Verification of WRF-ARW convective-resolving forecasts over Southeastern South America

C. MATSUDO, Y. GARCIA SKABAR*, J. J. RUIZ**, L. VIDAL*** and P. SALIO****

National Meteorological Service, Argentina

*National Scientific and Technical Research Council (CONICET), Argentina

**Centro de Investigaciones del Mar y la Atmósfera (CIMA-CONICET), Argentina

***Department of Atmospheric and Ocean Sciences, University of Buenos Aires, Argentina

****Instituto Franco Argentino sobre Estudios del Clima y sus Impactos

UMI - IFAECI (CNRS/UBA/CONICET), Argentina

e mail : matsudo@smn.gov.ar

सार – अर्जेनटीना की राष्ट्रीय मौसम सेवा (नेशनल वेदर सर्विस) द्वारा नवम्बर-दिसम्बर 2012 में उच्च विभेदन (4 कि.मी.-38 कि.मी. ऊँचाई के स्तरों), संवहन से 48 घंटों का डब्ल्यू आर एफ-ए आर डब्ल्यू पूर्वानुमान दिए गए थे। इस शोध पत्र का उद्देश्य घंटेवार मात्रात्मक वर्षा के पूर्वानुमानों का आकलन करना है ताकि इसके स्थान, तीव्रता, स्थानिक विभिन्नता और दैनिक चक्र को प्रस्तुत करके मॉडल के कार्य निष्पादन का पता किया जा सके। इसमें अर्जेनटीना के मध्य पूर्वी क्षेत्र और दक्षिणी ब्राजील पर ध्यान केंद्रित किया गया है। यह अध्ययन पूर्वानुमानित तथा क्षेत्रों में हुई आकलित कुल वर्षा की तुलना और CMORPH (सी पी सी मोरफिंग) उपग्रह वर्षा आकलनों आब्जेक्टीव स्कोर्स की गणना 8 कि.मी. विभेदन के संयोजन के आधार पर किया गया है। 6 एवं 7 दिसम्बर, 2012 को हुई संगठित संवहनीय घटनाओं की जाँच के द्वारा अतिरिक्त जानकारी प्राप्त की गई है। सिम्युलेटेड मॉडल परावर्तकता का उपयोग करते हुए संवहनीय विशेषताओं का मूल्यांकन करने के लिए रेडार डेटा को एक पूरक के रूप में लिया गया है। परिणाम दर्शाते हैं कि पूर्वानुमान के डब्ल्यू आर एफ मॉडल बड़ी संवहनीय घटनाओं की स्थिति और समय को अच्छी तरह अभिग्रहण करता है तथापि इसमें घटनाओं के परिमाण को कम महत्व दिया गया है। सत्यापन क्षेत्र और छोटी अवसीमाओं के अंतर्गत आने वाले क्षेत्र की कुल मात्राओं के औसत को कम महत्व दिया गया है। सामान्यत: इन परिणामों से पता चला है कि संवहन-अनुमत्य डब्ल्यू आर एफ- ए आर डब्ल्यू मॉडल में किसी क्षेत्र के लिए अल्प अवधि पूर्वानुमान दने में सुधार लाने की क्षमता है यद्यपि इससे लंबी अवधि का मूल्यांकन किया जाना चाहिए।

ABSTRACT. During November-December 2012, high-resolution (4 km-38 vertical levels), convection-allowing 48 hours WRF-ARW forecasts were produced at the National Weather Service of Argentina. The aim of this paper is to evaluate hourly quantitative precipitation forecasts to assess the model performance on representing its location, intensity, spatial variability and diurnal cycle. The focus is on the central-east region of Argentina and south of Brazil. The study is based on a combination of visual comparison of forecasted and estimates accumulated precipitation fields and objective scores calculated employing 8-km resolution CMORPH (CPC MORPHing technique) satellite rainfall estimations. Additional insight is gained by examining an organized convective event occurred during 6th and 7th December, 2012. As a complement, radar data is considered to evaluate convective features using simulated model reflectivity. Results show that WRF model forecast captures quite well the position and timing of the major convective events, even though the magnitude of events was underestimated. Total amounts averaged over the verification domain are underestimated as well as the areal coverage for small thresholds. In general, results suggest that convection-allowing WRF-ARW model has the potential to improve short range forecasts over the region although it should be evaluated over a longer period of time.

Key words – Verification, Performance, High-resolution, Rainfall, Satellite estimates of precipitation, CMORPH, Convective-resolving.

1. Introduction

Southeastern South America (SESA) is one of the regions most affected by intense deep moist convection usually associated to the development of large mesoscale convective systems (MCSs). These systems account for over 80% of the austral summer precipitation (Salio *et al.*, 2007) and most of the severe weather phenomena over the

region (Matsudo and Salio, 2011). MCS development over the region is favored by moisture transport from tropical latitudes produced by the South American Low Level Jet (SALLJ, Vera *et al.*, 2006). Because of the predominant convective origin of warm season precipitation, Quantitative Precipitation Forecasts (QPFs) are quite challenging during this season. One possible way to address this problem, as a first step, is the use of high-resolution and convection-resolving models using a "cold start" approach (*i.e.*, the model is initialized from an atmospheric state estimated in a low resolution analysis system and little or no information of the mesoscale circulation is present in the initial conditions). This is the approach that has been adopted at the National Weather Service (NMS) of Argentina given the lack of a data assimilation system that can incorporate high resolution observations (*i.e.*, radar and some satellite observations). The first attempt was performed using a regional Numerical Weather Prediction (NWP) model, the Brazilian Regional Atmospheric Modeling System (BRAMS), reaching a 2-km resolution, for a very small area of around 300 × 300 km domain (García Skabar *et al.*, 2011).

Since the Weather Research and Forecasting-Advanced Research WRF model (WRF-ARW) is being widely used in the region and by the scientific community worldwide in general, which produces a huge progress in its development, comparison experiments between BRAMS and WRF models were performed. An objective evaluation for a warm season using similar high resolution configuration for both models showed a better performance for WRF-ARW forecasts (Dillon *et al.*, 2013).

During November-December 2012, the CHUVA Santa Maria Experiment was carried out over SESA, as part of an International Research Project focused on Brazil (http://chuvaproject.cptec.inpe.br). This research project deployed conventional and special observations in order to create a database to describe the cloud processes of the main precipitating systems in the region. This experiment had also a strong modeling component in which the main goal was the evaluation of convective resolving models over the region. In this context, different institutions produced convective-resolving NWP runs for the region to generate ensemble precipitation forecasts from convective allowing models. Particularly, the NMS- Argentina produced the high-resolution forecasts using WRF-ARW that are evaluated in this work.

Aiming to design a forecast system, it is necessary to perform a quantitative verification to assess model performance based on the weather systems that are intended to predict. In particular, verification of mesoscale

systems is usually challenging because of the limited predictability in this scale as well as for the high spatio-temporal variability of some variables (*i.e.*, precipitation) and for the intermittency of some phenomena like convection. To address this problem, particularly for rainfall forecasts, the World Meteorological Organization (WMO) issued recommendations for the verification and intercomparison of QPFs from operational NWP models (WWRP/WGNE Joint Group on Verification 2009). Research in this subject has been done by verifying numerical forecasts with resolutions lower than 10 km where convection is explicitly resolved (Fowle and Roebber, 2003; Weisman *et al.*, 2008; Kain *et al.*, 2008, among others). All of them, applying different methodologies, validate high resolution forecasts towards information derived from radar and/or dense spatio-temporal rain gauge networks. These works agree in that standard methodologies may give false representation of a convective-resolving forecast value. Furthermore, the use of traditional metrics can lead to poor scores because of the difficulty in forecasting the exact location of precipitation systems. Particularly, when those precipitation features are comparable in scale to the model grid resolution. In this sense, there are techniques that are more representative of the spatial and temporal variability of the characteristic mesoscale systems such as "fuzzy" verification scores. Ebert (2008) reviewed a variety of "fuzzy techniques available for different applications.

In particular, over some parts of South America (like for example Northern and central Argentina), a major concern for the verification of high resolution forecasts, is the scarcity of in-situ, surface and upper level observations. One of the alternatives to overcome this issue is the use of remote sensing data. Satellite information becomes crucial to provide appropriate spatial and temporal resolution accordingly to the scale of the phenomena forecasted. For precipitation, CPC MORPHing technique (CMORPH, Joyce *et al.*, 2004) estimates are a useful tool to evaluate the model skill in representing the spatial pattern and timing of precipitation forecasts. Another valuable source of highly frequent high resolution information is provided by weather radars. In terms of the relevance of mesoscale verification and due to their higher spatial and temporal resolutions, radar data could be more helpful as it provides more detailed internal structure of precipitating systems.

The aim of this paper is to evaluate the performance of high resolution precipitation forecasts produced with WRF model at the National Meteorological Service in Argentina, for a 30 days period during CHUVA Santa Maria experiment. In section 2 data and methodology are described, in section 3 validation results are presented and finally in section 4 the conclusions are summarized.
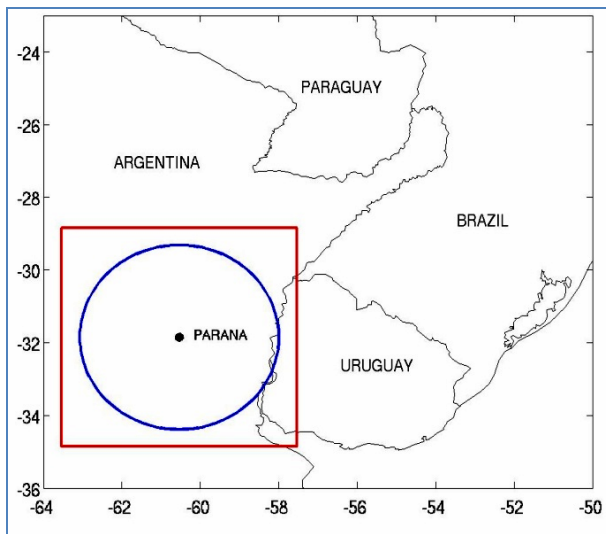
Fig. 1. Model domain, verification domain (red box, 29° S - 34.5° S; 63.5° W - 57.5° W) and 240-km range of Parana radar (blue circle)

2. Data and methodology

2.1. *Model description*

WRF-ARW model version 3.4.1 (Skamarock *et al.*, 2008) 48-hr forecasts were initialized twice a day (at 00 and 1200 UTC) at the NMS-Argentina from November to December 2012. The main features include 4 km grid spacing, 38 vertical sigma-p levels, 1-hour temporal resolution and convection explicitly resolved over the domain shown in Fig. 1. Model parameterizations includes the Yonsei University (YSU) planetary boundary layer (PBL) scheme (Hong *et al.*, 2006) and the WRF single-moment (WSM), particularly, the WSM6 microphysics scheme (Hong and Lim, 2006) while surface processes are modeled using the 4-layer NOAH LSM (Chen and Dudhia, 2001). YSU PBL scheme is a nonlocal scheme that includes counter gradient flux terms that enables realistic development of a well-mixed layer, while WSM6 scheme is appropriate for examining heavy rainfall events by including graupel as another prognostic variable in addition to rain and snow. Regarding the soil model, the NOAH scheme is a four-layer model that forecasts soil moisture and temperature which includes a time-varying green vegetation fraction, soil type and snow cover with up to two vertical layers. Similar configurations, as the one chosen form this work, are widely used in different institutions around the globe running the WRF model in a real-time basis. As an example, the WRF model with WSM6 microphysics and YSU boundary layer scheme provided useful information on high-resolution weather phenomena over U.S. (Schwartz *et al.*, 2010).

Initial and boundary conditions are provided by the GFS (Global Forecast System) forecasts with 0.5° horizontal grid available every 3 hours. In this work, hourly rainfall and maximum reflectivity forecasts initialized only at 0000 UTC are analyzed for forecast lead times between 12 and 36 hours from 20th November to 19th December, 2012 defined, as follows, as the verification period.

The simulated reflectivity (SCR) has been computed from the forecasted concentration of precipitation-size hydrometeors, assuming Rayleigh scattering by spherical particles and a Marshall-Palmer distribution for the hydrometeors sizes (Kain *et al.*, 2008).

2.2. *Remote sensing data*

2.2.1. *Satellite precipitation estimates*

CMORPH passive microwave precipitation estimates (http://www.cpc.ncep.noaa.gov/products/janowiak/cmorp h description.html) with 8 km horizontal resolution and 30 minutes temporal resolution version (Joyce *et al.*, 2004) are used as the observational data for the forecast verification.

Although there are currently other satellite precipitation estimates, CMORPH has been selected because the use of microwave data results in a better performance (Ebert *et al.*, 2007; De Maria *et al.*, 2011; Salio *et al.*, 2014) becoming an excellent alternative dataset to the coarse rain gauge networks available over South America as those included in the Global Telecommunication System (GTS, black triangles in Fig. 2). Particularly, among other existing estimates, CMORPH is the most adequate for verification purposes according to the model resolution. Nevertheless, some cautions should be taken when considering the use of remote sensing data instead of surface observations. Few previous studies included a validation of this 8-km CMORPH version over South America. Nonetheless, some results were addressed using coarser resolutions (0.25° × 0.25° horizontal resolution) of 24-hr accumulated values from special rain gauge networks. For example, Ruiz (2009) compared the 0.25° CMORPH version against a denser pluviometric network than the GTS one specially provided for the SALLJ Experiment (Vera *et al.*, 2006). Examining the 2002-2003 warm season, they showed that CMORPH overestimates the observed rainfall over their northeastern region [their Fig. 2(c)] similar to our region of interest. They found that systematic errors are not uniformly distributed and therefore, in order to apply a calibration, also seasonal variability should be considered. Salio *et al.* (2014) evaluated the 0.25° version of CMORPH data and other five different 24-hr
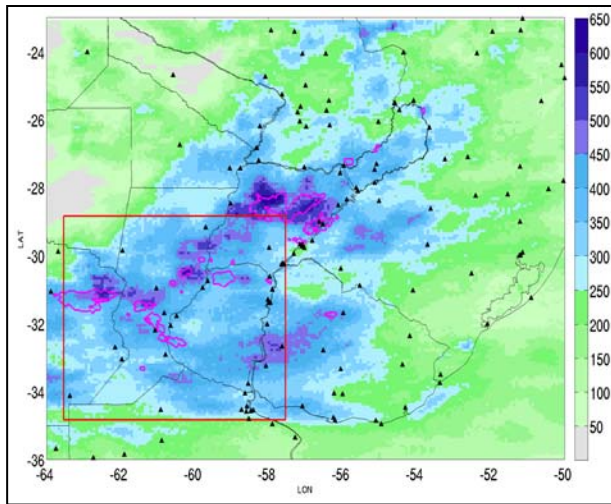
Fig. 2. Total rainfall for the period estimated by CMORPH (in mm, shaded), areas where rainfall estimates were greater than 20 mm/day over more than 25% of the days (purple contours) and verification domain (red box, 29° S - 34.5° S; 63.5° W - 57.5° W). GTS network is shown in black triangles

accumulated rainfall estimates for South America, including products like those of the Tropical Rainfall Measuring Mission 3B42 (Huffman *et al.*, 2007) Version 6, Version 7 and Real Time, Hydroestimator (HYDRO, Vila *et al.*, 2002; Scofield and Kuligowski, 2003) and the Combined Scheme algorithm (CoSch, Vila *et al.*, 2009). As well as Ruiz *et al.* (2009), they found large overestimations detectable in CMORPH performance, principally for extreme values over plain areas. POD and BIAS scores revealed overestimations for thresholds over 20 mm while correlations of 0.63 for annual 24-hr accumulated data were found for 0.25° CMORPH. Although Cosch achieved better performances than CMORPH, due to the correction with surface observations, CMORPH obtains comparable results with other estimates and has the advantage of providing a high resolution version of the data more suitable for this work purposes. Regarding these issues, both fixed and quantile thresholds will be used to address CMORPH biases with further explanation of this methodology described at section 2.3.

Acknowledging the limitations associated with the scarcity of available surface rain gauges and the CMORPH dataset performance over the region, deeper assessments should be carried out in order to achieve more accurate calibration methods to reduce systematic errors related to rainfall climatological regimes and spatial variability.

2.2.2. *Maximum reflectivity*

In this study, radar information was provided by the C-Band (frequency: 4-8 GHz, 12 elevations and 1° beam width) dual-polarization operational radar from the Agricultural Technology National Institute(INTA) located in Parana city (31.84° S, 60.53° W). Volume scans of reflectivity every 10 minutes up to 240 km of range were used to compute fields of maximum reflectivity in the vertical column (ColMax). ColMax reflectivity positions were transformed from radial format (*i.e.*, azimuth-range) to corresponding latitude-longitude locations. Then, the latitude-longitude locations were mapped to a 0.5×0.5 km grid of 480×480 km. Some caution should be taken since the radar resolution degrades with distance from the transmitter, shown by Pappalardo *et al.* (2014), and C-band radars are affected by attenuation. An automatic clutter filter is applied but no attenuation correction is made. Quality controls and data validation are still in progress for INTA radars.

ColMax reflectivity has become a common post-processing product for both research and operational models. For our purposes, the use of ColMax reflectivity detects the location of stronger convective cells although the information regarding its vertical structure is lost. It is important to note that this product is a way to "summarize" reflectivity information from different levels reducing the complexity between the observed and simulated data comparison. As stated by Koch *et al.* (2005) it is not possible to make a strict comparison between composite reflectivity computed from a model grid point and that measured by scanning radar due to several problems as ground clutter, anomalous propagation, bright bands, among others.

2.3. *Verification scores*

For the objective verification of precipitation, 1-hr rainfall accumulations and CMORPH precipitation estimates were linearly interpolated to common grid with 8-km horizontal resolution. Note that this is the same as the native resolution of the CMORPH data but is lower than the horizontal resolution used in the model.

The WWRP/WGNE Joint Group on Verification (WWRP 2009 1, 2009) report lists the recommended verification measures for forecast categories as : (*i*) forecasts of rain occurrence meeting or exceeding specific thresholds, (*ii*) forecasts of rain amount, (*iii*) probability forecasts of rain meeting or exceeding specific thresholds, (*iv*) verification of ensemble probability distribution. In this work, QPFs verification analysis has been conducted using some traditional grid-point scores of the (*i*) category as a first approach.
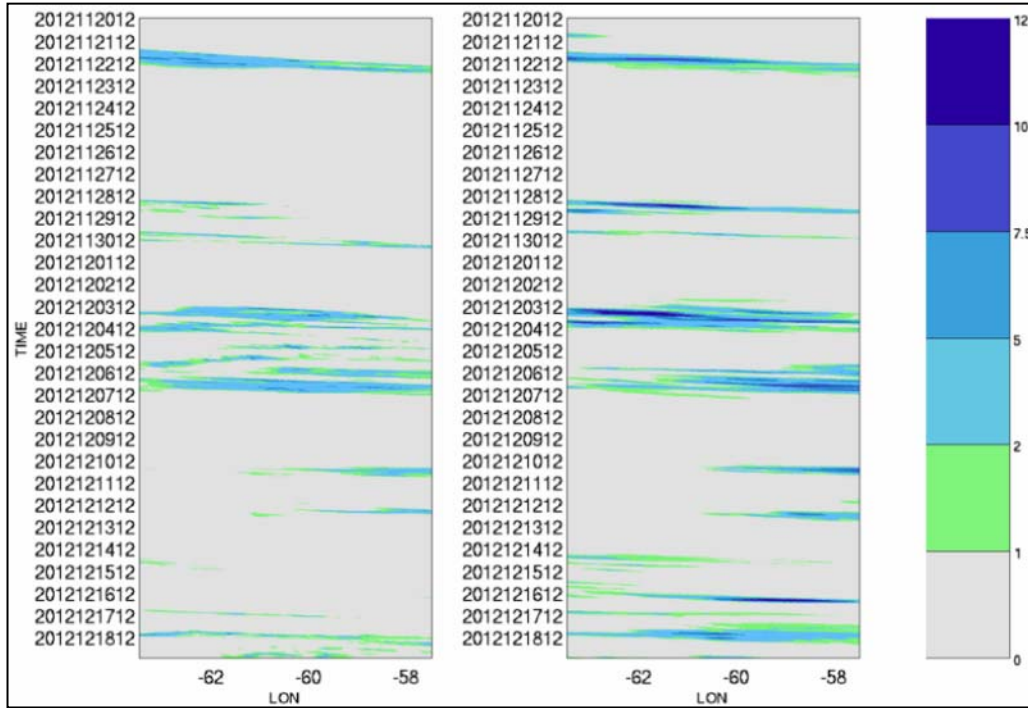
Fig. 3. Hövmoller diagram of 1-hour rainfall (mm) for the 30 day period: 12-36 hr lead time forecasts by WRF (left) and estimated by CMORPH (right). Forecasts correspond to runs initialized at 0000 UTC. Values were averaged in latitude (between 29° S and 34.5° S), over the verification domain

Equitable threat score (ETS), frequency bias (BIAS), false alarm ration (FAR) and probability of detection (POD) were computed (details can be found in the textbooks of Wilks 2006). These traditional categorical metrics are based on evaluating gridded data against point observations or, in our case, the forecasted value of rainfall in a grid box to the one estimated by CMORPH.

The fraction skill score (FSS, Roberts and Lean, 2008), a fuzzy verification approach is also calculated. This score assumes that a slightly displaced forecast can still provide useful information. Therefore, it is more suitable for model evaluation at the mesoscale where spatial and temporal displacements are common. For the computation of the FSS, both the forecast and observed rainfall fields are filtered by converting them into binary grids using a certain accumulated precipitation threshold. Grid boxes with values higher than the threshold are assigned a value of 1 and those with lower values are assigned a value of 0. Then a radius of influence (smoothing radius, *r)* is specified to build a neighborhood around each grid box. FSS compares the fractional coverage of events in the forecast and in the observations, in neighborhoods of different sizes as expressed in the following equation:

$$FBS = \frac{1}{N}\sum_{j=1}^{N}\left(p_j - o_j\right)^2$$

$$FBS = 1 - \frac{FBS}{\frac{1}{N}\left[\sum_j^N p_j^2 + \sum_j^N o_j^2\right]}$$

where p_j and o_j are the forecasted and observed frequency of the event in the neighborhood of the j-grid point.

Values for p_j and o_j fall between 0 and 1 and *N* is the number of pixels in the neighborhood area defined by *r*. As N varies, also the neighborhood size, defining the spatial scale over which the model is being evaluated. The FSS calculation was computed for different rainfall thresholds and forecast lead times.

Additionally, calculations of these scores are done using quantile thresholds as defined by Jenkner *et al.* (2008) instead of fixed amplitude thresholds allowing some calibration due to precipitation intensity biases between observational and forecasted datasets. The use of
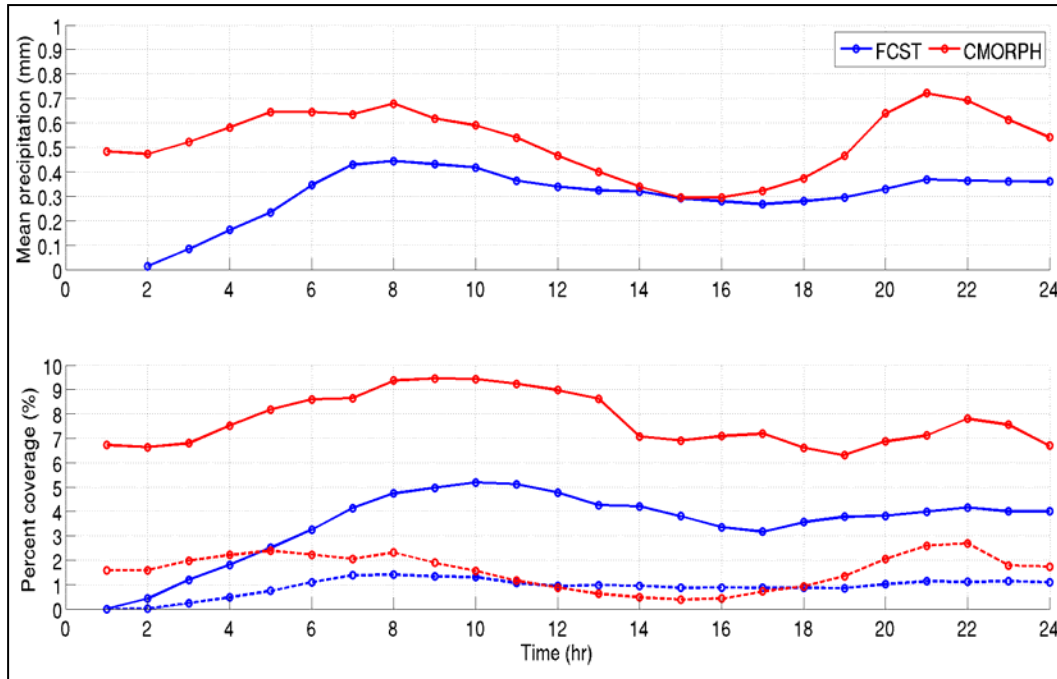
Fig. 4. Top panel: Spatially and temporally averaged precipitation rates (mm/hr) as function of forecast lead time.
Bottom panel: Fractional grid coverage of hourly precipitation values exceeding 1 mm/hr (solid lines) and
10 mm/hr (dashed lines) as a function of forecast lead time (solid lines) averaged over all days. For both
panels, the blue and red lines represent forecasted and CMORPH estimated values respectively. Forecasts
correspond to runs initialized at 0000 UTC

fixed intensity thresholds splits the precipitation distribution into unknown percentiles, making not obvious whether a certain threshold value represents common or rare events.

3.   Results and discussion

As a first insight into the spatial variability of CMORPH rainfall estimates, Fig. 2 shows the accumulated values over the entire 30-day verification period of the CHUVA-Santa Maria experiment. Areas over 500 mm can be found over the verification domain in agreement with the annual mean climatology distribution of CMORPH values over this region (Salio *et al.*, 2014, their Fig. 3). Moreover, there are some areas over which more than the 25% of the days presented values over 20 mm/day (contours in Fig. 2). However, to see how these precipitation accumulations were distributed over the experiment period, Fig. 3 shows that, although there were only a few convective events, at least five had intensities over 10 mm/hr. While the most intense convective events were well captured by the model, forecasted precipitation rates were generally lower.

First, in order to validate the model forecasts, time-average forecasted precipitation is examined to explore the presence of systematic biases as a function of a possible diurnal cycle. To achieve this, hourly precipitation rates were averaged over the verification domain (red box in Fig. 1, 29° S - 34.5° S; 63.5° W - 57.5° W) for the entire verification period (Fig. 4). Large differences between the forecast and estimated precipitation are found on the first 6 to 7 hours due to the spinup period of the mesoscale circulation in the model (Kain *et al.*, 2008). This is mainly a consequence of the lack of information about mesoscale circulation in the initial conditions. These circulations need to develop in the model and this takes certain time that is usually around 6 hours. No liquid water or ice is present in the model initial conditions, so clouds also need to develop during the model spinup period. Apart from the bias in the first hours, Fig. 4 shows that the model does not capture the shape of the diurnal cycle of precipitation with two distinct maximum biases around 8 and 2100 UTC. In fact, the total precipitation and the amplitude of the diurnal cycle is under predicted, especially during the morning (8 - 10 - hr forecast time) and early nighttime (20-22 hr
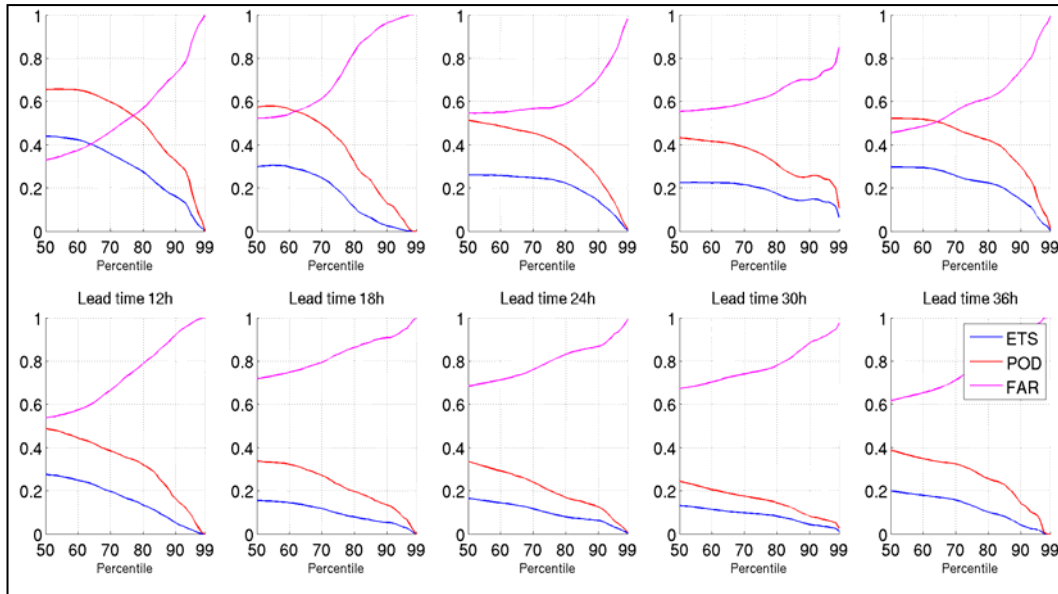
Fig. 5. ETS, FAR and POD for 6-hr (top) and 1-hr (bottom) accumulated rainfall using percentile thresholds for the 30-day period for different lead times

forecast time) when the model produces nearly half of the estimated precipitation. These results could be related to the diurnal cycle of mature stage MCSs during SALLJ events (Salio *et al*., 2007). A different model performance in representing rainfall diurnal cycle was found by Kain *et al*. (2008) for similar convective cases over the United States Great Plains. This could correspond to the fact that they were able to evaluate the model performance against stage II precipitation dataset which optimally combines both radar data and rain gauge data (Baldwin and Mitchell, 1998). Bottom panel in Fig. 4 shows the fractional coverage of hourly rainfall over 1mm/hr and 10 mm/hr averaged over all days. The fractional coverage in the model is systematically lower than in the estimations suggesting that, at least part of the bias previously discussed, can be due to a smaller size of the precipitation systems produced by the model. Considering 1 mm/hr threshold, differences between the forecasted and estimated area of precipitation are larger than for a higher threshold. The model overestimates the area with rainfall rates over 10 mm/hr during the early afternoon (13-18-hr forecast time) and fails to capture the diurnal cycle in the size of areas with precipitation over 10 mm/hr.

It is not clear at this point if differences between the forecasted and the estimated precipitation for the 1 mm/hr threshold comes from model deficiencies alone or if they are strongly affected by CMORPH biases which are known to be important for weak rainfall rates. For this reason, quantile-based thresholds are used for computing the objective scores, in order to reduce the impact of biases present both in the forecast and in the estimated

rainfall values. Moreover, the remaining calculations were focused on the 12 to 36 hr lead time period to avoid problems related to the model spinup period. Fig. 5 presents the results obtained with the traditional scores. Using thresholds based on the quantiles from the precipitation distribution improves the results as it removes the systematic component of the error before comparing the forecasts and the estimates. In general, when considering 6-hr accumulated values, higher skills are present (Fig. 5). POD and ETS values reflect that the model shows poorer performance as the lead time increases, where for the 36-hr lead time a smaller improvement can be noticed reaching values similar to those found at 24-hr lead time. This could be partially explained by the diurnal cycle of summer convection over this region which has relative maximum activity during nighttime hours around 0600 UTC (30 hr lead time) so that it becomes harder to predict accurately. In the ETS case, this could be associated to the fact that this score is not as "equitable" as revised by Hogan *et al*. (2010) and Hamill and Juras (2006). Therefore, this score is highly dependent on the climatology and sample size and also dependent on the locations where precipitation was more frequent.

An alternative to the previous point-based scores is the use of a smoothing filtering method applied to both the forecast and observed fields and then to calculate verification scores on the filtered field (Gilleland *et al*., 2009). One approach is presented as a "fuzzy" or neighborhood technique where values of a forecast in space windows are compared relative to a point in the
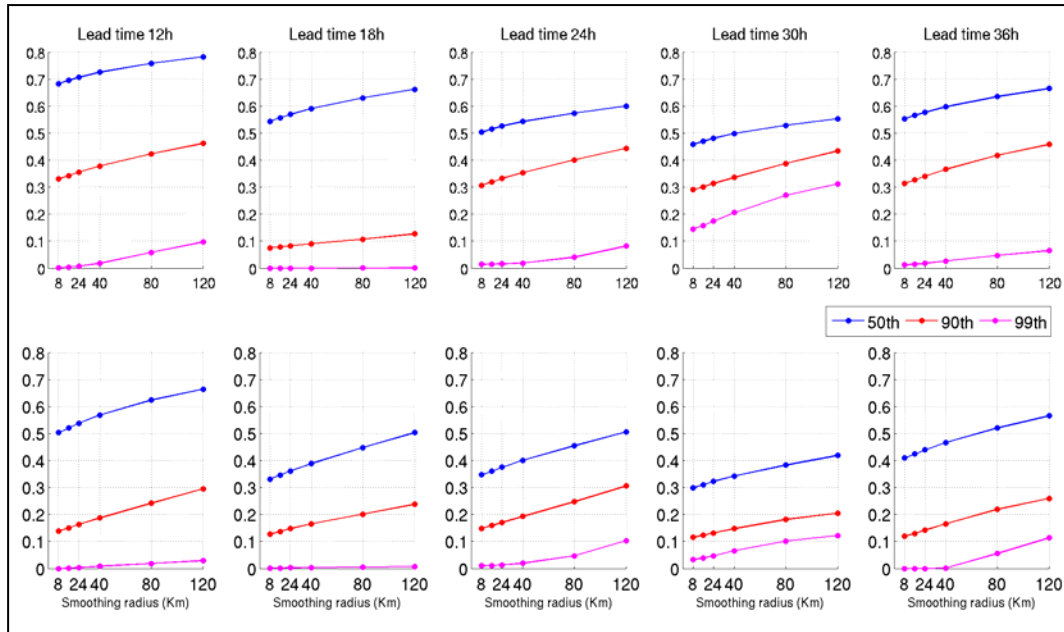
Fig. 6. FSS using percentile thresholds for 6-hr (top) and 1-hr (bottom) accumulated precipitation averaged over the verification domain using percentile thresholds as function of lead time

observation field. Thus, these methodologies are appropriate for high resolution forecasts where an evaluation of the skilful scales at which forecasts are useful can be achieved. In fact, to analyze the model performance at different horizontal scales, FSS was computed for different percentile thresholds for 12 to 36 hour forecast lead times. The size of the radius of influence considered varied from 8 km (1 grid point) up to 120 km (15 grid points). For a radius equal to 8 km (1 grid point), caution should be taken as it pairs CMORPH resolution.

Mittermaier and Roberts (2010) analyzed the skilful spatial scales using the FSS for idealized geometric and realistic cases. A 0.5 FSS value is found to be a critical threshold where greater values represent skilful scales and values tending to zero are more representative of random forecasts. They also show that forecast biases affect the FSS behavior in the way that as the neighborhood increases the FSS doesn't reach the optimum value of 1. On the other hand, they show that timing errors also result in the degradation of the FSS magnitude as well as if the verification domain has large areas with no rainfall.

Fig. 6 shows the FSS values obtained for different quantile-based thresholds and forecast lead times, for the 6-hr and 1-hr accumulated precipitations. As expected, FSS increased as the radius of influence increased. However, as thresholds increased, the FSS worsened at all

scales indicating that the model had poorer skill at forecasting heavy precipitation events. These results are in accordance with those presented by Schwartz *et al.* (2009). However, as well as the ETS behaviour, the FSS show higher values at the 36-hr compared to 30-hr lead time, which could be related to the sample size of events analyzed, location and spatial distribution maximum frequencies. Lower temporal resolutions like 6-hr accumulated forecasts (top panel) present some skilful for the lowest percentile threshold. On the other hand, hourly forecasts (bottom panel), present very poor performance with almost no skill between 18 hr and 30 hr lead times.

In the period studied, WRF model captures precipitating systems but it underestimates the amount of rainfall. Results are overall encouraging since are similar to those obtained with similar forecasting systems in other regions of the world (Kain *et al.*, 2008) and also with those reported for the same region at comparable model resolutions (Dillon *et al.*, 2013). Nevertheless, more research should be made to address the uncertainty in the forecast fields considering verification techniques against other sources of observations according to the model resolution.

Case study : 5th December, 2012

In this section a forecast evaluation of the most intense convective system observed during the verification
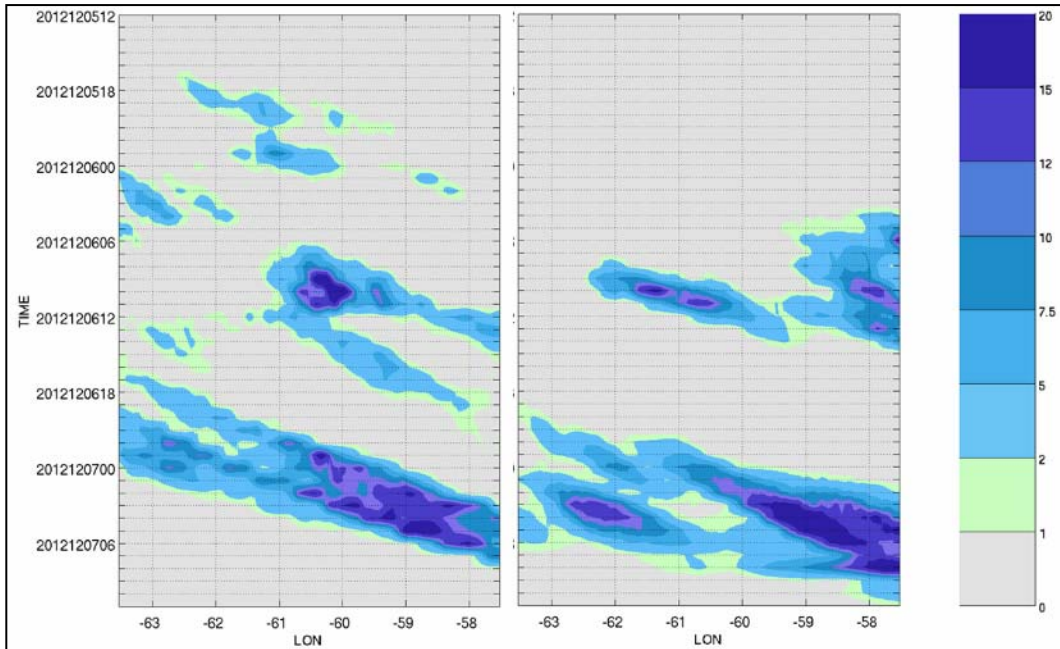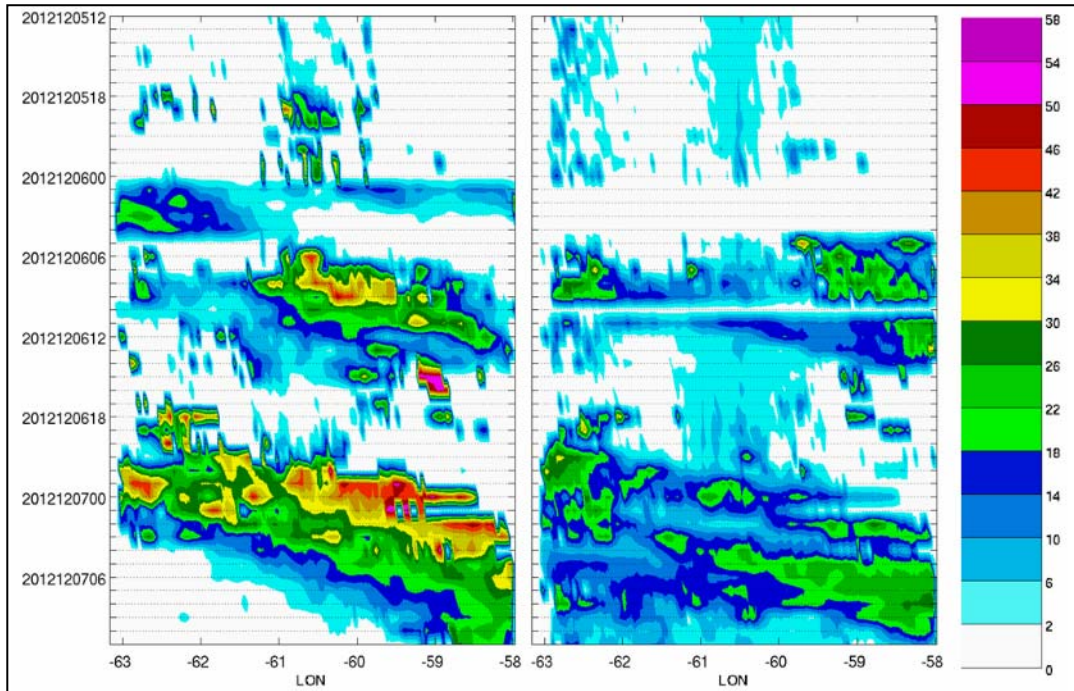
Fig. 7. Hövmoller diagrams for simulated (left) and estimated by CMORPH (right) 1-hr accumulated precipitation calculated between 31.37°S and 32.82°S. Forecast outputs are valid for the 12-36 hr lead time for runs corresponding to 5th and 6th December initialized at 0000 UTC. Values were averaged in latitude (31.37° S - 32.82° S) over the verification domain
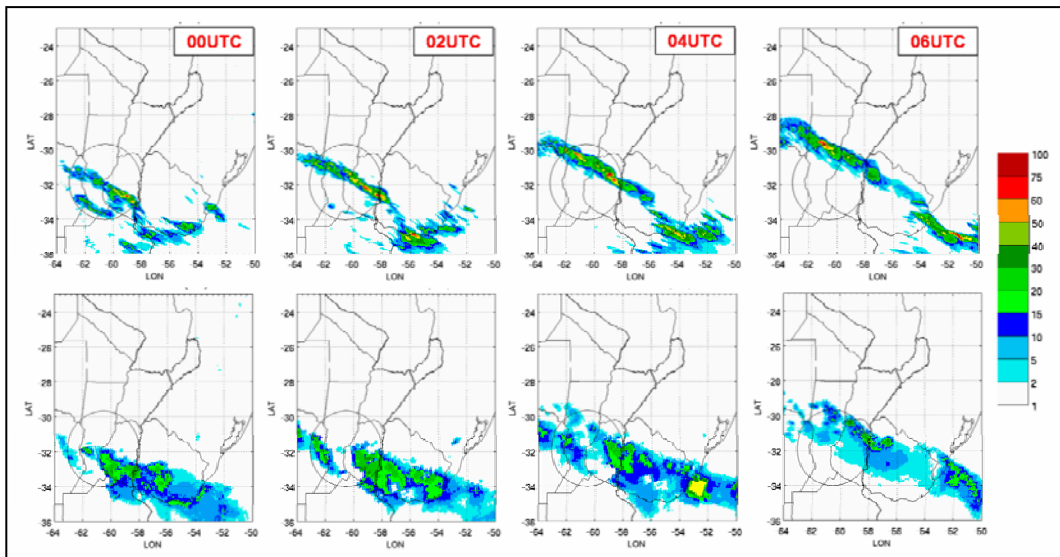
Fig. 8. Simulated (top panel) and corresponding CMORPH estimated (bottom panel) 1-hr rainfall (mm) for selected times on 7 December, 2012, associated with the model forecasts initialized at 0000 UTC 6 December, 2012. The black circle depicts the 240-km range of Parana radar

period is performed. Convective activity began on the 4th December, embedded in a synoptic situation dominated by the presence of a South American Low Level Jet (Salio *et al*., 2007). To illustrate the model performance in representing precipitation, 1-hr simulated precipitation is shown in Fig. 7. The maximum precipitation rates in this

period (over 15 mm/hr) are well captured by the model. It can be noticed that there were two most distinctive convective pulses, the first one developed during the 6th December reaching its maximum at 0600 UTC with values over 40 mm/hr. During the next hours, several convective cores developed ahead of an advancing cold

Fig. 9. Hövmoller diagrams for simulated (left) and observed (right) ColMax reflectivity (dBz) averaged between 31° S and 33° S. Forecast outputs are valid for the 12-36 hr lead time for runs corresponding to 5th and 6th December initialized at 0000 UTC. Observed values were calculated within the 240-km radar range (black circle in Fig. 8)

front. By 1800 UTC convection was better organized and aligned with the cold front moving towards the northeast.

The model developed a quasi-linear convective system oriented primarily northwest-southeast ahead of a cold front advancing to the northeast. Although the placement was accurately forecasted the rainfall area is narrower compared to that estimated by CMORPH, which is consistent with the systematic biases previously discussed. In this regard, the spatial structure of both simulated and estimated precipitation are quite different, CMORPH shows a broader band with lighter values and not as continue as the one forecasted by the model at 0000 UTC. Moreover, the model overestimated the intensity of heavy precipitation with small cores exceeding 60 mm/h while areas with less than 10 mm/hr are underestimated (Fig. 8).

Fig. 9 shows the simulated ColMax reflectivity from 5th to 7th December, 2012 at 1200 UTC, averaged between 31°S and 33°S. The observed averaged reflectivity from Parana radar is also shown for comparison. Spurious ColMax values around 2 dBz can be noticed in the radar data which are associated ground clutter. Most active convective activity is seemed to begin during the 6th December at 1800 UTC up to 1200 UTC on the 7th over the Parana radar area and southeastern Uruguay. The evolution of the highest echoes is well captured by the model though the highest echoes are overestimated. This could be associated to the fact that, during warm season, reflectivity values of Parana radar are underestimated during 2009-2013 heavy precipitation events as found by Pappalardo *et al*. (2014) using TRMM information. The west to east propagation speed of the system is also well represented by the model. Additional insight was gained as the convective structure was successfully captured by the model though the most intense echoes were overestimated and positioned ahead of the extended line (not shown). These results are comparable to those found by Kain *et al*. (2008) where they used base reflectivity at a fixed height instead of the column maximum values, although they also found similar systematic biases in the simulated reflectivity fields.

4.   Summary

During November-December 2012 as part of the CHUVA Santa Maria experiment, different institutions produced high resolution numerical forecasts over SESA in order to generate ensemble precipitation forecasts from convective allowing models. In this paper, an evaluation of the performance of convective-allowing WRF model precipitation forecasts produced at NMS-Argentina with 4-km horizontal resolution for the region was performed.

A 30-day verification period, as well as a particular case study, were analyzed. By examining traditional verification scores, it seems that WRF model forecast captures reasonably well the position and timing of the convective events, even though the magnitude was underestimated. Accumulated rainfall averaged over the verification domain are underestimated by the forecast as well as the areal coverage for small thresholds. It is important to note that these biases can be partially associated to biases present in the CMORPH precipitation estimates that has been documented in previous studies. In order to remove the biases present in the forecast and in the CMORPH estimates, verification metrics were computed using thresholds based on the quantiles of the rainfall rates distributions.

Were the methodologies correct to analyze the model performance? To understand the results obtained we should consider some limitations of the scores analyzed here. For example, the use of scores as ETS, POD or FSS cannot explain biases due to differences in timing and position or intensities of estimated and forecasted rainfall.

These were preliminary results for only a month period where only a few convective events were observed. An objective verification of convective-allowing forecasts over the region has to be carried out over a longer period in order to extend the analysis presented in this work and in order to obtain more statistically robust measures of forecast skill. Furthermore, a comparative analysis with the operational lower-resolution parameterized-convection forecasts will provide additional insight on the advantages of convective-allowing forecasts over the region. Moreover, accordingly to the time-space scale of the convective events analyzed, other verification techniques could be applied. Spatial verification approaches should be considered such as object-oriented methods like the Contiguous Rain Areas Method (CRA, Ebert and Mc Bride, 2000) in order to understand other features of the forecast evaluation. Exploring other methodologies as discussed in Gilleland *et al.* (2009) in agreement with the observational dataset available could provide more information about the model's performance and precipitation small-scale features. On the other hand, calibration techniques, in agreement with the available observations, should be explored in order to account for a bias reduction.

Nowadays, a forecast system similar to the one implemented for the experiment is running operationally at the NMS-Argentina. A forecast database is being generated, so a validation of precipitation forecast for a longer period could be achievable. Other spatially-based or "fuzzy" verification techniques could present an alternative for the high resolution rainfall forecasts

evaluation enabling a better representativeness of the uncertainty of these highly variable fields. A similar approach as the one applied here could be explored by the use of timing windows for the FSS calculation. On the other hand, more knowledge about the model's skill could be gained by examining seasonal variability and focusing the analysis over other sub-regions based on rainfall regimes. Pursuing and assessing different operational calibration strategies of radar-based information and satellite rainfall estimates could lead to better results.

References

Baldwin, M. E. and Mitchell, K. E., 1998, "Progress on the NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research", Preprints, Second Symp. on Integrated Observing Systems, Phoenix, AZ, *Amer. Meteor. Soc*., 10-11.

Chen, F. and Dudhia, J., 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.

Demaria, E. M. C., Rodriguez, D. A., Ebert, E. E., Salio, P. , Su, F. and Valdes, J. B., 2011, "Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach", *J. Geophys. Res*., **116**, D08103.

Dillon, M. E., García Skabar, Y. and Nicolini, M., 2013, "Desempeño del pronóstico de modelos de alta resolución, en un área limitada: análisis de la estación de verano 2010-2011", *Meteorologica*, **38**, 2, 69-89 (in Spanish).

Ebert, E. E., 2008, "Fuzzy verification of high resolution gridded forecasts: A review and proposed framework", *Meteor. Appl.*, **15**, 51-64.

Ebert, E. E., Janowiak, J. E. and Kidd, C., 2007, "Comparison of near-real-time precipitation estimates from satellite observations and numerical models", *Bull. of the Amer. Met. Soc*., **88**, 47-64.

Ebert, E. E. and McBride, J. L., 2000, "Verification of precipitation in weather systems: Determination of systematic errors", *J. Hydrology*, **239**, 179-202.

Fowle, M. A. and Roebber, P. J., 2003, "Short-range (0-48 hr) numerical prediction of convective occurrence, mode, and location", *Wea. Forecasting*, **18**, 782-794.

GarcíaSkabar, 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**, 9-10.

Gilleland, E., Ahijevych, D., Casati, B. and Ebert, B., 2009, "Intercomparison of spatial forecast verification methods", *Wea. Forecasting*, **24**, 1416-1430.

Hamill, T. and Juras, J., 2006, "Measuring forecast skill: Is it real skill or is it the varying climatology?", *Quart. J. Roy. Meteor. Soc.*, **132**, 2905-2923.

Hogan, R. J., Ferro, C. A. T., Jolliffe, I. T. and Stephenson, D. B., 2010, "Equitability Revisited: Why the "Equitable Threat Score" Is Not Equitable", *Wea. Forecasting*, **25**, 710-726.

Hong, S. Y., Noh, Y. and Dudhia, J., 2006, "A new vertical diffusion package with an explicit treatment of entrainment processes", *Mon. Wea. Rev*., **134**, 2318-2341.

Hong, S. Y. and Lim, J. O. J., 2006, "The WRF single-moment microphysics scheme (WSM6)", *J. Korean Meteor. Soc*., **42**, 129-151.

Huffman, George, J. and Coauthors, 2007, "The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales", *J. Hydrometeor*, **8**, 38-55. doi: http://dx.doi.org/10.1175/ JHM560.1

Jenkner, J., Frei, C. and Schwierz, C., 2008, "Quantile-based short-range QPF evaluation over Switzerland", *Meteorol. Z.*, **17**, 827-848.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. and Xie, P., 2004, "CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution", *J. of Hydromet*., **5**, 487-503.

Kain, J. S., Weiss, S. J., Bright, D. R., Baldwin, M. E., Levit, J. J., Carbin, G. W., Schwartz, C. S., Weisman, M. L., Droegemeier, K. K., Weber, D. B. and Thomas, K. W., 2008, "Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP", *Wea. Forecasting*, **23**, 931-952.

Koch, S. E., Ferrier, B., Stoelinga, M. T., Szoke, E., Weiss, S. J. and Kain, J. S., 2005, "The use of simulated reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts", Preprints, 11th Conference on Mesoscale Processes / 32nd Conference on Radar Meteorology, Albuquerque, New Mexico, *Amer. Met. Soc*., CD-ROM.

Matsudo, C and P. Salio, 2011, "Severe weather reports and proximity to deep convection over Northern Argentina", *Atmospheric Research*, **100**, 4, June 2011, 523-537, Special Issue 5th European Conference on Severe Storms.

Mittermaier, M. and Roberts, N., 2010, "Intercomparison of spatial verification methods: identifying skillful spatial scales using the fractions skill score", *Wea. and For*., **25**, 343-354.

Pappalardo, L., Salio, P. and Vidal, L., 2014, "Estudio de la organización y condiciones del entorno de los eventos altamente precipitantes en la zona central de la Argentina", University of Buenos Aires Dissertation. In Spanish.

Roberts, N. M. and Lean, H. W., 2008, "Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events", *Mon. Wea. Rev*., **136**, 78-97.

Ruiz, J., 2009, "CMORPH precipitation estimates calibration and verification over South America", *Revista Brasileira de Meteorologia*, **24**, 474-488 (in Spanish).

Salio, P., Hobouchian, M. P., Garcıa Skabar, Y. and Vila, D., 2014, "Evaluation of High-Resolution Satellite Precipitation Estimates over Southern South America using a Dense Rain Gauge Network", *Atmo. Res*., doi:10.1016/j.atmosres.2014.11.017.

Salio, P., Nicolini, M. and Zipser, E., 2007, "Mesoscale convective systems over southeastern South America and their relationship with the South American Low-Level Jet", *Mon. Wea. Rev*., **135**, 1290-1309.

Schwartz, C. S. Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J., Coniglio, M. C. and Wandishin, M. S., 2010, "Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership", *Wea. Forecasting*, **25**, 263-280.

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., Thomas, K. W., Levit, J. J. and Coniglio, M. C., 2009, "Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing", *Mon. Wea. Rev*., **137**, 3351-3372.

Scofield, R. A. and Kuligowski, R. J., 2003, "Status and Outlook of Operational Satellite Precipitation Algorithms for Extreme-Precipitation Events", *Wea. Forecasting*, **18**, 1037-1051. doi: http://dx.doi.org/10.1175/1520-0434(2003)018<1037:SAOOOS >2.0.CO;2

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W. and Powers, J. G., 2008, "A Description of the Advanced Research WRF Version 3", Technical Note TN-475 + STR. NCAR: Boulder, CO; p125.

Vera, C., Baez, M., Douglas, M., Emmanuel, C. B., Marengo, J., Meitin, J., Nicolini, M., Nogues-Paegle, J., Paegle, J., Penalba, O., Salio, P., Saulo, C., Silva Dias, M. A. and Zipser, E., 2006, "The South American Low-Level Jet Experiment (SALLJEX)", *Bull. of the Amer. Met. Soc*., **87**, 63-77.

Vila, D. A., Scofield, R. A. and Davenport, J. C., 2002, "Satellite rainfall estimation over South America: Evaluation of two major events", AMS 16th Conference on Hydrology, 33-36.

Vila, D. A., L. de Goncalves, Toll, D. L. and Rozante, J. R., 2009, "Statistical Evaluation of Combined Daily Gauge Observations and Rainfall Satellite Estimates over Continental South America", J. Hydrometeor, **10**, 533-543. doi: http://dx.doi.org/ 10.1175/2008JHM1048.1.

Weisman, M. L., Davis, C., Wang, W., Manning, K. W. and Klemp, J. B., 2008, "Experiences with 0–36-h explicit convective forecasts with the WRF-ARW Model", *Wea. Forecasting*, **23**, 407-437.

Wilks, D. S., 2006, "Statistical Methods in the Atmospheric Sciences. An Introduction", 2nd ed., Academic Press, San Diego, U.S.A., p467.

WWRP/WGNE Joint Working Group on Verification, 2009, "Recommendations for the verification and intercomparison of QPFs and PQPFs from operational NWP models", WMO Tech Doc. 1485, WWRP 2009-1, revision 2, 37 pp. Available at http://www.wmo.int/pages/prog/arep/wwrp/new/documents/W WRP2009_1.pdf, 2009.