; the time evolution of all variables is shown in one figure	Precipitation	Basins: LPB, PAY, MUP, LOP, URU
		Accumulated precipitation; soil moisture at 0.4- and 2.0-m depths; evapotranspiration	
		Runoff	

Figure 8a presents the upper LPB area-averaged observed daily precipitation during the 2014 austral spring. Compared to the precipitation climatology (continuous black curve), there was a lower frequency and lower intensity of rain events than usual. (The observed precipitation climatology was computed as the mean annual cycle for the period 2001–10, in a similar way as was done for the model precipitation climatology.) The few cases where individual storms exceeded the climatology were not enough to offset the water deficit, as noted in the bottom panel of Fig. 8a. The corresponding model-based figures, complemented with other moisture budget terms, are shown in Fig. 8b. Both observations and model precipitation have a notable resemblance in the day-to-day evolution. However, model precipitation rarely exceeds the climatology, and thus the dry anomalies are larger. The repeated short-term dry spells noted in both datasets (e.g., during October, mid-November, and early December) lead to large negative anomalies, as shown by the accumulated rain series. The figure also suggests that the dry spring is related to a weak onset of the South American monsoon that develops at this time of the year (Zhou and Lau 1998; Liebmann and Mechoso 2011; Marengo et al. 2012).

The increasing model precipitation deficit during the austral spring is reflected in the other water budget terms, with negative anomalies for soil water content, evapotranspiration, and runoff. The water deficit continued to grow during the 2014/15 austral summer (not shown), when the drought was most intense. Figures 9a and 9b present the 2014 austral spring observed and model precipitation anomalies, respectively. Although WRF overestimates the dry anomalies to the north, the pattern has a spatial distribution that is similar to that of the observations (Fig. 9a). The model precipitation deficits over the upper LPB (Fig. 9b) lead to a reduction in soil moisture (Fig. 9c), and also in the evapotranspiration and runoff (Figs. 9d,e). The mid- to upper Paraná, where São Paulo

is located, is the most affected subbasin with the largest precipitation and runoff deficits, which is consistent with Escobar (2015), who discussed the impacts on the river flows and the provision of water to metropolitan areas.

The opposite situation is found south of ~30°S, where an excess of observed and model precipitation and positive anomalies for the other water budget terms are found (Fig. 9). The time evolution of the observed precipitation for the lower LPB (Fig. 8c) indicates frequent precipitation events that contributed to the increase in the accumulated wet anomalies during the season. These events are also identified by the model although with an overestimation of their intensities. The time evolution of the anomalies of the simulated hydroclimate variables (Fig. 8d) indicates that most of the excess precipitation was partitioned between evapotranspiration, soil moisture accumulation, and runoff. Accumulation of water in the soil is detected as a slight wet anomaly in the first 40-cm layer (where crops develop roots). The larger effect is in the evapotranspiration while the remainder was converted into runoff. Runoff responds almost immediately to each precipitation event (see, e.g., 29–30 October and 29–30 November). The season's anomalies show, as did the time evolution, that the excess of rainfall is balanced mostly through the increased evapotranspiration with a smaller contribution from runoff (Figs. 9d,e).

Contrary to the dry anomalies that affected the northern regions with strong negative social and economic impacts, the wet anomalies toward the south and the well-distributed spring rainfall favored the planting of corn in September, improving its yield. It also favored the planting of soybean and corn in December in high and well-drained terrain, leading to historical records for soybean production. Conversely, over the lowlands these conditions caused declines in yields and losses in the December crops as a result of the excess in accumulated water (El Cronista 2014).
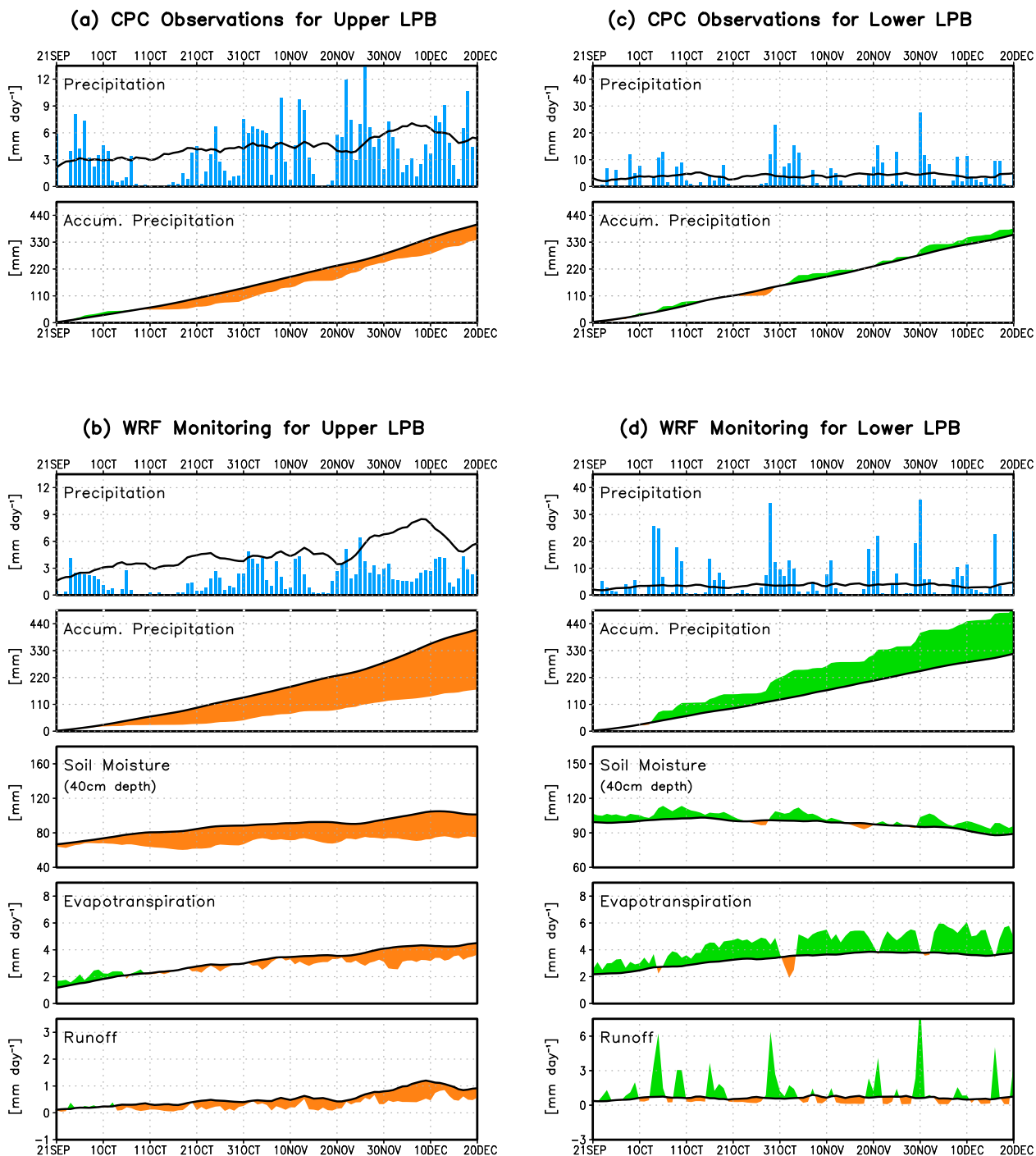
FIG. 8. Time series for the period 21 Sep–19 Dec 2014 for hydroclimate variables area averaged over the (left) upper and (right) lower LPB. (a),(c) CPC-observed gridded precipitation time series and (b),(d) WRF-derived time series. Precipitation is represented by bars while accumulated precipitation, soil moisture, evapotranspiration, and runoff series are presented as anomalies with respect to a 2001–10 model climatology (black line). Green shading denotes water excess while brown shading indicates water deficit. All model time series were created from the 24-h daily forecasts.
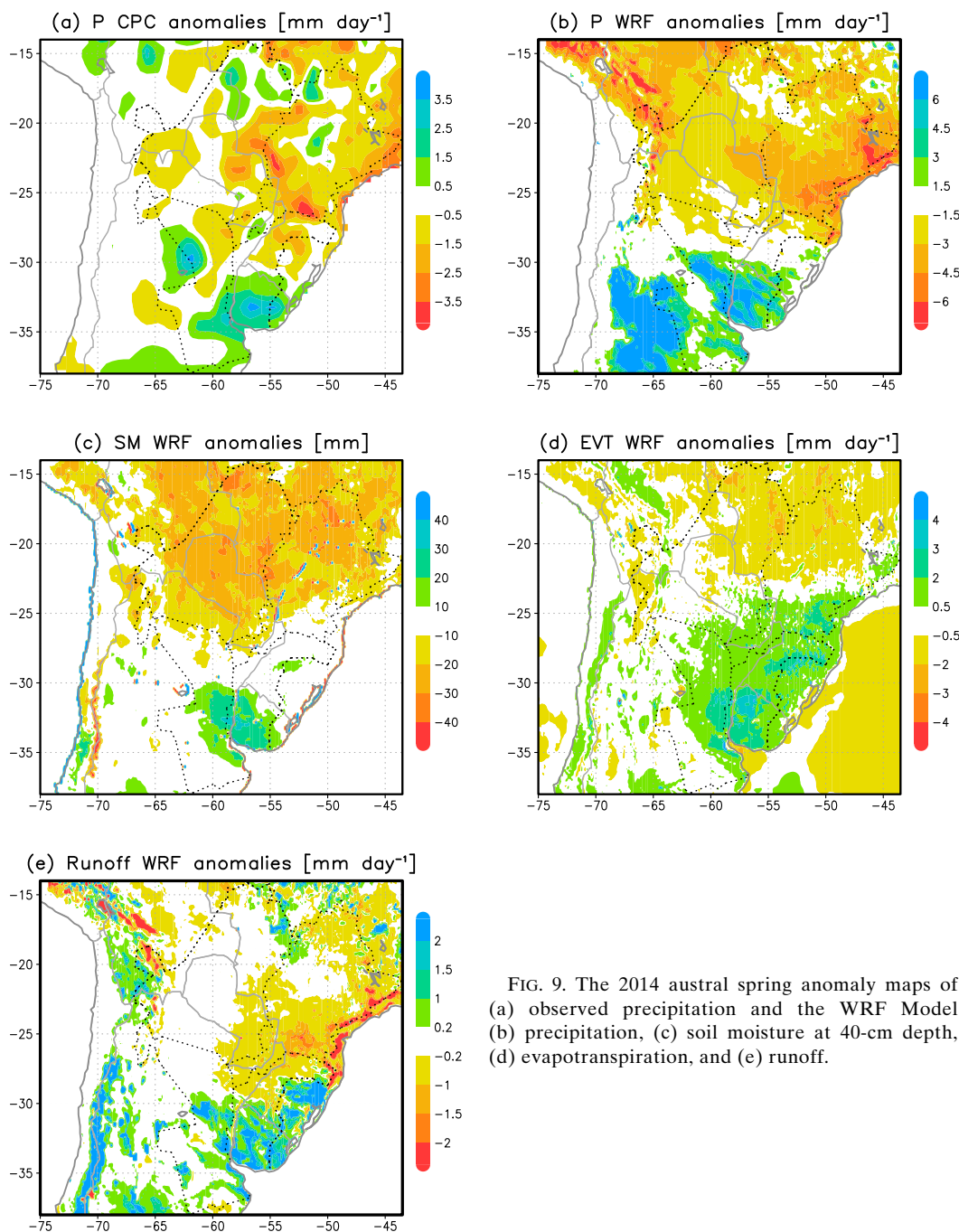
FIG. 9. The 2014 austral spring anomaly maps of (a) observed precipitation and the WRF Model (b) precipitation, (c) soil moisture at 40-cm depth, (d) evapotranspiration, and (e) runoff.

Monitoring based on observations is already being done at the RCC-SSA and other places. The quality of the regional model forecasts gives us confidence in their reliability for use as a complement to such observation-based systems. One advantage is that forecasts are available in real time, while observations usually have a latency inherent in the quality control process. Currently, the system is being transferred to the Regional Climate Center for Southern South America as a complement for their own activities.

6. Summary and conclusions

A system based on regional model simulations was developed with the objective of providing useful information for agriculture and water and risk management

activities in the La Plata basin in southern America. This study seeks to support operations in climate monitoring and weather forecasting at the Regional Climate Center for Southern South America. The first step in this process is the evaluation, through an analysis of statistical scores, of the WRF Model precipitation and temperature forecasts.

A set of high quality observations with records that cover at least 90% of the days were employed to evaluate precipitation forecasts in terms of accuracy, probability of detection that identifies potential rain days and should be of interest for risk management support in urban areas, false alarm ratio, frequency bias, and the Heidke skill score. The combined analysis of these scores is needed to avoid misrepresentations, like large verification scores due to the high volume of correct negatives (Murphy 1996). The precipitation forecasts maintain high skill up to forecast day 7 (the maximum length analyzed here). The model performance is best over the LPB, and the scores tend to be better over humid climates than over arid-to-semiarid climates. Compared to the arid–semiarid climate, the humid climate has a higher probability of detection and fewer false alarms. The ranges of the skill scores are similar to those found over the United States, suggesting that the proper choice of parameterizations leads to no loss of performance by the model.

Daily mean, minimum, and maximum forecast temperatures are highly correlated with observations up to 7-day lead time. The best performance is for daily mean temperature, followed by minimum temperature and a slightly weaker performance for the maximum temperature over arid regions. In general, forecast temperatures tend to correspond better with observations over the temperate climate and subtropical regions. The largest temperature biases and lower skill scores are found toward the northern warmer climate and over semiarid regions characterized by large diurnal amplitude.

The evaluation of the precipitation and temperature forecasts suggests that the WRF Model performs satisfactorily in the domain, and particularly within the LPB. The quality of the precipitation and temperature 7-day forecasts does not show a large decay with time, indicating that the period could be extended into the second week (but remaining within the ranges of predictability of the dynamical system). The analysis suggests that at least parts of the errors are systematic in nature, and thus they can be reduced by applying bias correction methods.

The extraordinary drought affecting southern Brazil and simultaneous beneficial conditions for agriculture to the south were discussed to demonstrate the usefulness of the WRF Model and the system products to describe the anomalous conditions in real time. Monitoring based on observations is already being performed at the RCC-SSA and other places. Rather, our goal was to present the basic elements of a monitoring system based not on observations that have a latency inherent to quality controls, but on systematic, reliable regional model forecasts.

Future plans include the implementation of the verification as a routine task of the system. Continued evaluation would help detect possible model performance changes due to improved inputs, model updates, or the increased number and quality of the observations. We intend to extend the continuous evaluation to cover precipitation intensity and additional in situ observations such as soil moisture, winds, and humidity.

REFERENCES

Berbery, E. H., and E. A. Collini, 2000: Springtime precipitation and water vapor flux over southeastern South America. *Mon. Wea. Rev.*, **128**, 1328–1346, doi:10.1175/1520-0493(2000)128<1328:SPAWVF>2.0.CO;2.

——, and V. Barros, 2002: The hydrologic cycle of the La Plata basin in South America. *J. Hydrometeor.*, **3**, 630–645, doi:10.1175/1525-7541(2002)003<0630:THCOTL>2.0.CO;2.

Campana, K., and P. Caplan, Eds., 2006: Technical procedures bulletin for the T382 Global Forecast System. NOAA/NCEP/EMC. [Available online at http://www.emc.ncep.noaa.gov/gc_wmb/Documentation/TPBoct05/T382.TPB.FINAL.htm.]

Casarin, D. P., and V. E. Kousky, 1986: Precipitation anomalies in southern Brazil and variations in the atmospheric circulation (in Portuguese). *Rev. Bras. Meteor.*, **1**, 83–90.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, M., W. Shi, P. Xie, V. Silva, V. E. Kousky, R. Wayne, R. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, doi:10.1029/2007JD009132.

Cheng, W. Y. Y., and W. J. Steenburgh, 2005: Evaluation of surface sensible weather forecasts by the WRF and the Eta Models

over the western United States. *Wea. Forecasting*, **20**, 812–821, doi:10.1175/WAF885.1.

Chou, S. C., J. F. Bustamante, and J. L. Gomes, 2005: Evaluation of Eta Model seasonal precipitation forecasts over South America. *Nonlinear Processes Geophys.*, **12**, 537–555, doi:10.5194/npg-12-537-2005.

Collini, E. A., E. H. Berbery, V. Barros, and M. Pyle, 2008: How does soil moisture influence the early stages of the South American monsoon? *J. Climate*, **21**, 195–213, doi:10.1175/2007JCLI1846.1.

de Goncalves, L. G. G., W. J. Shuttleworth, B. Nijssen, E. J. Burke, J. A. Marengo, S. C. Chou, P. Houser, and D. L. Toll, 2006: Evaluation of model-derived and remotely sensed precipitation products for continental South America. *J. Geophys. Res.*, **111**, D16113, doi:10.1029/2005JD006276.

de Moura, R. G., D. L. Herdies, D. Mendes, and M. Damiao Mendes, 2010: Evaluation of the Eta regional model using the analysis of CPTEC and NCEP (in Portuguese). *Rev. Bras. Meteor.*, **25**, 46–53, doi:10.1590/S0102-77862010000100005.

Dillon, M. E., Y. García Skabar, and M. Nicolini, 2013: Performance of forecast from high resolution models, in a limited area: Analysis of the 2010-2011 summer season (in Spanish). *Meteorologica*, **38**, 69–89.

Dimitrijevic, M., and R. Laprise, 2005: Validation of the nesting technique in a regional climate model through sensitivity tests to spatial resolution and the time interval of lateral boundary conditions during summer. *Climate Dyn.*, **25**, 555–580, doi:10.1007/s00382-005-0023-6.

Dravitzki, S., and J. McGregor, 2011: Predictability of heavy precipitation in the Waikato River basin of New Zealand. *Mon. Wea. Rev.*, **139**, 2184–2197, doi:10.1175/2010MWR3137.1.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

El Cronista, 2014: Las lluvias ayudaron a la siembra de soja, que está casi terminada (in Spanish). [Available online at http://www.cronista.com/negocios/Las-lluvias-ayudaron-a-la-siembra-de-soja-que-esta-casi-terminada-20141231-0025.html.]

Environmental Modeling Center, 2003: The GFS atmospheric model. NCEP Office Note 442, EMC/Global Climate and Weather Modeling Branch, Camp Springs, MD, 14 pp. [Available online at http://www.lib.ncep.noaa.gov/ncepofficenotes/files/on442.pdf.]

Escobar, H., 2015: Drought triggers alarms in Brazil's biggest metropolis. *Science*, **347**, 812–812, doi:10.1126/science.347.6224.812.

Gandin, K. S., and A. H. Murphy, 1992: Equitable scores for categorical forecasts. *Mon. Wea. Rev.*, **120**, 361–370, doi:10.1175/1520-0493(1992)120<0361:ESSFCF>2.0.CO;2.

García Skabar, Y., L. Vidal, P. Salio, and M. Nicolini, 2011: Experimental high-resolution forecast in a region of Argentina. *Research Activities in Atmospheric and Oceanic Modelling (Blue Book)*, A. Zadra, Ed., WMO/Working Group in Numerical Experimentation, 9–10.

——, P. Salio, and M. Nicolini, 2012: BRAMS forecast verification over a region of subtropical South America (in Spanish). *Rev. Bras. Meteor.*, **27**, 291–306, doi:10.1590/S0102-77862012000300004.

Getirana, A., 2016: Extreme water deficit in Brazil detected from space. *J. Hydrometeor.*, **17**, 591–599, doi:10.1175/JHM-D-15-0096.1.

Guardian, 2015: Brazil struggles with drought and pollution as Olympics loom large. [Available online at http://www.theguardian.com/sustainable-business/2015/may/04/brazil-drought-2016-olympic-games-rio-de-janiero-rio-20-climatechange/.]

Hamill, T., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi:10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

Hogan, R. J., and I. A. Mason, 2012: Deterministic forecast of binary events. *Forecast Verification—A Practitioner's Guide in Atmospheric Science, Second Edition*, I. T. Jolliffe and D. V. Stephenson, Eds., John Wiley and Sons, 31–59.

——, C. A. T. Ferro, I. T. Jolliffe, and D. B. Stephenson, 2010: Equitability revisited: Why the "equitable threat score" is not equitable. *Wea. Forecasting*, **25**, 710–726, doi:10.1175/2009WAF2222350.1.

International Business Times, 2015: São Paulo drought 2015: Photos of historic water crisis in Brazil show city on the brink of collapse. [Available online at http://www.ibtimes.com/sao-paulo-drought-2015-photos-historic-water-crisis-brazil-show-city-brink-collapse-1912767/.]

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, doi:10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

——, 1996: The surface layer in the NCEP Eta Model. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 354–355.

——, 2000: Comments on "Development and evaluation of a convection scheme for use in climate models." *J. Atmos. Sci.*, **57**, 3686–3686, doi:10.1175/1520-0469(2000)057<3686:CODAEO>2.0.CO;2.

Kok, K., B. W. Schreur, and D. Vogelezang, 2008: Valuing information from mesoscale forecasts. *Meteor. Appl.*, **15**, 103–111, doi:10.1002/met.54.

Kousky, V. E., and I. F. A. Cavalcanti, 1988: Precipitation and atmospheric circulation anomaly patterns in the South American sector. *Rev. Bras. Meteor.*, **3**, 199–206.

Lee, S.-J., 2010: Impact of land surface vegetation change over the La Plata Basin on the regional climatic environment: A study using conventional land-cover/land-use and newly developed ecosystem functional types. Ph.D. thesis, University of Maryland, College Park, College Park, MD, 153 pp. [Available online at http://drum.lib.umd.edu/handle/1903/10831.]

——, and E. H. Berbery, 2012: Land cover change effects on the climate of the La Plata basin. *J. Hydrometeor.*, **13**, 84–102, doi:10.1175/JHM-D-11-021.1.

Liebmann, B., and C. R. Mechoso, 2011: The South American monsoon system. *The Global Monsoon System: Research and Forecast*, 2nd ed., C. P Chang et al., Eds., World Scientific Series on Asia-Pacific Weather and Climate, Vol. 5, World Scientific, 137–158.

Marengo, J., and Coauthors, 2012: Recent developments on the South American monsoon system. *Int. J. Climatol.*, **32**, 1–21, doi:10.1002/joc.2254.

Menéndez, C. G., and Coauthors, 2010: Downscaling extreme month-long anomalies in southern South America. *Climatic Change*, **98**, 379–403, doi:10.1007/s10584-009-9739-3.

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, doi:10.1029/97JD00237.

Müller, O. V., 2015: Hydrometeorological forecast and their sensitivity to land cover changes. Ph.D. thesis, Universidad Nacional del Litoral, Santa Fe, Argentina, 132 pp.

——, E. H. Berbery, D. Alcaraz-Segura, and M. B. Ek, 2014: Regional model simulations of the 2008 drought in

southern South America using a consistent set of land surface properties. *J. Climate*, **27**, 6754–6778, doi:10.1175/JCLI-D-13-00463.1.

Murphy, A., 1996: The Finley affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11**, 3–20, doi:10.1175/1520-0434(1996)011<0003:TFAASE>2.0.CO;2.

Nogués-Paegle, J., and K. C. Mo, 1997: Alternating wet and dry conditions over South America during summer. *Mon. Wea. Rev.*, **125**, 279–291, doi:10.1175/1520-0493(1997)125<0279:AWADCO>2.0.CO;2.

Pappenberger, F., A. Ghelli, R. Buizza, and K. Bodis, 2009: The skill of probabilistic precipitation forecasts under observational uncertainties within the generalized likelihood uncertainty estimation framework for hydrological applications. *J. Hydrometeor.*, **10**, 807–819, doi:10.1175/2008JHM956.1.

Pei, L., N. Moore, S. Zhong, L. Luo, D. W. Hyndman, W. E. Heilman, and Z. Gao, 2014: WRF model sensitivity to land surface model and cumulus parameterization under short-term climate extremes over the southern Great Plains of the United States. *J. Climate*, **27**, 7703–7724, doi:10.1175/JCLI-D-14-00015.1.

Pennelly, C., G. Reuter, and T. Flesch, 2014: Verification of the WRF model for simulating heavy precipitation in Alberta. *Atmos. Res.*, **135–136**, 172–192, doi:10.1016/j.atmosres.2013.09.004.

Rauscher, S. A., A. Seth, B. Liebmann, J.-H. Qian, and S. J. Camargo, 2007: Regional climate model-simulated timing and character of seasonal rains in South America. *Mon. Wea. Rev.*, **135**, 2642–2657, doi:10.1175/MWR3424.1.

Rogers, E., M. Ek, Y. Lin, K. Mitchell, D. Parrish, and G. DiMego, 2001: Changes to the NCEP Meso Eta analysis and forecast system: Assimilation of observed precipitation, upgrades to land-surface physics, modified 3DVAR analysis. NWS Tech. Procedures Bull. 488. [Available online at http://www.emc.ncep.noaa.gov/mmb/mmbpll/eta12tpb/.]

Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction*, S. C. Michaelides, Ed., Springer Science and Business Media, 419–452.

Ruiz, J. J., C. Saulo, and J. Nogués-Paegle, 2010: WRF Model sensitivity to choice of parameterization over South America: Validation against surface variables. *Mon. Wea. Rev.*, **138**, 3342–3355, doi:10.1175/2010MWR3358.1.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Sanchez, E., and Coauthors, 2015: Regional climate modelling in CLARIS-LPB: A concerted approach towards twenty-first century projections of regional temperature and precipitation over South America. *Climate Dyn.*, **45**, 2193–2212, doi:10.1007/s00382-014-2466-0.

Saulo, C., M. Seluchi, C. Campetella, and L. Ferreira, 2001: Error evaluation of NCEP and LAHM regional model daily forecasts over southern South America. *Wea. Forecasting*, **16**, 697–712, doi:10.1175/1520-0434(2001)016<0697:EEONAL>2.0.CO;2.

——, S. Cardazzo, J. Ruiz, C. Campetella, and A. Rolla, 2008: Experimental forecast system at the research center for the sea and the atmosphere (in Spanish). *Meteorologica*, **33**, 83–97.

Seluchi, M. E., and S. C. Chou, 2001: Evaluation of two Eta Model versions for weather forecast over South America. *Geofis. Int.*, **40**, 219–237.

——, Y. V. Serafini, and H. Le Treut, 1998: The impact of the Andes on transient atmospheric systems: A comparison between observations and GCM results. *Mon. Wea. Rev.*, **126**, 895–912, doi:10.1175/1520-0493(1998)126<0895:TIOTAO>2.0.CO;2.

Stone, R. C., and H. Meinke, 2005: Operational seasonal forecasting of crop performance. *Philos. Trans. Roy. Soc.*, **360B**, 2109–2124, doi:10.1098/rstb.2005.1753.

Suaya, M., 2004: First objective evaluation of the ETA-SMN and NCEP GFS-operative forecasts during year 2003 (in Spanish). Licentiature thesis, Dept. of Atmosphere and Oceans Sciences, University of Buenos Aires, Buenos Aires, Argentina, 68 pp.

Tartaglione, N., 2010: Relationship between precipitation forecast errors and skill scores of dichotomous forecasts. *Wea. Forecasting*, **25**, 355–365, doi:10.1175/2009WAF2222211.1.

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, doi:10.1175/2007WAF2007005.1.

Werth, D., and A. Garrett, 2011: Patterns of land surface errors and biases in the Global Forecast System. *Mon. Wea. Rev.*, **139**, 1569–1582, doi:10.1175/2010MWR3423.1.

Wilks, D. S., 2011: Forecast verification. *Statistical Methods in the Atmospheric Sciences*, D. S. Wilks, Ed., Academic Press, 301–393.

WMO, 2014: Implementation plan of the global framework for climate services. World Meteorological Organization, 70 pp. [Available online at http://gfcs.wmo.int/sites/default/files/implementation-plan//GFCS-IMPLEMENTATION-PLAN-FINAL-14211_en.pdf.]

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, doi:10.1175/WAF-D-13-00135.1.

Wyszogrodzki, A. A., Y. Liu, N. Jacobs, P. Childs, Y. Zhang, G. Roux, and T. T. Warner, 2013: Analysis of the surface temperature and wind forecast errors of the NCAR-AirDat operational CONUS 4-km WRF forecasting system. *Meteor. Atmos. Phys.*, **122**, 125–143, doi:10.1007/s00703-013-0281-5.

Zhou, J., and K.-M. Lau, 1998: Does a monsoon climate exist over South America? *J. Climate*, **11**, 1020–1040, doi:10.1175/1520-0442(1998)011<1020:DAMCEO>2.0.CO;2.

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**, 1057–1071, doi:10.1175/BAMS-87-8-1057.