Routine Forecasting and Monitoring System for Agricultural and Hydrological Applications on La Plata Basin

Omar V. Müller[1], and Ernesto H. Berbery[2]

[1] Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral (UNL), Ciudad Universitaria. Ruta Nacional N° 168 - Km 472,4. (3000) Santa Fe, Argentina.
[2] Earth System Science Interdisciplinary Center/Cooperative Institute for Climate and Satellites-Maryland, University of Maryland , College Park, MD 20752.

ovmuller@unl.edu.ar

ABSTRACT

La Plata Basin is usually affected by extreme events. Nevertheless, there is not enough climate information available to assist operational decisions at real-time. An operational system based on simulations of the Weather and Research Forecasting Model was developed with the objective of offering information as support for agriculture, hydrology and risk management. Every day simulations are carried out over South America, and a nested domain covering La Plata Basin. The forecasts module shows the expected evolution of each variable in the next seven days. The system has also a monitoring module that shows the behavior of hydrometeorological variables for the last 90 days. This set of tools allows easily to characterize the last months and what can be expected for the next days in order to plan activities, and reduce impacts of anomalous conditions. The verification of the forecast skill shows an average accuracy of about 70% for rainy days, and 85% of correlation with observed temperatures. Currently, the system outputs are used as input of an early warning system for flows and also, to support extended forecasts. In both cases the users value positively the utility of the system and recognize it as a key tool for their applications.

Keywords: hydrometeorological forecasts, hydrometeorological monitoring, decision-making.

RESUMEN

La cuenca del Plata es usualmente afectada por eventos extremos. Sin embargo, no existe suficiente información climática disponible para la toma decisiones en tiempo real. Un sistema operacional basado en simulaciones del modelo climático Weather and Research Forecasting Model fue desarrollado con el objetivo de ofrecer información para la gestión agrícola, hidrológica y de riesgo. Diariamente, se hacen simulaciones sobre Sudamérica, y un dominio anidado que cubre la cuenca del Plata. El módulo de pronósticos muestra la evolución esperada de cada variable en los próximos 7 días. El sistema tiene además un módulo de monitoreo que muestra el comportamiento de variables hidrometeorológicas en los últimos 90 días. Este conjunto de herramientas permite fácilmente caracterizar los últimos meses y determinar lo esperable en los próximos días para planificar actividades y reducir impactos ante condiciones anómalas. La verificación de la habilidad del pronóstico muestra una presición aproximada del 70% para días lluviosos y un 85% de correlación con temperaturas observadas. Actualmente, las salidas del modelo son utilizadas como entrada en un sistema de alerta temprana de caudales, y además, como soporte para pronósticos extendidos. En ambos casos los usuarios valoran la utilidad del sistema y lo reconocen clave para la toma de decisiones.

Palabras clave: Pronóstico hidrometeorológico, monitoreo hidrometeorológico, toma de decisiones.

1. INTRODUCTION

The climate of La Plata Basin is vital for the quality of life and economy of the region. This basin is one of the most densely populated regions of South America, where harvests and livestock are among the region’s most important assets. In the last decades, different parts of the basin were affected by climate change that modified precipitation frequency and intensity patterns (Giorgi, 2002). These changes favored the agricultural expansion towards the west, but also increased the occurrence of floods and droughts. Zipser et al. (2006) evaluated precipitation data derived from the

Tropical Rainfall Measuring Mission (TRMM) Satellite and detected that La Plata Basin, as well as the plains in central United States (east of Rocky Mountains) are the regions with highest occurrence of intense precipitation events worldwide.

Despite the features of the region, i.e., the high variability of the regional climate and its influence in economic activities, there is not enough practical information of hydrometeorological variables on real-time for the users. Forecasting and monitoring systems, linked with appropriate decision and discussion support tools, could substantially improve operational decision making in agricultural and water management (Stone and Meinke 2005). In the region of interest, there are a few initiatives for developing these kind of systems. The Servicio Meteorológico Nacional (SMN) of Argentina publishes daily a 24hr weather forecast of Argentina generated with the Eta Model, and a 18hr forecast with the Brazilian Regional Atmospheric Modeling System (BRAMS) over the Province of Buenos Aires (García Skabar et al., 2011). Also, SMN provides once a month monitoring information of some variables over LPB based on observations. The Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) of Brazil offers two products of forecasts over South America. They are based on numerical simulation with Eta and BRAMS. The Eta product provides an extended 11 days forecast, and the BRAMS product a 3 days forecast. In terms of monitoring, CPTEC has abundant information but limited to the Brazilian territory. Both centers (SMN and CPTEC) are pioneers in the region in providing weather information based on numerical models, and the products offered are based on Eta and BRAMS models. The products offered are spread across their websites, making them difficult to use. Our effort uses the community developed WRF model, which is the one of the most (if not the most) advanced regional model in existence. In addition, our products are tailored specifically for users in northeastern Argentina and other regions of high agricultural productivity and water needs.

The objective of this work is to develop predictive capabilities by implementing a real-time forecasting and monitoring system to produce practical information for the users and stakeholders, particularly in agricultural and hydrological management activities. The development aims to overcome the limitations of other systems in the region offering tools with graphical information in a unified system based on numerical simulation with the Weather Research Forecasting (WRF) Model, which is the most used by the worldwide community of climate modelers.

2. METHODOLOGY

2.1 Model Simulations

All products that make up the system are based on daily routine simulations with the WRF model. The model simulates a period of 168hr (7 days) with output every 3hr. It ingests initial and 6hr boundary conditions from the Global Forecast System (GFS) and is run over two nested domains. The parent domain covers South America and has a resolution of 45km, while the nested domain is centered over the LPB. The selected outputs include variables of social, meteorological, agricultural and hydrological interest such as precipitation, temperature, winds, pressure, soil moisture, evapotranspiration, runoff, and heat fluxes.

Some variables are shown as total values, and others as anomalies. The latter are computed as the difference with climatological variables obtained from an 11-year simulation. The long term simulation was carried out with the same model parameterizations and domains used on this system. The climate simulation was forced by the initial and 6hr lateral boundary conditions from the NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010). The period of simulation starts on January 1, 2000 and extends to December 31, 2010. Then, the mean annual cycle of each variable is considered as climatology. The first year, i.e., year 2000, is considered as spin-up year and therefore it is discarded from the computations.

2.2 System Overview

The system comprises two modules: forecasting and monitoring. All system figures are updated daily with the results of a routine simulation. It is available at http://www.atmos.umd.edu/~berbery/research/forecasts.html. All products t are available for the two domains: South America and LPB.

The forecasting module offers three types of products:

- Three hour maps: still images and an animation of the forecast maps for the next 168hr at 3hr interval are presented. The variables include precipitation and pressure, temperature at 2m (land-only) and winds at 10m. The animations give an idea of the evolution in time and the spatial distribution of the variables. In terms of events, they allow for the identification and tracking of potential centers of high pressure (anticyclones) and low pressure (cyclones), fronts, convective storms, windstorms, and the possible occurrence of heat waves and cold episodes.
- Seven day average maps: the module presents two maps showing the average forecast of the next 168hr and the anomalies with respect to the same period of the climatology. The variables include precipitation, evapotranspiration, soil moisture at 0.4m and 2.0m depth, runoff, maximum and minimum temperature, and winds at 10m. These maps show the expected behavior of the water balance, temperatures and winds in the next days. In the case that some part of the region is suffering an extreme event, a drought for instance, this tool may be useful when is combined with monitoring maps (explained later) to know whether the event evolution could be reverted or intensified.
- Meteograms: they are time series showing the evolution in the next 168hr of forecast variables for a given location. Precipitation bins and curves for evapotranspiration, soil moisture at 0.4m and 2.0m depth, runoff, temperature and wind intensity and direction are summarized in one figure, providing a quick report of several variables for a specific place. Usually the analysis of this figure facilitates an integrated view of the current meteorological event, showing, for example, the consistent changes in wind direction preceding a precipitation event. Then, after the event, the rest of the variables usually reflect the corresponding changes (changes in temperature, stability conditions, soil moisture, etc.).

The monitoring module is composed of:

- Maps: Similar to the 7-day average maps of the forecasting module, it shows average and anomaly maps of variables for the last 30, 60 and 90 days. The maps are computed with the model forecast variables instead of observations as it is the only way that allow a consistent comparison with the climatology and the rest of the system. The selected variables are precipitation, evapotranspiration, soil moisture at 0.4m and 2.0m depth, runoff, and maximum and minimum temperature. From this graphics, the user can easily detect if any region of the domain is experiencing persistent positive or negative anomalies of the given variable. In the case of precipitation, the periods of water excess or deficit can easily be identified.
- Times-series: the figures show the areal average evolution of the surface water balance variables for selected basins. The basins are LPB, Paraguay, Mid-Upper Paraná, Lower Paraná and Uruguay. Precipitation bins and curves for evapotranspiration, soil moisture at 0.4m and 2.0m depth, and runoff are shown from top to bottom. The curves are plotted as shaded bands limited by the forecast variables in the last 30, 60 and 90 days and the climatology during the same period of the year. When a variable is higher than its climatological value, the band shows green shading; when the variable is below its climatological value the band get brown. In this way the users know at a glance when a basin has excess or deficit of water.

2.3 Forecasts Verification

Since the entire system is based on daily simulations, it is essential to evaluate the quality of the forecasts with observations. The evaluation presented here is restricted to the inner domain and to daily values of precipitation and temperature at 2m, as they are the most commonly used variables. The period of evaluation begins on August 01, 2012 and ends on April 20, 2014, i.e., 628 days.

The observation data set (hereinafter called OBS) is from the National Climate Data Center (NCDC) of the United States, which gathers daily summaries of gauge station data from more than 9000 worldwide stations. Among the stations located inside the domain, we select those that pass a threshold of availability, i.e., those with the fewest undefined values for the period of evaluation are used, as to be considered statistically representative. The forecasts data set (hereinafter called FCST) is a collection of 7 individual data sets, one per forecast day. For instance, observations on 07-AUG-2012 have their corresponding forecasts from simulations issued on 01-AUG-2012 (7 days in advance), 02-AUG-2012 (6 days in advance), …, 07-AUG-2012 (1 day in advance). In all cases, each station data is compared to the forecast in the nearest grid point.

2.3.1 Precipitation Verification

The precipitation is evaluated in terms of occurrence of daily events. When observed precipitation for a given day is more than a preset rainfall threshold, it is considered a rainy day. Then, methods for dichotomous (yes/no) forecasts can be applied. To verify daily precipitation (P) events, contingency tables are calculated. A contingency table (see Table 1) shows the frequency of "yes" and "no" forecasts and occurrences for a given place. It is a useful way to see what types of errors are being made. The four combinations of forecasts (yes or no) and observation (yes or no), called the joint distribution, are:

- hit: event predicted to occur in forecast, and did occur in observations;
- miss: event not predicted in forecasts, but did occur in observations;
- false alarm: event predicted to occur in forecast, but did not occur;
- correct negative: event not predicted in forecast, and did not occur in observations.

Table 1 – Contingency Table

		OBS		
		YES	NO	TOT
FCST	YES	hits	False alarms	Forecasts yes
	NO	misses	Correct negatives	Forecasts no
	TOT	Observed yes	Observed no	total

A large variety of categorical statistics can be computed from the elements in thecontingency table to describe particular aspects of forecast performance. In this research three statistical values are selected:

$$Accuracy = \frac{hits+correct\ negatives}{total} \tag{1}$$

$$Probability\ of\ Detection = \frac{hits}{hits+misses} \tag{2}$$

$$False\ Alarm\ Ratio = \frac{false\ alarms}{false\ alarms+hits} \tag{3}$$

The accuracy (Eqn. 1) indicates what fraction of the forecasts were correct. The result can vary between 0 and 1, being 1 the perfect score. The Probability of Detection (POD) in Eqn. 2 shows what fraction of the observed "yes" events were correctly forecast. POD ranges from 0 to 1, being 1 the perfect score. The False Alarm Ratio (FAR) answer the question: What fraction of the predicted "yes" events actually did not occur (i.e., were false alarms)?. Perfect score is 0, but possible values range from 0 to 1.

2.3.2 Temperature Verification

Temperature is assessed in terms of mean daily temperatures (T). Unlike of precipitation verification where just yes/no events where evaluated, here the objective is to measures how the magnitudes of the forecasts differ from the magnitudes of the observations. This kind of verification methods include some selected scatter plots or box plots, as well as various summary scores. In this work, scatter plots are selected and the following statistics are defined:

$$Mean\ Absolute\ Error = \frac{1}{N}\sum_{i=1}^{N}|FCST_i - OBS_i| \quad (4)$$

$$Correlation\ Coefficient = \frac{\sum_{i=1}^{N}(FCST_i-\overline{FCST})-(OBS_i-\overline{OBS})}{\sqrt{\sum_{i=1}^{N}(FCST_i-\overline{FCST})^2}\sqrt{\sum_{i=1}^{N}(OBS_i-\overline{OBS})^2}} \quad (5)$$

where N is the total number of days, $FCST_i$ is forecasted T for the day i and OBS_i is observed T. $\overline{FCST}$ and $\overline{OBS}$ denote the mean value of the series. From Eqn. 4, the Mean Absolute Error (MAE) shows the average magnitude of the forecasts error, the ideal score is 0, but it can get any positive value. The Correlation Coefficient (r) in Eqn. 5 indicates the correspondence between forecast and observed values. It varies from -1 to 1, reaching a perfect score when it is equal to 1.

3. RESULTS AND DISCUSSION

Contingency tables for daily precipitation events and scatter plots for daily temperature, as well as their statistics, were computed for all stations in the domain. For practical reasons, they are shown here only for a sample station arbitrarily selected. Then, the behavior of the forecasts for all stations in the region of interest are summarized in skill curves and maps.

The sample station is “Sauce Viejo Aero” (id 873710) which is located at 31.7S, -60.82W, in Santa Fe Province, Argentina. The rainfall threshold is set to 0.25mm, coincident with the minimum measurable precipitation of most rain gauges (bucket type). The availability threshold is set to 523 days, it is more than 80% of the 628 days analyzed. Stations with more than 105 undefined daily values are discarded.

For the sample station, a contingency table per day in advance of forecast is presented in Table 2 (from 2a to 2g). In general, tables shows that successes in forecasts (hits+correct negatives) almost triple (2.82 in the worse case) the amount of failures (misses+false alarms). Among the failures, the most common are false alarms, which exceed the number of hits when the forecasts is 5 days prior to observations or more. Observed rainy days are not predicted (misses) in about 39% with 7 days of anticipation to 26% 1 day prior its occurrence. Note that, according to what is expected, the successes tend to increase and failures tend to decrease with the forecast time (when the observed day is closer to the forecast issue date).

The observed temperature during a day for the sample station is compared against the 7 forecasts of temperature for the same day in Fig. 1 with scatter plots (central column). As extra information are also added the scatter plots of daily minimum temperature (left column) and daily maximum temperature (right column). It is a common practice when working with model forecasts to define the minimum and maximum values from the 3-hr forecasts, therefore they cannot define exactly the precise moment of the occurrence (a 1.5 hr uncertainty is present). This uncertainty leads to more scattered points and lower correlation for minimum and maximum T when they are compared with mean T. Also, it is noted that as forecasts approach the observed date (looking at the pictures from bottom to top), the points tend to cluster together and the correlation coefficient improves. The coefficient are always greater than 0.8 for minimum and maximum temperature and greater or equal than 0.88 for mean temperature, reaching a correlation of 0.97 with 1 day in advance in this case.

Table 2 – Contingency Tables from (a) 1 days in advance forecasts to (g) 7 days in advance forecasts.

		OBS		
		YES	NO	TOT
FCST (-1)	YES	103	87	190
	NO	37	379	416
	TOT	140	466	606

(a)

		OBS		
		YES	NO	TOT
FCST (-2)	YES	106	91	197
	NO	36	371	407
	TOT	142	462	604

(b)

		OBS		
		YES	NO	TOT
FCST (-3)	YES	105	82	187
	NO	35	382	417
	TOT	140	464	604

(c)

		OBS		
		YES	NO	TOT
FCST (-4)	YES	98	91	189
	NO	42	373	415
	TOT	140	464	604

(d)

		OBS		
		YES	NO	TOT
FCST (-5)	YES	100	109	209
	NO	41	355	396
	TOT	141	464	605

(e)

		OBS		
		YES	NO	TOT
FCST (-6)	YES	94	103	197
	NO	45	363	408
	TOT	139	466	605

(f)

		OBS		
		YES	NO	TOT
FCST (-7)	YES	83	105	188
	NO	52	360	412
	TOT	135	465	600

(g)

As noted above, the positive results presented for forecast P and T takes into account just one sample station. To make a complete verification, all stations in the domain accomplishing the availability threshold are verified with contingency tables and scatter plots. The results are summarized in Fig. 2 for P and Fig. 3 for T.

The map of FCST accuracy for P (Fig. 2a) that averages the 7 days of forecasts evaluated in 112 gauge stations shows the lowest results in the northern part of the domain (mainly in Bolivia) with accuracies of about 0.6, but the forecast performance improves toward the southern part of the domain with accuracies up to 0.7. Notably, the best results are found in Chile and west Argentina where the topography implies an extra issue for the model. The accuracies values vary in this region from 0.8 to 1 indicating that more than the 80% of rainy or not rainy days were correctly forecast. The good performance in this region could be the result that resolutions considered in our model configuration satisfactorily represent the role of the strong local forcing of the topography on the region's precipitation regimes. Focusing on the LPB region the accuracies varies from 0.5 in the Paraguay and Mid-Upper Paraná sub-basins to 0.8 in the Lower Paraná and Uruguay sub-basins. Fig. 2b presents the skill evolution when the forecasts date moves away from the observed date. The curves average the results of all valid stations for LPB adding up 63 out of 112. The accuracy (black curve) shows almost no changes in time, with values around 0.7. That is, the model hits at least the 70% of the occurrence of rainy or dry days for LPB. The curve of POD starts with a value of 0.75 for 1 days in advance and decreases to 0.6 for 7 days in advance. In other words, the percent of observed rainy days that were correctly forecast varies from 60 percent with 7 days of anticipation to 75% with one day of anticipation. Finally, FAR (the orange curve) indicates that around the 60% of the forecast rainy days were false alarms (did not occur). This is also reflected in Table 2 for the sample station.

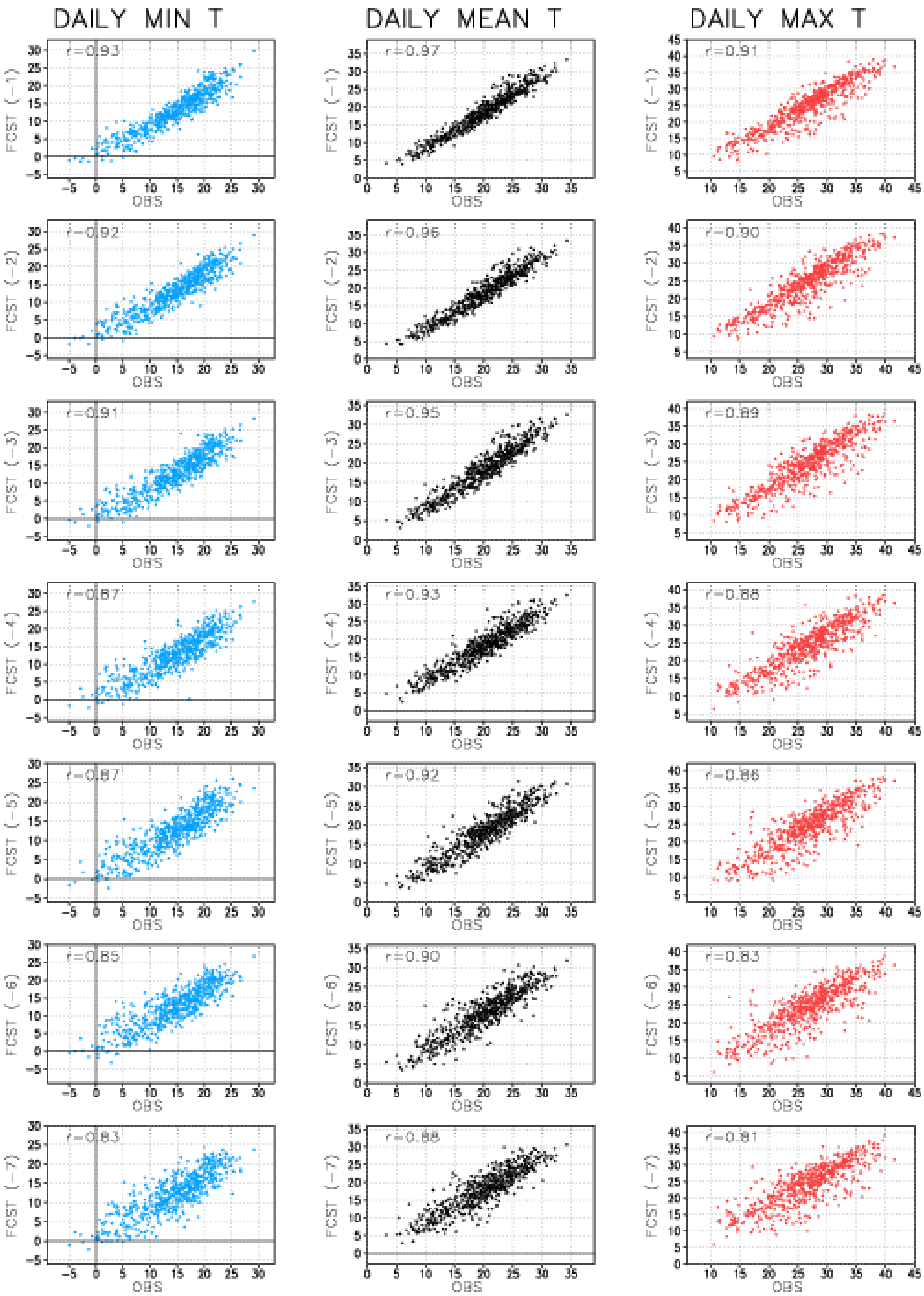

Figure 1 – Scatter plots for daily minimum temperature (left column), mean temperature (central column) and maximum temperature (right column) comparing forecasted values with observations in the sample station. Each row of scatter plots corresponds to each day in advance of forecasts (from 1 to 7) as indicated by FCST (-days in advance). Also, the correlation is showed in the top left corner of each scatter plot.

In terms of T, the correlation map (Fig. 3a), which is a mean in time of the 7 forecasts in the nearest grid points of 171 stations, shows an increasing correlation from top to bottom, i.e., from the north with correlation coefficients around 0.7 to the southern domain with correlation coefficients above 0.9. These results agree with the verification of P were the forecasts has better results towards the south. Specifically, the LPB (evaluated in 99 stations) has an almost homogeneous correlation of 0.9 or more for the southern sub-basins (Lower Paraná and Uruguay). Then, the Paraguay and the Mid-Upper sub-basins have most values varying from 0.6 to 1. According to the map it is found that the forecasts for LPB temperature have a correspondence with observations of more than 80% in general, with few exceptions in the northern border of the basin. From the point of view of forecast time (Fig. 3b), the correlation is of about 0.9 for FCST (-1) and decreases to 0.8 for FCST (-7). The correlation shows the correspondence between the model and observations, i.e., how well changes in observations are replicated by forecasts when the correlation coefficient is high. The magnitude of the forecast bias is shown in Fig. 3b (blue line). It indicates that the mean absolute error varies between 2°C and 3°C for LPB.

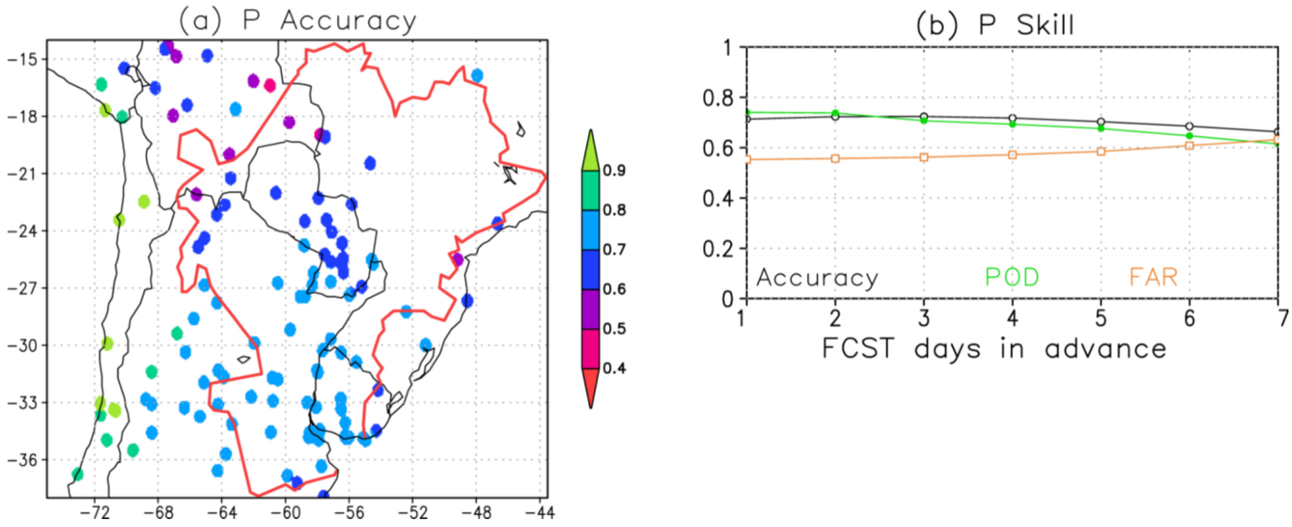

Figure 2 – Forecasts skill for P. The map (a) averages in time (from 1 to 7 days in advance) the accuracy of FCST with respect to observations for each station. The times-series (b) averages in space over LPB (highlighted in red line in panel a) the accuracy, the probability of detection and the false alarm ratio of all forecasts grid points under evaluation, i.e. points inside red line in panel a.

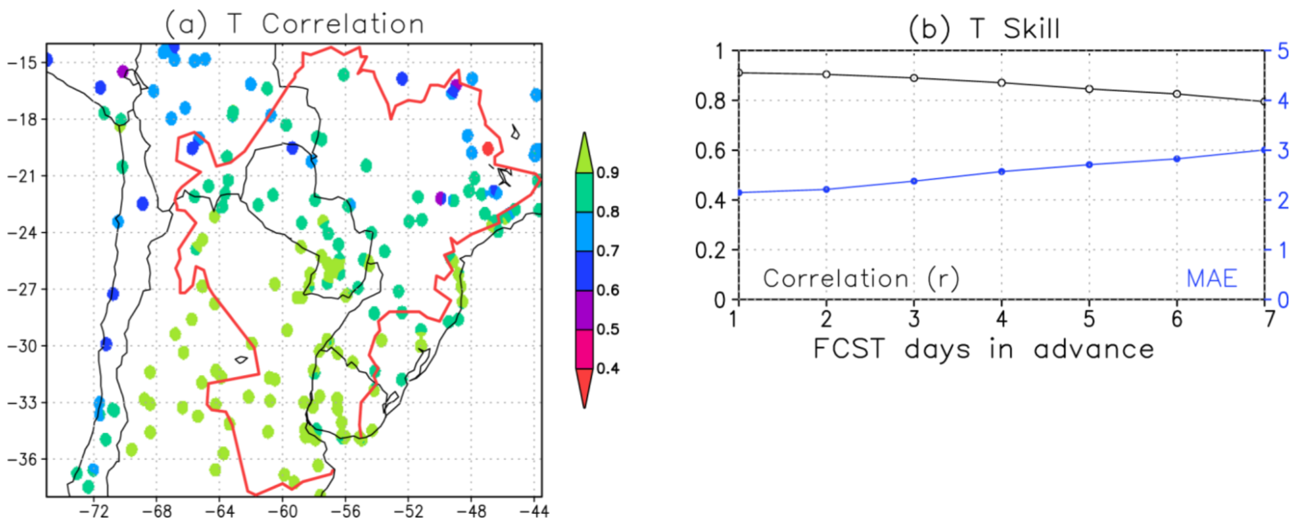

Figure 3 – Forecasts skill for T. The map (a) averages in time (from 1 to 7 days in advance) the correlation of FCST with respect to observations for each station. The times-series (b) averages in space over LPB (highlighted in red line in panel a) the correlation, and the mean absolute error of all forecasts grid points under evaluation, i.e. points inside red line in panel a.

4. CONCLUSIONS

Weather information is usually claimed and appreciated by different stakeholders in decisionmaking. In this work a comprehensive forecasting and monitoring system is presented and evaluated. The set of tools in the system allow an easy characterization and monitoring of the recent months climate, and predicts several days in advance, helping to plan activities, and reduce adverse impacts of anomalous conditions.

All system tools are originated by daily simulations of 7 days. Then, the results of routine model simulations were evaluated against gauge station data for two variables: the occurrence of daily precipitation and mean daily temperature. The forecasts of both variables show a good performance for the basin with better results in the southern part of the basin. In particular, the occurrence of daily precipitation is predicted with an accuracy of about 70% average (in time and space) for LPB. The model tends to generate false alarms of rainy days: Although for most applications a high amount of missed rainy events represents the major problem, a high rate of false alarms could be considered a weakness of the system depending on the use of the information. On the other hand, the forecasts of daily mean temperatures has an average correlation (in time and space) of about 85% with observations and present a mean absolute error of about 2.5°C. Note that the skill of the forecasts has a slow decrease with the increasing forecast time. This feature, observed for P and T, suggests that the period could be extended for more than 7 days of simulation (but remaining within the ranges of predictability of the dynamical system)

The system described in this article are not official forecasts of any institution and therefore do not carry any liability. Yet, It is noted that the system output is used as input of an early warning system for flows by the Instituto de Hidrología de Llanuras "Dr. Eduardo Usunoff" of UNICEN. Also, the system is used to support the daily weather report and extended forecasts of the Centro de Informaciones Meteorologicas of FICH, UNL. In both cases the users value positively the utility of the system and recognize it as a key tool for their applications. The system was recently developed and has less than two years on line. It is expected that the current errors could be reduced in the future applying bias correction methods. The future inclusion of new tools such as standardized precipitation index will facilitate the interpretation of the monitored variables.

ACKNOWLEDGEMENTS

This work was carried out with the aid of the project CAI+D 2011 "Previsión climática e hidrológica a diferentes escalas espaciales" supported by UNL and a grant from the Inter-American Institute for Global Change Research (IAI) CRN3035 which is supported by the US National Science Foundation.

REFERENCES

García Skabar, Y., Vidal, L., Salio, P. and Nicolini, M., 2011. Experimental high-resolution forecast in a region of Argentina. Working group in Numerical Experimentation (WGNE) Research Activities in Atmospheric and Oceanic Modelling (Blue book), 5: 09-10.

Giorgi, F., 2002. Variability and trends of sub-continental scale surface climate in the twentieth century. Part I: observations. Climate Dynamics, 18: 675-691.

Saha, S., et. al., 2010. The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., 91(8): 1015-1057.

Stone, R.C. and Meinke, H., 2005: Operational seasonal forecasting of crop performance. Phil. Trans. R. Soc. B, 360: 2109-2124.

Zipser, E.J., Cecil, D.J., Liu, C., Nesbitt, S.W. and Yorty, D.P., 2006: Where are the most intense thunderstorms on Earth? Bull. Of the Am. Met. Soc.: 1057-1071.