IMPACT OF ENRICHED ANALYSES ON REGIONAL NUMERICAL FORECASTS OVER SOUTHEASTERN SOUTH AMERICA DURING SALLJEX

YANINA GARCÍA SKABAR[1,2,3] Y MATILDE NICOLINI[3,4,5]

[1]Servicio Meteorológico Nacional, Buenos Aires, Argentina.
[2]CONICET, Buenos Aires, Argentina
[3]Instituto Franco Argentino sobre Estudios del Clima y sus Impactos (UMI), Buenos Aires, Argentina.
[4]Universidad de Buenos Aires (UBA), Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos, Buenos Aires, Argentina
[5]Centro de Investigaciones del Mar y la Atmósfera, CONICET, UBA, Buenos Aires, Argentina.

yanina@smn.gov.ar; nicolini@cima.fcen.uba.ar

ABSTRACT

During the warm season 2002-2003, the South American Low-Level Jet Experiment (SALLJEX) was carried out in southeastern South America. Taking advantage of the unique database collected in the region, a set of analyses is generated for the SALLJEX period assimilating all available data. The spatial and temporal resolution of this new set of analyses is higher than that of analyses available up to present for southeastern South America. The aim of this paper is to determine the impact of assimilating data into initial fields on mesoscale forecasts in the region, using the Brazilian Regional Atmospheric Modeling System (BRAMS) with particular emphasis on the South American Low-Level Jet (SALLJ) structure and on rainfall forecasts. For most variables, using analyses with data assimilated as initial fields has positive effects on short term forecast. Such effect is greater in wind variables, but not significant in forecasts longer than 24 hours. In particular, data assimilation does not improve forecasts of 24-hour accumulated rainfall, but it has slight positive effects on accumulated rainfall between 6 and 12 forecast hours.

As the main focus is on the representation of the SALLJ, the effect of data assimilation in its forecast was explored. Results show that SALLJ is fairly predictable however assimilating additional observation data has small impact on the forecast of SALLJ timing and intensity. The strength of the SALLJ is underestimated independently of data assimilation. However, Root mean square error (RMSE) and BIAS values reveal the positive effect of data assimilation up to 18-hours forecasts with a greater impact near higher topography.

Keywords: Data assimilation; numerical forecasts, SALLJEX

RESUMO: IMPACTO DA ANALISES ENRIQUECIDAS SOBRE PREVISÕES NUMÉRICAS SOBRE O SUDESTE DA AMÉRICA DO SUL DURANTE O SALLJEX

Durante a temporada quente de 2002-2003 foi realizado o *South American Low-Level Jet Experiment* (SALLJEX) no sudeste da América do Sul. Aproveitando-se da base de dados única coletados na região, um conjunto de análises foi gerado para o período SALLJEX, assimilando todos os dados disponíveis. A resolução espacial e temporal deste novo conjunto de análises é maior do que a das análises disponíveis até hoje para o sudeste da América do Sul. O objetivo deste trabalho é determinar o impacto de assimilação de dados em campos iniciais das previsões de mesoescala na região, utilizando o *Brazilian Regional Atmospheric Modeling System* (BRAMS), com ênfase particular na estrutura do Jato em Baixos Níveis na América do Sul (LLJSA) e nas previsões de precipitação. Para a maioria das variáveis, o efeito de utilizar as análises numéricas com dados assimilados nos campos iniciais é positivo nas previsões de curto prazo. Esse efeito é maior nas variáveis de vento, porém não é significativo nas previsões superiores a 24 horas. Em particular, a assimilação de dados não

melhora as previsões de 24 horas da precipitação acumulada, mas tem pequenos efeitos positivos na precipitação acumulada entre 6 e 12 horas de previsão.

Como o foco principal é a representação do LLJSA, o efeito de assimilação de dados na sua previsão foi explorado. Os resultados mostram que LLJSA é bastante previsível, entretanto quando dados observados adicionais são assimilados, o impacto é pequeno na previsão durante o tempo de integração e na previsão da intensidade do LLJSA. A intensidade do LLJSA é subestimada independente da assimilação de dados. No entanto, o erro quadrático médio (RMSE) e os valores do Viés revelam os efeitos positivos da assimilação de dados até 18 horas de previsão, com maior impacto perto de topografia.

Palavras-Chave: assimilação de dados; previsão numérica;SALLJEX

1. INTRODUCTION

Human activities are affected by major mesoscale atmospheric phenomena which are not detected by synoptic-scale analyses or forecasts. Increasingly reliable and accurate forecasts are needed for different activities. Consequently, the accuracy of numerical rainfall forecasts is an important challenge for meteorological prediction. Numerical forecast models fail to predict rainfall intensity and the spatial and temporal distribution of precipitating systems. So far the international scientific community has not been able to solve this problem and research continues, mainly using two basic tools: observations and numerical modeling.

Regional and mesoscale numerical forecast models have been extensively used in many countries to support research, and spatial and temporal resolution is being constantly improved (Weisman et al, 2008; Kain et al, 2008; Pielke, 2002; Warner et al., 1989; among others). In southern South America, several experiments suggest that the use of this kind of models represents a step forward in simulating and forecasting mesoscale phenomena (Saulo et al, 2007; Nicolini and Saulo, 2006; Salio et al., 2006; Paegle et al., 2004; Nicolini et al., 2002; among others). However, deficiencies are observed, particularly in rainfall forecast. Saulo et al. (2001) and Saulo and Ferreira (2003), identified rainfall deficits in the forecasts for the northern part of the La Plata Basin. In addition, the scarcity of data makes it difficult to assess model performance.

An important aspect for further studies is the impact of assimilating data into the initial fields on the quality of numerical model forecasts (Harms et al., 1992; Warner et al., 1989). In the last years, mesoscale case studies were performed in different regions in the world, evaluating the impact on forecasts of assimilating different datasets such as rawinsonde observations, satellite data, surface data and special measurements made during field campaigns (Cherubini et al., 2006; Cucurull et al., 2004; Barker et al., 2004; Davolio and Buzzi, 2004; Faccani et al., 2003). In all cases, results indicate that models perform better when an enriched initial field is used, although the impact is not always as good as expected or it is not observed on all the variables.

Enhancement of the observational network particularly in southeastern South America, is crucial because there are few regular observation stations. In Argentina, for instance, the current rawinsonde network is composed of 5 stations where only two observations a day are performed. A larger number of data would allow to generate a set of analyses to represent more accurately the state of the atmosphere, and to be used as initial condition in forecast models while making it possible to improve model performance assessments. Consequently, during the warm season 2002-2003, the South American Low-Level Jet Experiment (SALLJEX) was carried out in southeastern South America, aimed at gathering a set of data that would contribute to advance documentation and characterization of the low level jet east of the Andes (SALLJ) as well as the associated rainfall in the region. The operational rainfall network and the spatial and temporal density of rawinsonde and pilot balloon observations were enhanced during SALLJEX. A description of the data obtained during SALLJEX can be found in Vera et al. (2006). A set of analyses was generated using that database and the Brazilian Regional Atmospheric Modeling System (BRAMS) assimilating all available data. The horizontal resolution achieved is 20 km and the temporal resolution is 3 hours, as described in García Skabar and Nicolini (2009).

The impact of data assimilation on forecasts for the period 15 - 24 January 2003 and using SALLJEX data were examined by Cavalcanti and Herdies (2004). Those authors assimilated only the special SALLJEX rawinsonde observations. Pilots balloons were not included in that study. They applied the data assimilation system Global Physical-space Statistical Analysis System (GPSAS) with horizontal resolution of 100km. The authors assessed the impact of data assimilation on the analyses and forecasts. They found that both, analyses and forecasts have a positive impact on the representation of vertical structure of humidity and meridional wind component in a site close to Santa Cruz de la Sierra, Bolivia, matching rawinsonde observations better after data assimilation.

As to the impact of data assimilation on the BRAMS model forecasts, Hallak and Silva Dias (1996) found that the assimilation cycle of additional rawinsonde observations and surface data significantly improved model performance

by resolving the spatial dimensions of mesoscale systems. In that case study authors assimilated additional data every 6 hours. Another case study in the region also determined that assimilating the data available from the operational network and observations from seven pilot balloons from the PACS/SONET network into NCEP reanalyses was not enough to achieve satisfactory improvements in BRAMS rainfall forecasts (Nicolini et al, 2002).

In southeastern South America, studies on the impact of data assimilation on forecasts have been restricted to case studies or short periods. Previous effort was focused on generating a set of enriched analyses in which mesoscale data obtained during SALLJEX had been assimilated. Present work addresses how sensitive are BRAMS model forecasts to data assimilation instead. The main objective of this paper is to determine the impact of using enriched analyses as initial condition on BRAMS model forecast. Special emphasis is placed on rainfall and SALLJ structure forecasts. Results were evaluated in the three-month period of the field experiment.

Section 2 describes the data and methodology. Section 3 describes the verification methodology applied and discusses the results. Finally, conclusions are included in Section 4.

2. DATA AND METHODOLOGY

2.1 Methodology for analysis and data assimilation

In this section a brief description of the methodology applied to generate enriched analyses is presented. A more detailed description of observations included and methodology used to generate the different sets of analyses can be found in García Skabar and Nicolini (2009).

The SALLJEX was conducted from 15 November 2002 to 15 February 2003, deploying a dense rain-gauge network and enhancing the spatial and temporal density of rawinsonde (RAOBS) and pilot balloon observations (PAOBS) in southeastern South America. The location of observational stations is shown in Figure 1 and Table 1 presents the number of assimilated data from each of the subsets in the entire period.

Enriched analyses were generated using version 3.2 of the BRAMS model, that was developed originally at the University of Colorado (Cotton et al., 2003), and modified jointly by ATMET (Atmospheric, Meteorological and Environmental Technologies), IME-IAG/USP (Institute of Mathematics and Statistics -Institute of Astronomy, Geophysics and Atmospheric
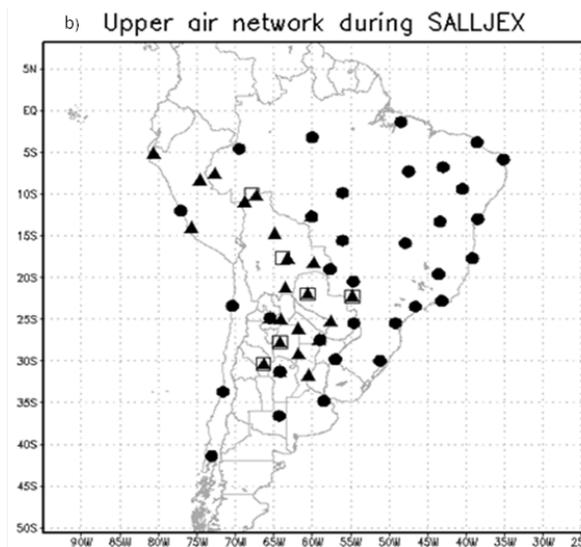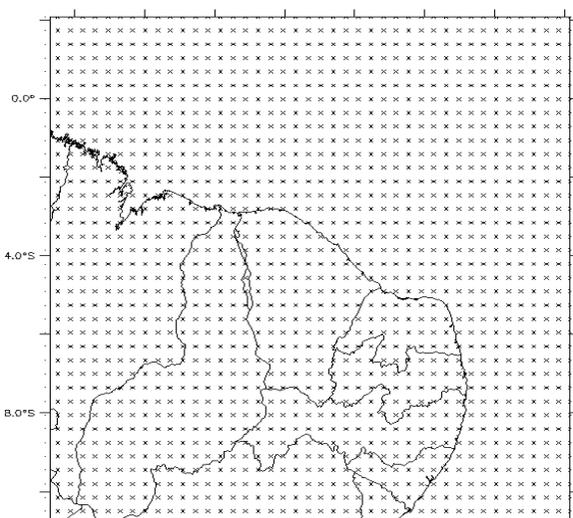
Figure 1 - Spatial distribution of assimilated observations.a) Surface network b)Upper air network during SALLJEX; circles: operational rawisondes;square:SALLJEX rawisondes and triangles:SALLJEX pilot ballons. (Adapted from Figure 2, Garcia Skabar and Nicolini,2009)

Table 1 - Number of assimilated observations. Adapted from Table 1, García Skabar and Nicolini(2009).

Data Source	Nº of data during the whole SALLJEX period
Operational surface network	329000
Operational rawisonde network	2850
SALLJEX rawisonde	815
SALLJEX Pilot balloons	3880

Sciences /University of Sao Paulo (USP)), and Weather Forecast and Climate Study Center (CPTEC). Specific characteristics of the model can be found in Freitas et al (2009) and on the BRAMS website (http://brams.cptec.inpe.br/).

The analyses were generated for the entire SALLJEX period starting from 13 November 2002. Soil moisture information was updated daily and sea surface temperature, weekly. The process was developed uninterruptedly to allow analyses keep memory of previously assimilated data.

Operational analyses generated by the Global Data Assimilation System (GDAS) of the National Oceanic and Atmospheric Administration/ National Centers for Environmental Prediction (NOAA/NCEP) with 26 vertical levels, 1º horizontal resolution and 6-hour temporal resolution were used as first guess analyses. In a first step, a set of analyses was generated for the entire period without assimilating data to achieve 3-hourly analyses -hereafter called ANASD. These analyses have higher spatial and temporal resolution than GDAS. In a second step, ANASD analyses were used as first guess and all available data were assimilated. The set of analyses obtained was called ANACD. While the ANASD were obtained from GDAS analyses; the ANACD were generated using ANASD as first guess and modified with all the available data.

The process resulted in two 3-hourly analysis sets (ANASD and ANACD) covering the 3 months period. Resolution was 80 km for almost all South America and 20 km resolution in the center and north of Argentina and Chile, Uruguay, Paraguay and the south of Bolivia and Brazil (Figure 2).

2.2 Forecast setup

The model setup was defined after several numerical tests with different physical options for specific cases. In particular, the same two nested grids that were used to generate analyses were applied to obtain the forecasts (Figure 2). 72 hour forecasts were performed, starting at 1200 UTC every day, for the period between 15 November 2002 and 15 February 2003 using vertical coordinate shaved-eta (Adcroft et al,1997). Nine soil levels are used as well as 29 vertical levels in the atmosphere, with 18 levels in the first 12 km and the top of the model located at approximately 23 km.. At both resolutions, parameterizations were activated for both shallow (Souza and Silva, 2002) and deep convection. Deep convection was represented by Grell's parameterization and closure (Grell and Devenyi. 2002). The "bulk water" scheme was used to represent the microphysics of 8 water species. Radiation was parameterized as described in Chen and Cotton (1983). Vertical and horizontal sub-grid turbulence was simulated through the schemes of Mellor and Yamada (1982) and Smagorinsky (1963) respectively. Topographic data with a resolution of

about 1km provided by the United States Geological Survey (USGS); land use and vegetation cover data with a resolution of 1km from the International Geosphere-Biosphere Programme (http://edcdaac.usgs.gov/glccc) and soil types with spatial resolution of 5 km from FAO/UNESCO (http://daac.gsfs.nasa. gov) were fed into the model. Other data used were weekly sea surface temperature fields provided by NCEP version OI.v2 with a resolution of 100 km and described in Reynolds et al. (2002); daily soil moisture fields from USP (Gevaerd and Freitas, 2006). Boundary conditions were taken from Global Forecast System (GFS) of NOAA/NCEP every 12 hours.

Two forecast sets were obtained for the whole SALLJEX period using the same setup but different initial conditions. One with no assimilated data (ANASD), where only interpolation from GDAS analysis was performed, and the other with the set of enriched analyses, that assimilated all surface and upper-level data from the operational network, as well as rawinsonde and special pilot balloon observations from SALLJEX (ANACD).

3. FORECAST VERIFICATION

### 3.1 Temperature, humidity and wind

3.1.1 Methodology

The statistics used to analyze the difference between forecasts are the root mean square error (RMSE) and the bias (BIAS), which are defined as:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} \left(X_f^i - X_o^i \right)^2 \right]^{1/2}$$

$$BIAS = \frac{1}{N} \sum_{i=1}^{N} \left(X_f^i - X_o^i \right)$$

where N is the total number of observations, subscript f indicates forecast and *o,* observed. X, a variable. RMSE and BIAS values were calculated for the following variables: temperature, moisture and zonal and meridional wind components at 15 fixed vertical levels (100, 300, 600, 900, 1200, 1700, 2300, 2900, 3700, 4600, 5600, 6700, 8700, 10700 and 12700 m.a.m.s.l.) and at the surface. Verifications were made on the highest resolution domain.

Forecast BIASes and RMSEs are calculated respect to observations and ANACD analyses. In the case of observations, for each variable and level, calculations using all the available data in the period resulted in one RMSE and one BIAS value. For each observed vertical profile the closest grid point was determined to generate a predicted profile. Each statistic value was obtained out of a maximum of about 800 data. However, the number of data varies depending on the verification
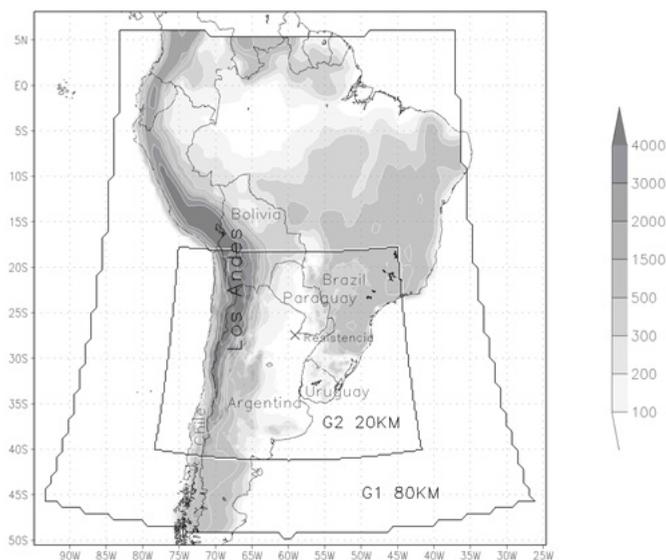
Figure 2 - Nested domains. Topography is shaded and indicated in meters. (Adapted from Figure 2, Garcia Skabar and Nicolini,2009).

time and the vertical level, particularly in the case of pilot balloon observations. In the case of ANACD, the number of data involved in each statistic value was quite different and calculations were made at each point of the highest resolution grid (91x86 points). One RMSE and BIAS value was obtained for each variable, each level and each analysis time; all these were then averaged over the period. For statistics calculated at the grid points, both forecasts –those initialized with ANASD and ANACD- were compared with ANACD analyses.

Statistics were calculated separately for pilot balloon and rawinsonde observations. Verification against observed data was made for the time when the greatest number of data was available. For rawinsonde observations, these times are 0000 and 1200 UTC, which means that forecasts were evaluated at 12, 24, 36, 48, 60 and 72 hours. The number of data available varies significantly between 0000 and 1200 UTC, with about 200 rawinsonde observations at 0000UTC and 600 at 1200UTC. Pilot balloon observations were considered at 2100 UTC instead of 0000 UTC. This means that forecasts were verified at 9, 33, and 57 hours instead of 12, 36 and 60 hours, with about 700 observations at 1200 UTC, and 850 at 2100 UTC. In the case of comparisons with enriched analyses, the verification was made every 3 hours, at all available times.

It is worth to mention that we compare forecasts with analysis or data that were not included as initial condition in the forecast, as they correspond to a later time than those used as initial condition for the specific time forecast we are assessing. Anyway, the use as reference of the same set of analyses that were used as initial condition in the forecast could be generating some minor benefits for the experiment with data assimilation.

3.1.2 Verification of vertical profiles against enriched analyses and observations

Figures 3 and 4 show the RMSE and BIAS vertical profiles respectively, for temperature, water vapor mixing ratio and wind components, compared with the ANACD analyses and the observations. Only the results obtained for 12-hour forecasts against ANACD analyses and rawinsonde observations that verified at 0000 UTC are shown. In the case of the comparison with pilot balloons, results are shown for 9 and 33-hour forecasts that verified at 2100 UTC.

Forecast verification respect to enriched analysis shows that initial fields with assimilated data have positive impacts on RMSE and BIAS values. Such effect can be observed in the wind field and it is not so evident in thermodynamic variables (Fig.2a and 3a) and is most observable in 12-hour forecasts.

The comparison of RMSE and BIAS vertical profiles with rawinsonde observations (Fig. 3b and 4b), showed no significant effects of data assimilation on any of the variables. The same exercise done with pilot balloon observations showed smaller RMSEs in 9 hour forecasts after data assimilation (Fig. 3c and 3d). As in the case of analyses, this difference also decreases with increasing forecast time. It is worth mentioning that in this case the positive effect of data assimilation on the meridional wind component can be observed even in 48 hour-forecasts. For all forecast times, the BIAS of both wind components is closer to zero when using enriched datasets (Fig. 4c and 4d).

For surface data (not shown), data assimilation only gave smaller wind variable RMSES in the first 6 forecast hours. No impact of data assimilation for thermodynamic variables was observed.
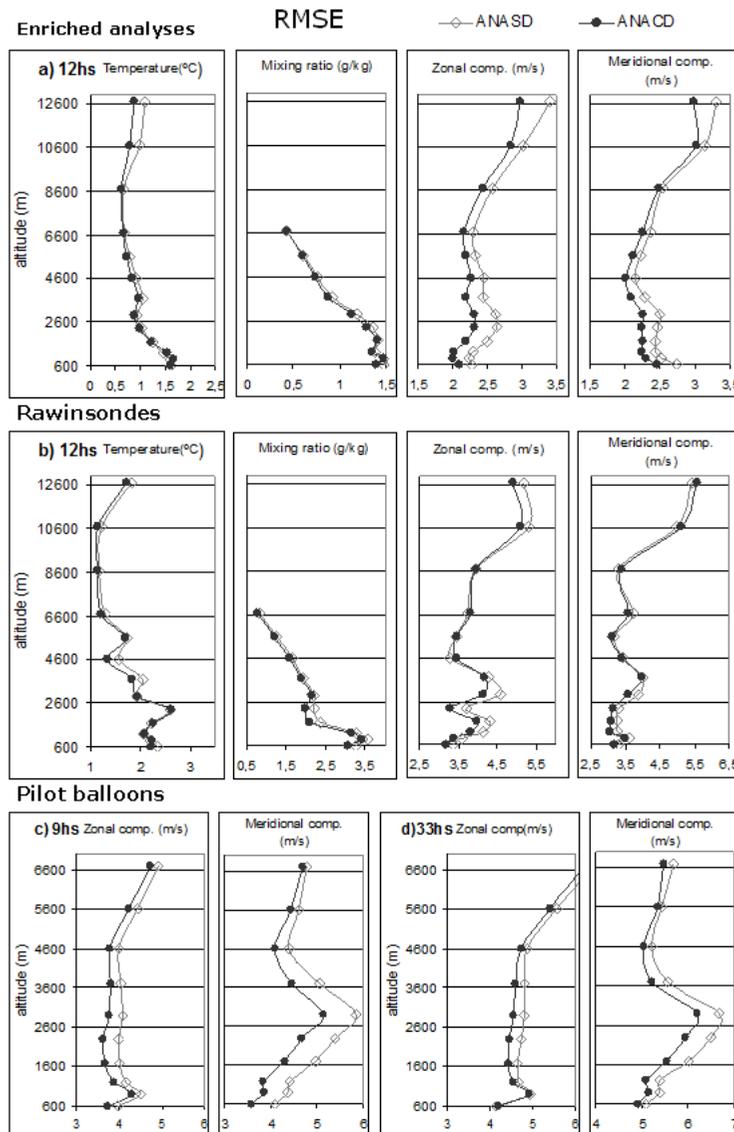
Figure 3 - RMSE vertical profiles for the SALLJEX period for 12-hour forecasts verified at 0000 UTC using different initial fields a) with respect to ANACD and b) with respect to rawinsonde observations. c) and d) 9 and 33-hour forecasts verified at 2100 UTC against pilot balloons.

BIAS profiles (Fig. 4) show that the model underestimates temperature at low-levels and overestimates it at upper levels with a consequent stabilization effect along the temperature profile, but this is observed in both forecasts. The mixing ratio is underestimated at levels above 2000 m. The BIAS of the zonal and meridional wind components shows that westerly winds are underestimated at all vertical levels. BIASes obtained for the meridional wind component reflect an underestimate of northerly wind below 7000m. This means that the strength of the SALLJ is underestimated by both model forecasts, but the underestimation is slightly smaller when initial fields with assimilated data are used.

Figure 5 shows the difference (%) between RMSE of forecasts initialized with ANACD and ANASD respect to

their average at different forecast times at 1700 and 2900m for temperature and the meridional wind component. The figure shows that the impact of data assimilation on the forecast decreases with increasing forecast time. After 24 hours this difference represents less than 5% of the error. The rapid weakening of the impact of assimilating data in the initial forecast field may be related to the size of the nested domain. If the horizontal extent of the domain was greater the impact would have been observed in longer forecast times. This may be related to propagation out of the domain after 24 hours of systems that have been better represented through data assimilation at the initial conditions.

All the comparisons show that data assimilation in the initial fields affects forecasts. The effect is greater on the wind
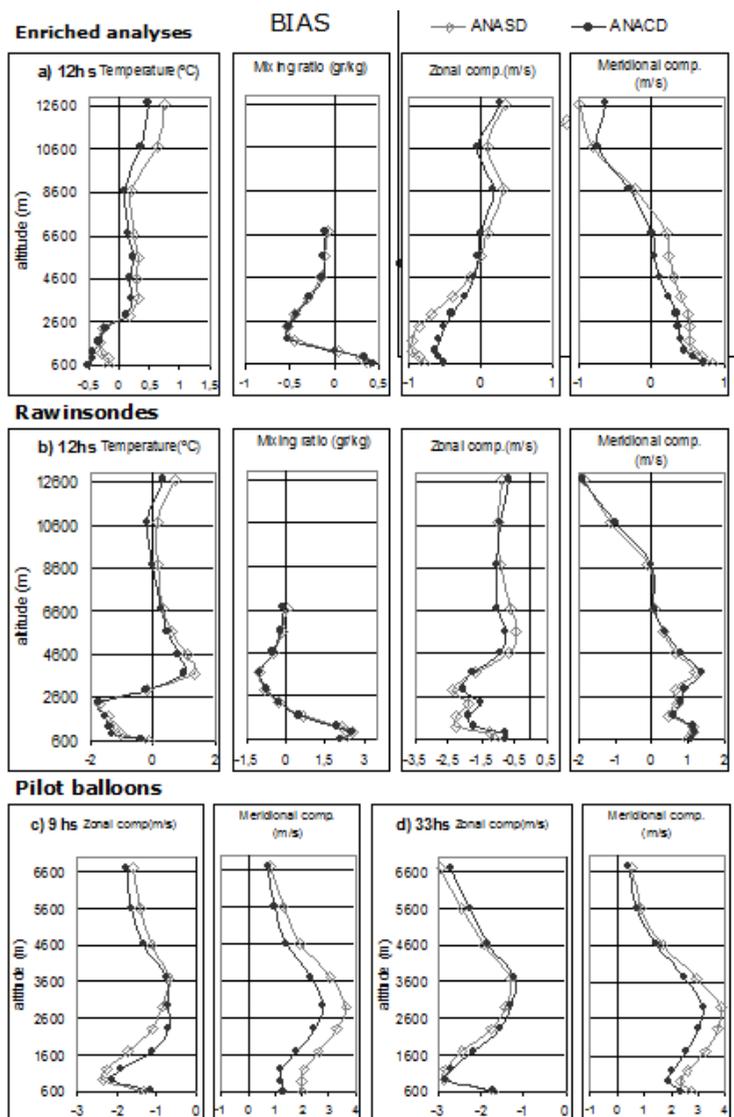
**Figure 4** - Idem Figure 3 but for BIAS vertical profiles.

field than on temperature and humidity. The impact is even greater when results are compared with pilot balloobn data, showing the benefits of data assimilation even in 48 hour-forecasts. This is partly because rawinsonde observations of the operational network had been assimilated in the GDAS analyses used as first guess fields, contrary to SALLJEX pilot balloon data which had not been previously assimilated.

3.1.3 SALLJ forecast verification

The SALLJ is the northerly low level jet that occurs just east of the Andes. It represents an effective mechanism of moisture transport from tropical to subtropical latitudes, and therefore has a fundamental role on rainfall in southeastern South America (Berbery and Barros, 2002). SALLJ occurrence

can be identified by criterion 1 of Bonner (Bonner, 1968), i.e., when maximum speed is greater than 12 m/s at low levels (850 or 925 hPa) and wind shear is greater than 6 m/s above the low-level maximum. SALLJ events have different characteristics, a description and classification of different cases can be found in Nicolini et al (2006).

To evaluate if SALLJ forecasts can be improved using enriched initial fields, the wind field was analyzed in greater detail at the levels with maximum values in the area covered by the high resolution domain. Hovmöller diagrams were made with the forecasts and the analyses at fixed latitudes for the different variables that describe the SALLJ. Figure 6 shows Hovmöller diagrams comparing the ANACD meridional wind component at 1700 m with 6 and 12 hours forecasts initialized with ANASD and ANACD at 20ºS latitude. The meridional wind
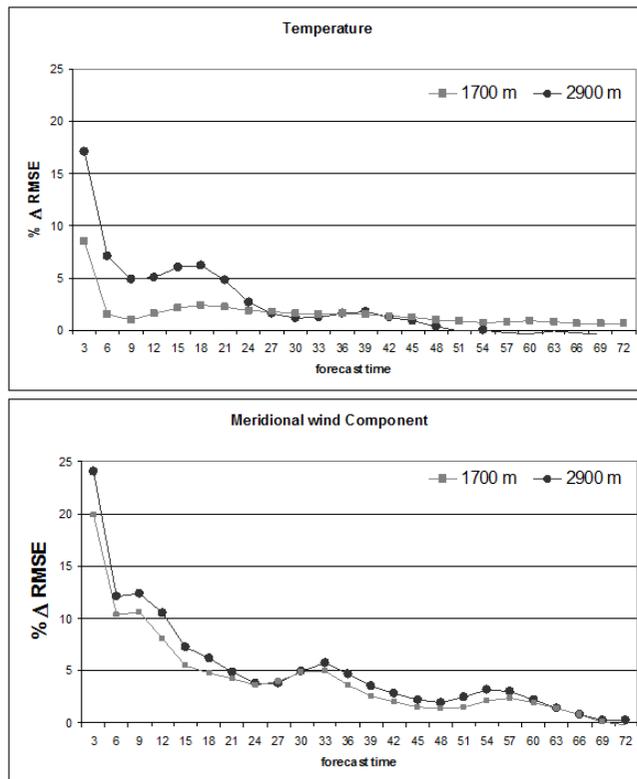
Figura 5 - Percentage of RMSE difference between forecast initialized with ANACD and ANASD respect RMSE average at different forecast times, at 1700 and 2900m.

component is satisfactorily represented in both forecasts, but the expected differences between them are almost not observable. The diagrams show that both forecasts follow satisfactorily the sequence of events with northerly and southerly winds. The peaks observed in the analyses can also be found in the forecasts. The speed of the meridional wind component is represented correctly in the forecasts. The comparison of the forecasts with and without assimilated data shows a slight positive impact of data assimilation during the first 6 forecast hours (Figures 6a-6c). As the forecast time advances, the differences decrease (Figures 6c and 6f). These results show that SALLJ is fairly predictable and that assimilating additional observations has small impact on the forecast of SALLJ timing and intensity.

To evaluate the effect of data assimilation throughout the domain, spatial distributions of RMSE and BIAS were also obtained between both forecasts and the enriched analyses. The fields were analyzed in the region located between 55 and 64º W and 32 and 20º S. Figure 7 shows RMSE and BIAS fields for the 12 hour-forecasts of the meridional wind component at 1700m. RMSE and BIAS differences between forecasts initialized with ANASD and ANACD reveal the positive effect of data assimilation in the entire field. A greater impact is observed in the west side of the domain, where topography is higher.

Eighteen-hour forecasts (Figure not shown) still show an impact of data assimilation, with some areas within the domain having smaller RMSE values. The impact is no longer observed in 24-hour forecasts (Figure not shown). Wind speed was analyzed in the same way, with similar results (Figure not shown).

During the first days of February an intense SALLJ event occurred, identified as a SALLJ Chaco according to definition of Nicolini and Saulo (2000), which reached subtropical latitudes up to at least 25ºS. At that time an intensive observation period was being carried out, with 3 or 4 rawinsondes and 8 pilot balloons per day at selected sites along the SALLJ axis. Figure 8 shows the evolution of 18 hours forecasts of the vertical wind profile in Resistencia from 3 to 8 February 2003 at 0600 UTC. Four profiles are observed in the figure, i.e., forecasts initialized with ANASD and ANACD, observations and ANACD. Forecasts initialized with enriched analyses better represent the SALLJ than forecasts initialized with ANASD, although all forecasts fail to predict the maximum observed SALLJ speed.

Based on these results, we find that the model is able to forecast SALLJ events, but it underestimates maximum wind speeds. Although the impact of data assimilation on the SALLJ related fields is weak, a positive impact can be observed in the first 24 hours, the most significant being the first 12 forecast hours. Considering that SALLJ speed peaks occur at night (Nicolini et al, 2006) and the short duration of the positive impact of data assimilation on the forecast, a more convenient forecast initialization time would be 0000 UTC instead of 1200 UTC.

3.2 RAINFALL

3.2.1 Methodology

The region selected to verify rainfall forecasts stretches from 53 to 63º W and 22º to 40º S (Figure 9). This domain was selected for the following reasons:
- It is completely covered by the highest resolution grid
- The rainfall database available in the region during the SALLJEX period has the greatest spatial density (about 1200 additional daily rain gauge observations).
- Intense rainfall associated with significant SALLJ events occurs frequently within this region.

To perform forecast verifications, data measured over an irregular observation network were converted into regularly distributed points. To do so, observations were averaged in 0.5º latitude by 0.5º longitude boxes within the domain. Boxes with missing data were assigned indefinite rainfall values and were not included in the verification. The percentage of days with indefinite values per box in the SALLJEX period is shown in Figure 9. This methodology differs from the interpolation

techniques usually applied to represent observed rainfall fields such as kriging (Davies, 1973), which smoothes rainfall and generates rainfall values in areas lacking observations. Thus, comparison is stricter since no verification is made at the points where data are not available, and no interpolations but just averages are used in the boxes containing data. Liebmann and Allured (2005) used a similar methodology to generate a gridded historical database in South America.

To perform the verification the data and forecasts need to be on a common grid . So, forecast rainfall values were averaged at the grid points included in each box.

Forecast rainfall was verified using the Equitable Treath Score (ETS) and BIASPP, statistics following the definition of Mesinger, 1996:

$$ETS = \frac{H - CH}{F + O - H - CH} \qquad \text{where} \qquad CH = \frac{FxO}{N}$$

$$BIASPP = \frac{F}{O}$$

where F is the number of boxes where rainfall is forecast, O is the number of boxes where it is observed, H is the number of boxes where it is both forecast and observed and N is the number of boxes with data in the entire verification domain. These statistics were used by Saulo and Ferreira (2003) to verify rainfall forecasts in southern South America.

Both the ETS and the BIASPP were calculated for eight rainfall thresholds (0.25, 2.50, 6.30, 12.7, 19.0, 25.4, 38.1 and 50.8 mm). Most data are accumulated rainfall in 24-hour periods –coinciding with the pluviometric day. Values obtained for each forecast were compared with rainfall persistence.

A quantitative verification of rainfall forecasts was made by generating fields of accumulated rainfall in the entire period for the observations as well as for the two forecasts. Forecasts of 24 hours accumulated rainfall were evaluated taking into account the first 24 hs, the period between 24 and 48 hours and between 48 and 72 hours.

Since the impact of enriched analysis on the forecasts is greater in the first forecast times, accumulated rainfall forecasts were assessed also in 6-hour periods. Raingauge observations were considered from 13 Argentine meteorological stations from the operational network of the National Weather Service which informs the amount of rainfall accumulated in 6 hours. The differences between model forecasts and observations are due to model errors and partly to the scale of observations. The values
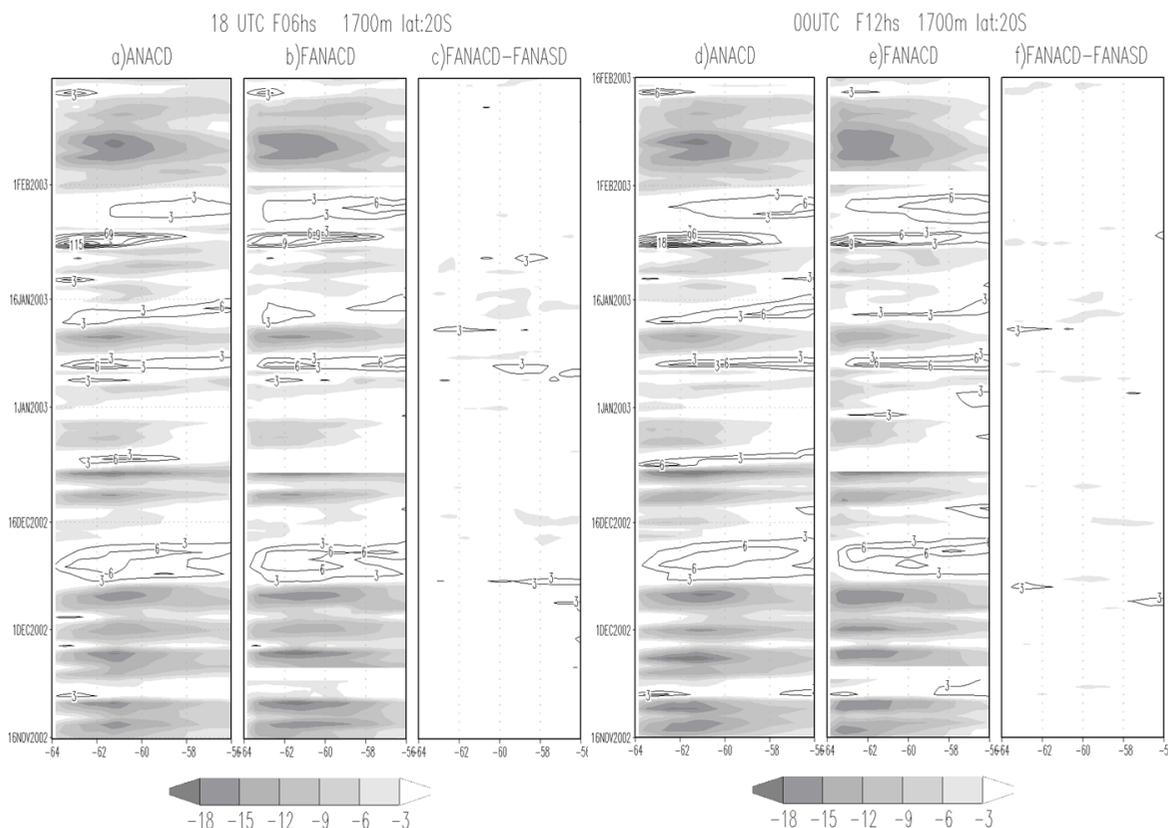
Figure 6 - Hovmöller diagrams for meridional wind component at 1700 m at 20º S from 56 to 64ºW. a), b) and c) for 1800 UTC, and 6 hours forecast and d), e) and f) for 0000 UTC, and 12 hours forecast. a) and d) from ANACD, b) and e) from forecast initialized with ANACD and c) and f) difference between forecasts initialized with ANACD and ANASD. (Negative values shaded and positive values contoured)
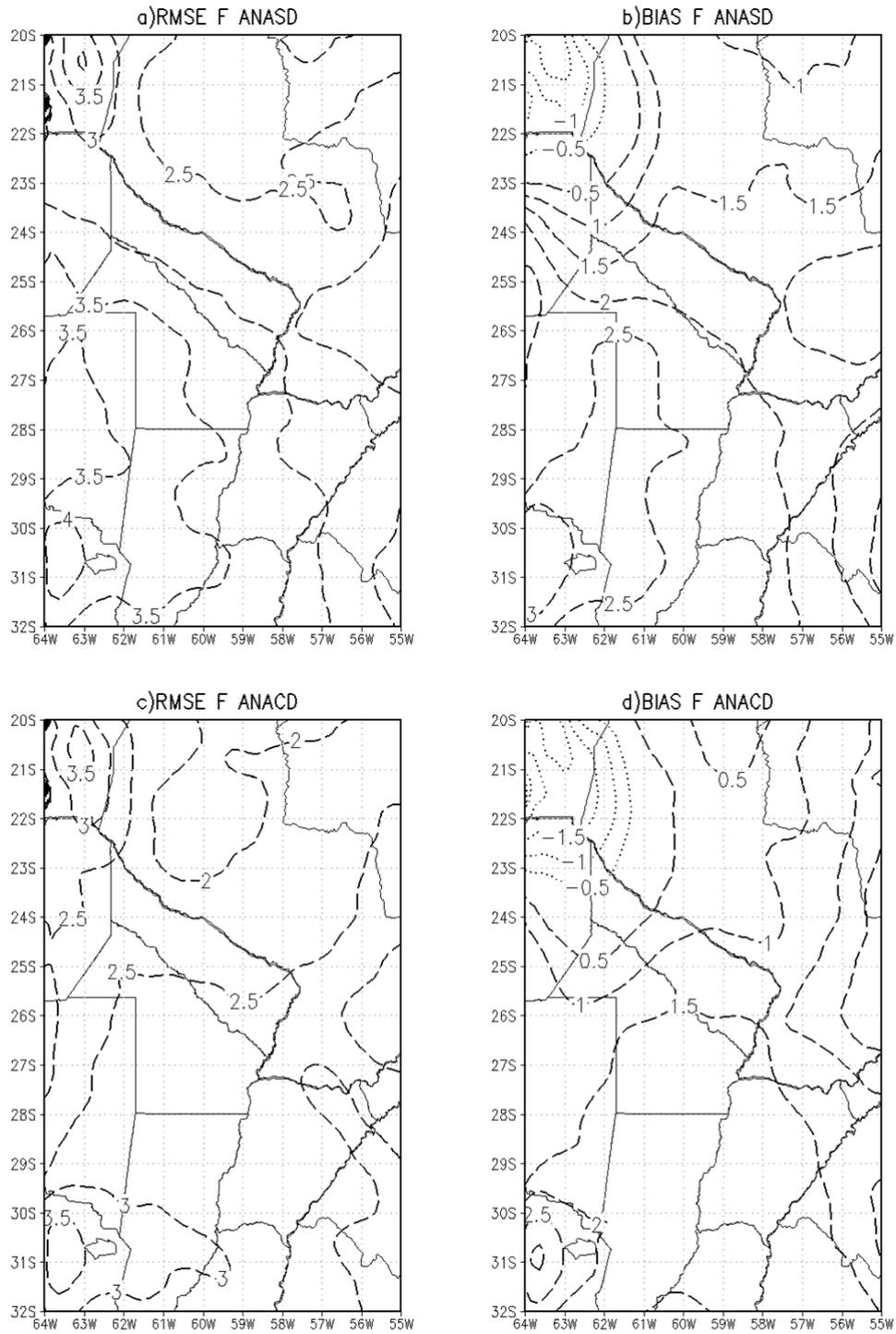
Figure 7 - RMSE and BIAS respect to enriched analyses of the meridional wind component at 1700m of 12 hour forecasts verified at 0000 UTC initialized with a) , b) ANASD, and c), d) ANACD. Contour interval is 0,5 m/s.
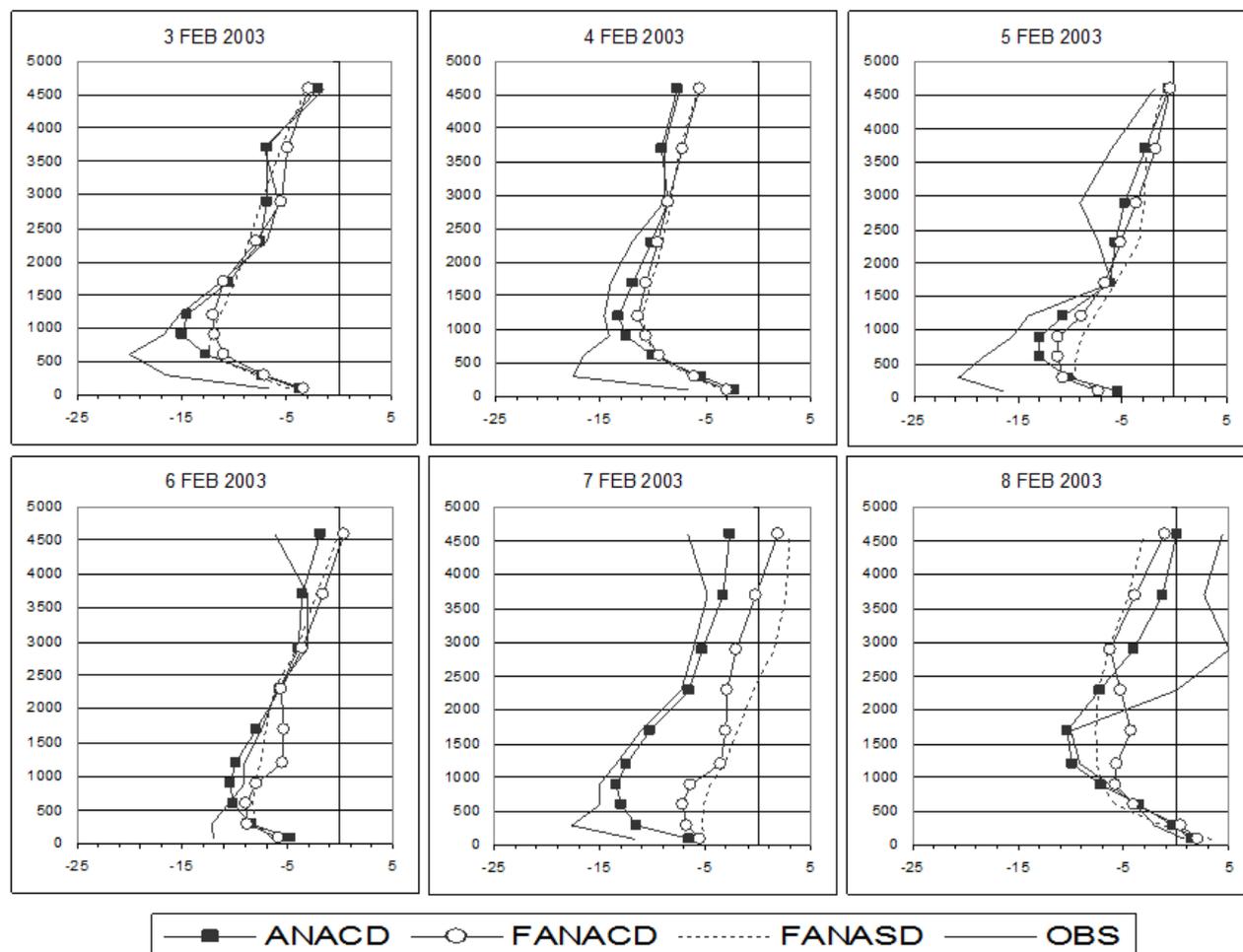
Figure 8 - Evolution of meridional wind component (m/s) vertical profile from 3 to 8 Feb. 2003 at Resistencia at 0600 UTC generated from ANACD, observations (OBS) and 18 hours forecast initialized with ANASD (FANASD) and ANACD (FANACD).

observed at a given station were compared with the averaged forecast of the 4 grid points closest to that meteorological station. The location of the stations is shown in Figure 9. RMSE and BIAS were calculated for 6-hour accumulated rainfall, in the 6, 12, 18 and 24-hour forecasts, only in the high-resolution grid.

3.2.2 Results

Figure 10 shows the values of ETS and BIASPP of 24- and 48-hour forecasts of 24 hour accumulated rainfall initialized with the different analyses in the SALLJEX period. The comparison of ETS values obtained for the different forecasts, in general, revealed that data assimilation did not improve model performance. Small positive impact of data assimilation on the 24- and 48-hour forecasts was only observed at low rainfall. For all forecast times, ETS values were greater than the persistence. As might be expected, ETS values decrease with increasing threshold and forecast time.

The analysis of BIASPP values shows that forecasts overestimate rainfall areas at low thresholds and underestimate them at high thresholds. In the 24-hour forecast BIASPP was smaller for all thresholds after data assimilation, which means that the forecast rainfall area is smaller when using an enriched initial field. This is good for low thresholds where forecasts overestimate the rainfall area, but at high thresholds using enriched initial fields results in even greater underestimates. The differences between the BIASPPs of forecasts initialized with and without enriched analyses are not observable in 48 or more hour forecasts.

Accumulated rainfall in the 92-day period is shown in Figure 11 for observations and for both 24-hour forecasts considered. In the regions with more data (Fig. 9), rainfall is underestimated in all forecasts and the underestimate increases after data assimilation, which agrees with the BIASSPP results. The same verification for rainfall accumulated between 24 and 48 hours and between 48 and 72 hours gives similar results
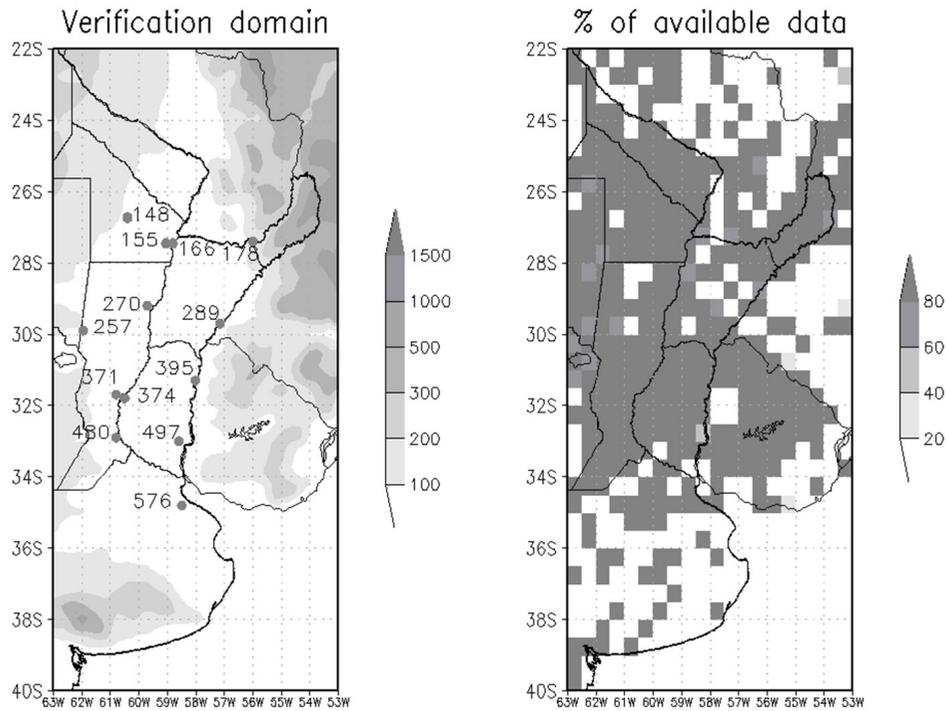
Figure 9 - The left panel shows the verification domain for precipitation (topography shaded and indicated in meters), circles indicate stations used to verify 6-hour accumulated rainfall forecasts, the last 3 digits of the WMO code of each station were included. Right panel indicates the percentage of days in the SALLJEX period on which the 0.5º x 0.5º boxes contain data.
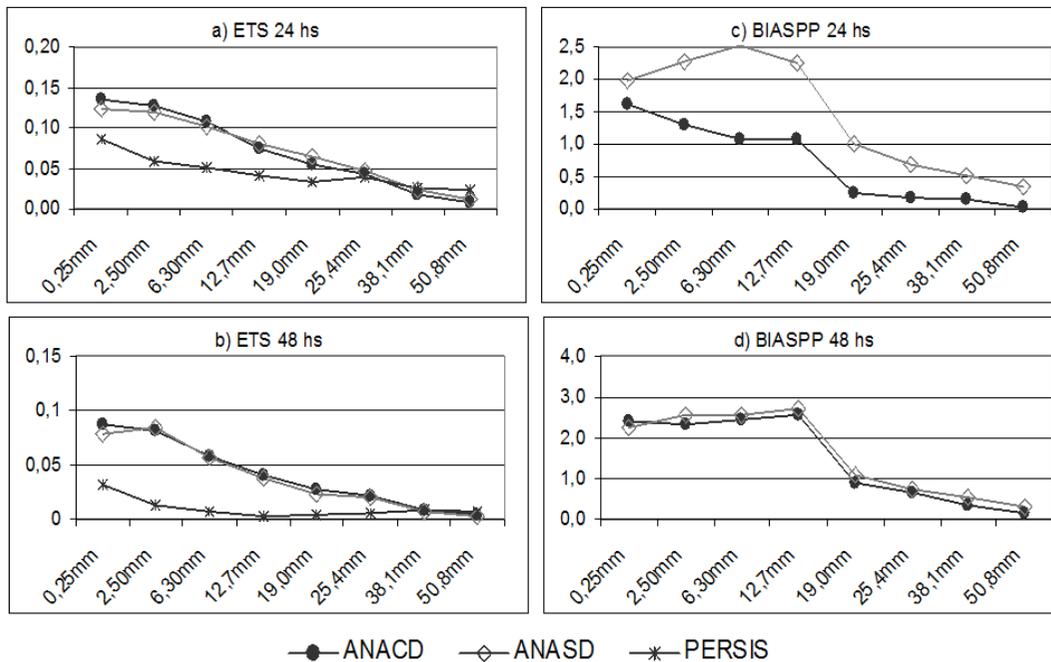
Figure 10. ETS on the left and BIASPP on the right, for a) and c) 24, b) and d) 48 hour forecasts of 24-hour accumulated rainfall using the different initial fields and the persistence.
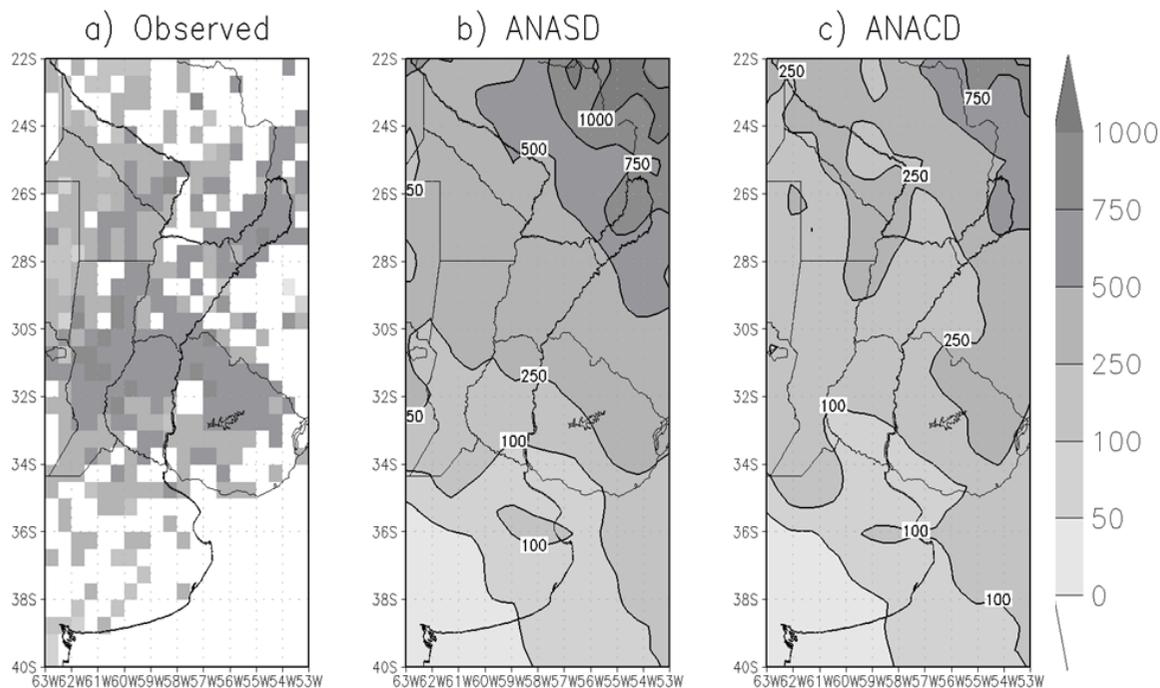
Figure 11 - Accumulated rainfall in mm for the whole SALLJEX period: a) observed, b) and c) forecasts initialized with ANASD and ANACD respectively, for the first 24 hour forecasts.

(Figure not shown). However, as the forecast time increases, differences between forecasts with and without assimilated data tend to decrease. BIAS temperature profiles in the area where precipitation forecasts were verified show that in the first 12 hour forecast temperature decreases at lower layers when ANACD was used as initial field stabilizing the temperature gradient (not shown). This might partially explain the smaller rainfall areas and the smaller rainfall amount forecasted.

The analysis showed that the impact of assimilating data into the model can only be observed in the first 24 hours. In the case of rainfall, starting with enriched initial fields does not improve model performance, at least when assessing rainfall accumulated in a 24-hour period. Forecast rainfall area and amount decrease when an enriched initial field is used.

In order to analyze if the impact of data assimilation on the rainfall forecast is greater in the shorter term, Table 2 presents the RMSE values obtained for the 13 meteorological stations considered, for rainfall accumulated in the last 6 hours of 6, 12, 18 and 24 hour forecasts, in generated forecasts using both enriched and not enriched initial fields. RMSEs of accumulated rainfall forecasts between 6 and 12 forecast hours decrease at all stations when forecasts are initialized with enriched analysis, except in Corrientes. In the remaining cases, results vary depending on the meteorological station considered. As mentioned before, it has to be taken into account

that rain gauges provide observations at a point and model results have greater spatial representativeness. For example, at two neighboring sites -Resistencia and Corrientes- different RMSE values were obtained, where due to spatial resolution of the model the same amounts of rainfall were forecasted at both locations. In general, BIAS values are negative (not shown), which indicates that rainfall is underestimated. However most of the calculated values were smaller than 1 mm. BIAS estimates did not reveal any improvements in model performance when runs were initialized with enriched analyses.

4. CONCLUSIONS

For most variables, a positive short term impact on forecasts was observed when using analyses with data assimilated as initial field. Such impact is observable until 24 hour forecasts. This short time impact might be due to the size of the nested domain Effects are stronger on wind variables than on thermodynamic ones. In addition, better results were obtained when pilot balloon wind observations were assimilated. This result is very useful to plan improvements in upper-level observation networks, considering that pilot balloons are much cheaper than rawinsondes.

Both forecasts reproduce the temporal and spatial location of the SALLJ meridional wind component maxima

Table 2 - RMSE of 6 hours accumulated precipitation forecast for 6, 12, 18 y 24 hours at different sites. Forecast initialized with ANASD and ANACD, respectively.

Stations	6 hs Forecast		12hs Forecast		18hs Forecast		24hs Forecast	
	ANASD	ANACD	ANASD	ANACD	ANASD	ANACD	ANASD	ANACD
87148	5,6	3,8	10,4	10,0	8,5	8,1	3,6	2,8
87155	7,6	8,7	6,6	6,1	4,9	4,5	8,7	8,7
87166	4,4	6,2	5,3	5,6	4,5	4,4	10,3	10,5
87178	9,2	9,4	6,3	4,0	8,3	8,5	7,6	7,7
87257	7,0	5,9	3,4	2,6	7,3	7,1	10,1	10,2
87270	6,8	7,1	4,3	2,9	6,8	9,0	9,8	10,2
87289	6,0	6,1	5,9	5,2	6,5	6,0	10,5	9,9
87371	8,0	7,2	5,4	5,1	9,8	9,6	13,7	13,1
87374	12,7	11,0	6,3	5,6	8,5	8,3	8,7	8,4
87395	4,9	3,6	6,7	6,1	7,8	7,4	15,5	15,6
87480	6,4	6,2	4,0	2,9	7,0	6,9	6,5	3,8
87497	8,0	7,6	5,4	4,9	4,0	3,7	5,3	5,4
87576	6,4	6,8	4,0	3,6	6,8	7,0	5,5	5,5

and minima fairly well, but they underestimate the maximum SALLJ speed. The analysis of the RMSE and BIAS of SALLJ associated variables shows that enriched initial fields produce better 6- and 12-hour forecasts. The impact is much weaker in 18- and 24-hour forecasts. Also the impact on the analyses increases with decreasing distance to the mountains, because models fail to represent adequately the fields of atmospheric variables in areas with complex topography. In mountainous regions, the errors of the analyses not assimilating data are greater than in flat regions. Enriched analyses contribute to obtaining more realistic forecasts in areas located closer to the Andes.

Positive impacts are not observable in the verification of 24-hour accumulated rainfall forecasts. Rainfall is underestimated in 24 hours forecasts and the underestimation increases when analyses with data assimilated are used as initial fields. For rainfall accumulated over a shorter period, the verification of 6-hour accumulated rainfall against observations shows that 12-hour forecasts represent rainfall better when using an enriched initial field. This result may be revealing that enriched initial fields have positive effects even on rainfall, though only in the first 12 forecast hours. Nicolini and Saulo (2006) and Salio and Nicolini (2006) found that the heaviest rainfall events and the most intense convective activity occur during nighttime in the study region. In view of all this, maybe

a way of improving predictions is initializing forecasts in the last hours of the afternoon and advancing 6 to 18 hour forecasts to cover more accurately the nighttime period. This will also make it possible to define the SALLJ more accurately at the time when its maximum intensity occurs, considering that the greatest impact of data assimilation is observed in the first 12 hours.

In summary, although positive impacts on the forecast of wind variables, temperature and humidity are observed when using enriched initial fields, this impact does not last longer than 24 hours and does not improve forecasts of 24-hour accumulated rainfall. A slightly positive impact on the rainfall forecast is observed in rainfall accumulated between 6 and 12 forecast hours. We found that starting with a set of analyses whose values are closer to the real field has slightly improved forecasts. The results obtained for this particular forecast set are difficult to generalize, given their strong dependence on the model, the selected setup and the assimilation methodology which may be the main cause of not achieving greater impacts on the forecasts.

5. ACKNOWLEDGMENTS

6. REFERENCES

ADCROFT, A.; HILL, C.; MARSHALL, J. Representation of topography by shaved cells in a height coordinate ocean model. Monthly Weather Review, v. 125, p. 2293-2315, 1997.

BARKER D. M.; HUANG, W. ;GUO, Y.-R ;BOURGEOIS, A.J.; XIAO, Q.N. A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. Monthly Weather Review, v. 132, p. 897-914, 2004.

BERBERY, H.; BARROS, V. The hydrologic cycle of the La Plata Basin in South America. Journal of Hydrometeorology, v. 3, p. 630–645, 2002.

BONNER, W. D., Climatology of the low level jet. Monthly Weather Review, v. 96, p. 833-850, 1968

CAVALCANTI, I.; HERDIES, D. Data Assimilation study using SALLJEX data. CLIVAR Exchanges. v. 9, n.1. p. 23-24, 2004.

CHERUBINI, T.; BUSINGER, S.; VELDEN, C.; OGASAWARA, R. The impact of satellite derived atmospheric motion vectors on mesoscale forecast over Hawaii. Monthly Weather Review, v. 134, p. 2009-2020, 2006.

CHEN, C.; COTTON, W. R.. A one dimensional simulation of the stratocumulus-capped mixed layer. Boundary-Layer Meteorology, v. 25, p. 289-321, 1983.

COTTON, W. R. et al. RAMS 2001: Current status and future directions. Meteorology and Atmospheric Physics, v. 82, p. 5-29, 2003.

CUCURULL, L.; VANDENBERGHE, F.; BARKER, D.; VILACLARA, E. ; RIUS, A. Three-Dimensional Variational Data Assimilation of Ground-Based GPS ZTD and Meteorological Observations during the 14 December 2001 Storm Event over the Western Mediterranean Sea. Monthly Weather Review, v. 132, p. 749-763, 2004.

DAVIES, J.C. Statistics and Data Analysis in Geology. J. Wiley Ed., New York, 550pp, 1973.

DAVOLIO, S.; BUZZI, A. A Nudging Scheme for the Assimilation of Precipitation Data into a Mesoscale Model. **Weather and Forecasting**, v. 19, p. 855-871, 2004.

FACCANI, C.; FERRETTI, R. ; VISCONTI, G. High-resolution weather forecasting over complex orography: Sensitivity to the assimilation of conventional data. **Monthly Weather Review,** v. 131, p. 136-154, 2003.

FREITAS, S. R. et al. The Coupled Aerosol and Tracer Transport model to the Brazilian developments on the Regional Atmospheric Modeling System (CATT-BRAMS). Part 1: Model description and evaluation. **Atmospheric Chemistry and Physics**, v. 9, p. 2843–2861, 2009

GARCÍA SKABAR, Y.; M. NICOLINI. Enriched Analyses with Assimilation of SALLJEX Data. **Journal of Applied Meteorology and Climatology**, v. 48, p. 2425-2440. doi: 10.1175/2009JAMC2091.1, 2009.

GEVAERD, R.; S.R. FREITAS. Estimativa operacional da umidade do solo para iniciação de modelos de previsão numérica da atmosfera parte I: descrição da metodologia e validação. **Revista Brasileira de Meteorología**, v. 31, n. 3, p. 59-73, 2006.

GRELL, G.A.; D. DEVENYI. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. **Geophysical Research Letters**, v.29, n. 14, p. 1963, 2002.

HALLAK R.;SILVA DIAS, P. L. Operationalization of regional forecast with RAMS in DCA/IAG/USP. Proc. IX Congresso Brasileiro de Meteorologia. **Anais...** FLorianópolis, 1996, p.795-799.

HARMS, D. E.; RAMAN, S.; MADALA, R. V. An Examination of Four-Dimensional Data-Assimilation Techniques for Numerical Weather Prediction. **Bulletin of the American Meteorological Society**, v. 73, n. 4, p. 425–440, 1992.

KAIN, J.S. et al.. Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP. **Weather and Forecasting**, v. 23, p. 931-952, 2008.

LIEBMANN, B.; ALLURED,D.. Daily Precipitation Grids for South America. **Bulletin of the American Meteorological Society**. v. 86, n. 11, p. 1567–1570, 2005.

MELLOR G.L.;YAMADA, T.. Development of a turbulence closure model for geophysical fluid problems. **Reviews of Geophysics and Space Physics**, v. 20, p. 851-875, 1982.

MESINGER, F. Improvements in quantitative precipitation forecast with the ETA regional model at the National Centers for Environmental Prediction: The 48 km upgrade. **Bulletin of the American Meteorological Society**, v. 77, p. 2637-2649, 1996.

NICOLINI, M.; SAULO, A. C. Eta characterization of the 1997-1998 warm season Chaco jet cases. Preprint, Sixth Int. Conf. on Southern Hemisphere Meteorology and Oceanography. **Annals...** Santiago, Chile, Amer. Meteor. Soc., 2000, p. 330-331.

NICOLINI, M.;GARCÍA SKABAR,Y.; ULKE,A.G.; SAULO, A. C. RAMS model performance in simulating precipitation during strong poleward low level jet events over northeastern Argentina. **Meteorologica - Special Issue on Variability of the South American Monsoon System**, v 27, n. 1-2, p. 89–98, 2002.

NICOLINI, M.; SAULO, A. C. Modeled Chaco low-level jets and related precipitation patterns during the 1997-1998 warm season. **Meteorology and Atmospheric Physics**, v. 94, p. 129-143, 2006.

NICOLINI. M.; SALIO, P; BORQUE, P. Thermodynamic and kinematic characterization of the low-level troposphere during SALLJEX under different large-scale environments. Proc. of the 8th International Conference on Southern Hemisphere Meteorology and Oceanography. **Annals...** Foz do Iguaçu, Brazil, April 24-28, 2006, p 1141-1148.

PAEGLE, J., BERBERY, E. H. ; GARREAUD, R. ; AMBRIZZI, T. ; PROFIRIO DA ROCHA, R. ; SILVA DIAS, P. L.; HERDIES, D.; MARENGO, J. ; SELUCHI, M; CAMPETELLA, C. ; MENENDEZ, C.; NICOLINI, M.; RUIZ, J.; SAULO, C. Modeling Studies Related to SALLJEX. **Clivar Exchanges**, v. 29, p. 20-22, 2004.

PIELKE, R. A., 2002: **Mesoscale Meteorological Modeling**. Academic Press, 676 pp.

REYNOLDS, R.W.; N.A. RAYNER; T.M. SMITH; D.C. STOKES; W. WANG. An Improved In Situ and Satellite SST Analysis for Climate. **Journal of Climate**, v. 15, p. 1609-1625, 2002.

SALIO, P.; NICOLINI, M. Seasonal characterization on the diurnal cycle of convection frequency over Southeastern South America under different low-jet conditions. 8th International Conference on Southern Hemisphere Meteorology and Oceanography (8ICSHMO). **Annals...** Foz do Iguaçu, Brazil, April, 24-28, 2006, p. 1157-1162.

SALIO, P.; CAMPETELLA, C.; RUIZ, J.; GARCÍA SKABAR, Y.; NICOLINI, M. Snow-fall over Southeast of Buenos Aires province: Synoptic climatology and a case study. **Meteorologica**, v. 31, n. 1-2, p. 67-84, 2006.

SAULO, A.C.; SELUCHI, M.; CAMPETELLA, C.; FERREIRA, L. Error evaluation of NCEP and LAHM Regional Model Daily Forecasts over Southern South America. **Weather and Forecasting**, v. 16, p. 697-712, 2001.

SAULO, A.C.; FERREIRA, L.. Evaluation of quantitative precipitation forecast over southern South America. **Australian Meteorological Magazine**, v. 52, p. 81-93, 2003.

SAULO, C.; RUIZ, J.J.; GARCÍA SKABAR, Y.. Synergism between the low level jet and organized convection at its exit region. **Monthly Weather Review**, v. 135, p. 1310-1326, 2007.

SMAGORINSKY, J., 1963: General circulation experiments with the primitive equations. Part I, The basic experiment. **Monthly Weather Review**, v. 91, p. 99-164, 1963.

SOUZA, E. P.; SILVA, E. M. Impacto da Implementação de uma Parametrização de Convecção Rasa em um Modelo de Mesoscala. Descriçao e Testeo de Sensibilidade do Esquema. **Revista Brasileira de Meteorologia**, v. 18, n. 1, p. 33-42, 2002.

VERA, C. et al. The South American low-level jet experiment. **Bulletin of the American Meteorological Society**, v. 87, p. 63-77, 2006.

WARNER T.; KEY, L.; LARIO, A. Sensitivity of mesoescale model forecast skill to some initial data characteristics: data density, data position, analysis procedure and measurement error. **Monthly Weather Review**, v. 117, n. 6, p. 1281 – 1310, 1989.

WEISMAN M. L.; DAVIS, C.; WANG, W.; MANNING, K. W.; KLEMP, J. B. Experiences with 0–36-h Explicit Convective Forecasts with the WRF-ARW Model. **Weather and Forecasting**, v. 23, p. 407-437, 2008.