



# AMERICAN METEOROLOGICAL SOCIETY

*Journal of Applied Meteorology and Climatology*

## **EARLY ONLINE RELEASE**

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/JAMC-D-15-0010.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Vendrasco, E., J. Sun, D. Herdies, and C. Angelis, 2015: Constraining a 3D-Var Radar Data Assimilation System with Large-scale Analysis to Improve Short-range Precipitation Forecast. *J. Appl. Meteor. Climatol.* doi:10.1175/JAMC-D-15-0010.1, in press.



1           **Constraining a 3D-Var Radar Data Assimilation System with**  
2           **Large-scale Analysis to Improve Short-range Precipitation Forecast**

3  
4  
5   Vendrasco<sup>1</sup>, Eder Paulo; Juanzhen Sun<sup>2</sup>; Dirceu Luis Herdies<sup>1</sup> and Carlos Frederico de Angelis<sup>3</sup>

6  
7                           <sup>1</sup>National Institute for Space Research (INPE)

8                           Cachoeira Paulista, Sao Paulo - Brazil

9                           <sup>2</sup>National Center for Atmospheric Research (NCAR)

10                          Boulder, Colorado - United States

11                          <sup>3</sup>National Center for Monitoring and Early Warning of Natural Disaster

12                          Cachoeira Paulista, Sao Paulo - Brazil

13  
14  
15  
16 \_\_\_\_\_

17   Corresponding author address: Eder Paulo Vendrasco - Rodovia Presidente Dutra, km 39, Cachoeira  
18   Paulista, São Paulo, CEP 16300-000 - Brasil. Email: eder.vendrasco@cptec.inpe.br

19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## Abstract

It is known from previous studies that radar data assimilation can improve the short-range forecast of precipitation, mainly when radial wind and reflectivity are available. However, from our experience the radar data assimilation, when using the 3D-Var technique, can produce spurious precipitation and large errors on the position and amount of precipitation. One possible reason for the problem is attributed to the lack of proper balance in the dynamical and microphysical fields. This work attempts to minimize this problem by adding a large-scale analysis constraint in the cost function. The large-scale analysis constraint is defined by the departure of the high resolution 3D-Var analysis from a coarser resolution large-scale analysis. It is found that this constraint is able to guide the assimilation process in such a way that the final result still maintains the large-scale pattern, while adding the convective characteristics where radar data are available. As a result, the 3D-Var analysis with the constraint is more accurate when verified against an independent dataset. The performance of this new constraint on improving precipitation forecast is tested using six convective cases and verified against radar-derived precipitation by four skill indices. All skill indices show improved forecast when using the methodology presented in this paper.

## 40 **1. Introduction**

41           In recent years, an increase of natural disasters such as tornados and extreme precipitation  
42 events has been globally observed (Coumou and Rahmstorf 2012). Therefore, there has been an  
43 increased demand for improved forecasting of severe convective weather and its associated  
44 hazards. An improvement can be achieved by either increasing the grid resolution, developing  
45 physical parameterization schemes suitable for high-resolution models, or improving the initial  
46 conditions. Roberts and Lean (2008), for example, conducted experiments using 12, 4 and 1-km  
47 horizontal resolutions for 10 days of summer convection cases and their results showed that the  
48 1-km model grid was more skillful than the 12-km model grid at all horizontal scales greater than  
49 15 km for heavy precipitation. Besides the model resolution, the initial conditions also play an  
50 important role in the skill of convective forecasting and, therefore, it is crucial to initialize the  
51 model using observations that describe not only the large scale, but also the convective scale.  
52 The use of convective scale observations such as radar data through data assimilation can  
53 potentially improve the initial conditions and lead to a better convective weather prediction. Data  
54 assimilation is a technique for generating an accurate image of the true state of the atmosphere at  
55 a given time in which the observed information is accumulated into the model state by taking  
56 advantage of consistency constraints with laws of time evolution and physical properties.

57           According to Bouttier (2009) the most challenging aspects of the fine scale data  
58 assimilation stem from the occurrence of complex, intermittent convective structures in the  
59 atmosphere (and in numerical models) that have a strong sensitivity to initial conditions and a  
60 potential for rapid error growth. On the one hand, kilometer scale numerical weather prediction  
61 (NWP) systems can better provide important background information needed for convective-  
62 scale data assimilation, including low-level winds, convergence lines, tropospheric humidity, and

63 preexisting convective systems that may influence subsequent convective triggering. Besides, the  
64 location and magnitude of these features can have significant differences from the reality of the  
65 atmosphere.

66 Doppler radar observations (reflectivity and radial velocity) have been used in complex data  
67 assimilation systems in order to improve the initial conditions for convection-permitting models  
68 (e.g., Caya et al. 2005; Sun et al. 2005; Tong and Xue 2005; Gao et al. 2007; Xiao et al. 2007;  
69 Ming et al. 2009; Aksoy et al. 2009, 2010; Wang et al. 2013a) since they are almost the only  
70 source of three-dimensional information at this scale (Aksoy et al. 2009). These data assimilation  
71 systems (e.g., Variational methods (3D-Var (Sasaki 1970; Lorenc 1986) and 4D-Var (Lorenc  
72 1988)), Ensemble Kalman Filter (Evensen 1994) and Hybrid methods (Zupanski 2005)) have  
73 been developed for both research and operational purposes and previous works using those  
74 systems have shown improvement in precipitation forecast (e.g., Snyder and Zhang 2003;  
75 Kawabata et al. 2007; Tong and Xue 2008; Shimizu et al. 2011; Sun et al. 2012; Sun and Wang  
76 2013; Wang et al. 2013b).

77 Although the above-mentioned studies have demonstrated that radar data assimilation can  
78 improve the forecast of precipitation, the positive impact tends to last only for a few hours. Using  
79 a 3D-Var radar data assimilation system, Sun et al. (2012) found that the impact of assimilating  
80 Doppler radars lasted only for three hours for forecasts initialized at early mornings when the  
81 convective systems are mostly scattered and radar coverage is small. From our experience with  
82 3D-Var, we have found that radar data assimilation can introduce large wind increments and  
83 spurious precipitation, which can potentially make the positive impact of radar data assimilation  
84 short-lived and even cause adverse effect on longer forecast. We believe one of the reasons for  
85 this behavior could be due to the lack of proper balance in the dynamical and microphysical

86 fields of the initial analysis. According to Reen (2007), the 3D-Var method tends to produce  
87 noise and a digital filter (Lynch, 1993) can be used in an attempt to diminish it. However,  
88 besides the fact that the digital filter may also remove real atmospheric features, it can only be  
89 implemented after the data assimilation process. Another methodology that tries to deal with the  
90 noise produced in a 3D-Var system is the Incremental Analysis Update (IAU; Bloom et al.  
91 1996). Lee et al. (2006) performed experiments using the fifth-generation Pennsylvania State  
92 University-NCAR Mesoscale Model (MM5) 3D-Var with the IAU methodology. The authors  
93 found that IAU could reduce the initial noise and the moisture (i.e., the model variables related  
94 with the water cycle) spun up efficiently because of the gradual incorporation of the analysis  
95 increment. However, the IAU method is unable to get rid of the noise in the analysis; rather, it  
96 reduces the impact of the noise by the gradual addition of the increment during the forecast.  
97 Although 4D-Var and hybrid (4D-Var+EnKF) systems may have advantages in producing better  
98 balanced analysis and hence a greater potential than the 3D-Var, their computational costs are  
99 still much higher than what many operational centers can afford.

100         Sugimoto et al. (2009) conducted experiments assimilating radar data using the Weather  
101 Research and Forecasting Model - WRF (Skamarock et al. 2008) three-dimensional variational  
102 data assimilation system - WRFDA 3D-Var (Barker et al. 2004) with observation simulation  
103 system experiments. Although encouraging results were shown in their study, they also found  
104 that the 3D-Var had limited ability in retrieving the unobserved variables by radar. This implies  
105 the wind increment resulting from the assimilation of radar radial velocity is not balanced by  
106 other fields at the convective scales and can contain significant noise. The noise can even  
107 contaminate the large-scale balance. Guo et al. (2007) and Ming et al. (2009) addressed the issue  
108 of the final analysis balance when using the WRFDA 3D-Var. The authors subjectively reduced

109 by half these background error variance in order to give more weight to the background during  
110 the assimilation process for the purpose of keeping the background dynamical balance.

111 Radar observations such as reflectivity and wind radial velocity carry information about  
112 convective forcing in the atmosphere and such high-resolution information ( $\approx 100$  m - 1km)  
113 cannot be accurately resolved in mesoscale models due to their lower resolutions and  
114 parameterizations. Therefore the initialization of the convective-scale flow has to largely rely on  
115 the radar observations. However, because radar observations consist of a large amount of data,  
116 they can dominate the analysis results by adding large imbalanced wind increments if the cost  
117 function is not properly constrained, especially when convective systems are present. Although  
118 the method to tune the variance and length scale (Guo et al. 2007; Ming et al. 2009) can alleviate  
119 the problem to some extent, it can also reduce the impact of the radar observations on the  
120 analysis and hence the precipitation forecast.

121 One of the challenges in convective-scale data assimilation using radar observations is to  
122 extract as much information as possible from the observations while maintaining the large-scale  
123 balance that is in the background. Or in other words, adding radar data to the initial conditions  
124 through a data assimilation system should not damage the large-scale pattern nor cause spurious  
125 convection. In this paper, we demonstrate a method to minimize the imbalance problem in a 3D-  
126 Var system by adding a constraint in the cost function using large-scale analyses. This large-  
127 scale analysis constraint (LSAC) is defined by the departure of a high resolution 3D-Var analysis  
128 from a coarser resolution large-scale analysis. We will show that LSAC can guide the  
129 assimilation process in a such way that the final result still maintains the large scale pattern by  
130 minimizing the noise, especially in the data sparse regions, while adding the convective  
131 characteristics from the radar data.

132 This paper is organized as follows. Section 2 describes the WRF 3D-Var assimilation system  
 133 employed in this study and the implementation of the large scale analysis constraint. The  
 134 methods used to evaluate the skill of precipitation forecast are also described in this section.  
 135 Section 3 presents the radar data used for this work and outlines the experimental setup. In  
 136 section 4, we first compare the results of precipitation forecasts between the experiments with  
 137 and without the LSAC for 6 convective cases and then conduct a detailed study on one case to  
 138 examine how the LSAC impacts the 3D-Var analysis. The findings from this study are  
 139 summarized in section 5.

## 140 **2. Methodology**

### 141 *a. WRFDA 3D-Var*

142 The assimilation system used in this study is the WRFDA 3D-Var (Barker et al. 2004)  
 143 version 3.4. It iteratively minimizes the cost function that is defined using the incremental  
 144 formulation (Courtier et al., 1994):

$$145 \quad J = J_b + J_o = \frac{1}{2} \mathbf{v}^T \mathbf{v} + \frac{1}{2} (\mathbf{d} - \mathbf{H}' \mathbf{U} \mathbf{v})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}' \mathbf{U} \mathbf{v}) \quad (1)$$

146 where  $J_b$  and  $J_o$  stand for the background and observation terms, respectively. The term  $\mathbf{v}$  is the  
 147 control variable (CV) defined by  $\mathbf{v} = \mathbf{U}^{-1}(\mathbf{x} - \mathbf{x}_b)$ , where  $\mathbf{U}$  is the decomposition of the background  
 148 error covariance  $\mathbf{B}$  via  $\mathbf{B} = \mathbf{U} \mathbf{U}^T$ ,  $\mathbf{x}$  is the full analysis variable, and  $\mathbf{x}_b$  is the background variable.  
 149 The innovation vector that measures the departure of the observation  $\mathbf{y}_o$  from its counterpart  
 150 computed from the background  $\mathbf{x}_b$ , is given by  $\mathbf{d} = \mathbf{y}_o - \mathbf{H}(\mathbf{x}_b)$ .  $\mathbf{H}'$  is the linearization of the  
 151 nonlinear observation operator  $\mathbf{H}$ , and  $\mathbf{R}$  is observation error covariance matrix.  
 152

153 The control variables used in this study are velocity components  $u$  and  $v$ , temperature  $T$ ,  
154 surface pressure  $P_s$ , and pseudo relative humidity  $RHs$  (humidity divided by its background).  
155 These are the same as previous radar data assimilation studies using WRFDA 3D-Var except for  
156 the absence of momentum variables. A recent study by Sun et al. (2014) compared the original  
157 momentum control variables of stream function and velocity potential ( $\psi$  and  $\chi$ ) with that of  
158 velocity components  $u$  and  $v$  and concluded that the  $u/v$  pair outperformed  $\psi/\chi$  for limited-area  
159 convective-scale data assimilation. They found that the use of the  $u/v$  control variables allowed  
160 closer fits to dense observations such as radar radial velocity and resulted in improved  
161 precipitation forecast. The  $u/v$  momentum control variables have been used in other variational  
162 convective-scale data assimilation systems in previous studies (i.e., Sun and Crook 1997; Zou et  
163 al. 1995; Gao et al. 1999). We tested both control variable options and obtained conclusions  
164 consistent with that of Sun et al. (2014). For reflectivity data assimilation, we follow the indirect  
165 assimilation scheme described by Wang et al. (2013a) in which the retrieved rainwater instead of  
166 the reflectivity itself is assimilated. The authors argued that assimilating the rainwater mixing  
167 ratio avoided non-linearity issues caused by the linearized observation operator for reflectivity  
168 required in the incremental formulation.

#### 169 *b. Large Scale Analysis Constraint*

170 The LSAC is introduced to the WRFDA 3D-Var by adding a new term  $J_c$  to Eq. 1 that  
171 measures the deviation of the 3D-Var analysis from a coarser resolution large-scale analysis for  
172 horizontal velocity components, temperature and humidity:

173

$$J = J_b + J_o + J_c \quad (2)$$

$$= J_b + J_o + \frac{1}{2}(\mathbf{d}_c - \mathbf{H}'\mathbf{U}\mathbf{v})^T \mathbf{R}_c^{-1}(\mathbf{d}_c - \mathbf{H}'\mathbf{U}\mathbf{v})$$

174

175 where  $\mathbf{d}_c$ , given by  $\mathbf{d}_c = \mathbf{y}_c - \mathbf{H}(\mathbf{x}_b)$ , is the innovation vector that measures the departure of the  
 176 LSAC  $\mathbf{y}_c$  from its counterpart computed from the background  $\mathbf{x}_b$ . The  $\mathbf{y}_c$  variable is the  
 177 meridional and zonal wind components ( $u$  and  $v$ ), the temperature ( $T$ ) and the water vapor  
 178 mixing ratio ( $q_v$ ) from the large scale analyses which are being assimilated as observations. In  
 179 this study, we use the National Center for Environment Prediction's (NCEP) Operational Global  
 180 Forecast/Analysis System (GFS) global model analysis.  $\mathbf{R}_c$  is the large-scale analysis error  
 181 covariance matrix for  $u$ ,  $v$ ,  $t$  and  $q_v$ . The  $\mathbf{R}_c$  matrix is constructed by considering constant  
 182 uncorrelated errors for each variable, being  $2.5 \text{ m s}^{-1}$  for wind components,  $2 \text{ }^\circ\text{C}$  for temperature  
 183 and  $3 \text{ g kg}^{-1}$  for water vapor mixing ratio. These values are chosen based on the diagnostics of  
 184 the performance of the NCEP's GFS available at the Environmental Modeling Center's webpage  
 185 (<http://www.emc.ncep.noaa.gov/GFS>). The sensitivity of the analysis and forecast with respect to  
 186 the error variance was evaluated by performing (not shown) experiments that increase or  
 187 decrease the errors by 10% and no significant sensitivity was found. The large-scale analysis  
 188 data used in LSAC is mapped to the closest 3D-Var analysis grid point by tri-linear interpolation.

### 189 *c. Methods for precipitation verification*

190 Our primary goal for implementing radar data assimilation is to improve the precipitation  
 191 forecast. Therefore, besides some diagnoses to evaluate the quality of the 3D-Var analysis, skills  
 192 in precipitation forecast are also used to measure the performances of the data assimilation  
 193 experiments. The Fractional Skill Score (FSS - Roberts and Lean 2008) and LRMSE (Local Root  
 194 Mean Square Error) are applied for the precipitation verification. The FSS is defined by

195

$$FSS = 1 - \frac{FBS}{FBS_w} = 1 - \frac{\frac{1}{N} \sum_{k=1}^N [P_{M(k)} - P_{O(k)}]^2}{\frac{1}{N} (\sum_{k=1}^N P_{M(k)}^2 + \sum_{k=1}^N P_{O(k)}^2)} \quad (3)$$

196

197 where  $P_{M(k)}$  and  $P_{O(k)}$  are the forecasted and observed fractional coverage of precipitation at the  
198  $k^{th}$  grid point that exceeds a given threshold value, and  $N$  is the total number of grid points in the  
199 verification domain. The thresholds used in the verification were 1 and 5 mm and a radius of 20  
200 and 10 km, respectively. A smaller radius was used in the case of 5 mm in order to be more  
201 rigorous in the verification of heavier precipitation. FSS is equal to 1 when the forecast is  
202 perfect, which occurs as  $P_M$  and  $P_O$  are equal. The LRMSE is defined by

203

$$LRMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (\overline{M}_k - \overline{O}_k)^2} \quad (4)$$

204

205 where  $\overline{M}$  and  $\overline{O}$  are the forecasted and observed precipitation averaged over all the grid points  
206 within the same radius of influence used for calculating the FSS. The subscript  $k$  represents the  
207  $k^{th}$  grid point and  $N$  the total number of grid points in the verification domain. In addition to the  
208 above two scores, the traditional skill scores of the False Alarm Ratio (FAR) and Probability of  
209 Detection (POD) are also used to verify the precipitation forecast. The FAR and the POD are  
210 defined by:

$$FAR = \frac{\text{false alarms}}{\text{false alarms} + \text{hits}} \quad (5)$$

211

$$POD = \frac{hits}{hits + misses} \quad (6)$$

212

213 and the contingency table is defined in the table 1.

214         The precipitation data used for this evaluation is estimated from radar reflectivity  
215 applying the reflectivity-precipitation relationship (*Z-R* relationship) giving by  $Z=aR^b$  (Marshall  
216 and Palmer 1948) where *a* and *b* are parameters estimated empirically. The advantage of using  
217 precipitation obtained from radar reflectivity is its spatial and temporal resolution, besides its  
218 spatial coverage. Despite the rain gauge gives the amount of precipitation directly, it does not  
219 have a spatial coverage, but only provide information for a specific location. It would be  
220 necessary a very dense rain gauges network to get a good spatial coverage. Since *a* and *b* are  
221 obtained empirically, they will vary for each type of precipitation, location and many other  
222 factors, for example, Battan (1973) provided a list of 69 different *Z-R* relationships based on  
223 climatic conditions for different parts of the world. There are also examples in recent literature  
224 describing the uncertainty and new methodologies of radar estimated precipitation (e.g., Maity et  
225 al. 2015). However, the traditional *Z-R* relationship is still widely used nowadays due to its  
226 simplicity and acceptable uncertainties, if applied correctly.

### 227 **3. Observations and experimental setup**

228         Assimilated observations in this study include radar data from the CHUVA project  
229 (Cloud processes of the main precipitation systems in Brazil: A contribution to cloud resolving  
230 modeling and to the Global Precipitation Measurement – GPM, Machado et al. (2014)) and the  
231 surface and upper-air conventional data from the Global Telecommunication System. The

232 CHUVA project aims to map the main precipitating systems in Brazil. So far, experiments have  
233 been performed at seven sites with various measurements including those with dual polarization  
234 radar, lidar, microwave radiometer, disdrometer, radiosonde, rain-gauge and various other  
235 instruments. Only data from the Paraíba Valley Experiment were used and six convective cases  
236 were chosen to examine the proposed new LSAC methodology. Figure 1 shows the topography  
237 of this region, the X-band radar site used in this study, and the surface network used for  
238 verification. Table 2 summarizes the six convective cases studied in this paper. Figure 2 shows  
239 the 3-km CAPPI (Constant Altitude Plan Position Indicator) for each case at 18 UTC observed  
240 by the X-band radar. These convective systems all produced localized heavy rainfall. Due to the  
241 transient and small-scale nature of storms, it is challenging for operational NWP models to  
242 accurately forecast them.

243         The São Paulo state (see Fig. 1) is one of the most important regions in Brazil, not only  
244 because of its economic prosperity, but also because of the high population density. São Paulo  
245 city, the state capital, is located in the eastern part of the São Paulo state and it is the biggest city  
246 in the country with around 12 million people. The Sao Paulo Metropolitan Area (SPMR) consists  
247 of 39 municipalities, including São Paulo city, and the total population is over 20 million. The  
248 Paraíba Valley is another important region in the São Paulo state, where a highly developed  
249 industrial park and a space station are located. Because of the importance of this region and the  
250 recurrent flood episodes during the summer season in the SPMR, it is very important to improve  
251 the data assimilation and forecast systems over this area to deliver the quantitative precipitation  
252 forecast (QPF) as precise as possible.

253         The conventional observations are obtained from GTS and they include airep, metar,  
254 synop, ship and sound in this region. Figure 3 shows the locations of the conventional data. As

255 can be seen, the amount of data is quite small and sparse, even though this region is very  
256 important due to the factors explained before. Particularly in this area, due to the limited amount  
257 of GTS data and the small domain, the impact of the radar data is very strong and can dominate  
258 over GTS observations. The parameterizations used in the WRF forecast include the WRF  
259 single-moment 6-Class microphysics (WSM6 - Hong and Lim 2006), the Rapid Radiative  
260 Transfer Model longwave radiation (Mlawer et al. 1997), the Dudhia shortwave radiation scheme  
261 (Dudhia 1989) and the Yonsei University (YSU) Planet Boundary Layer (Hong et al. 2006). The  
262 initial and boundary conditions are from the GFS 0.5°x0.5° analysis and the model domain is  
263 shown in Fig. 1. The horizontal grid resolution is 2-km and the vertical resolution is about 60-m  
264 near the surface and 1-km in higher levels with a total of 45 sigma levels. Previous studies has  
265 shown (e.g., Bryan and Morrison 2012) that 60-m vertical resolution is sufficient for modeling  
266 convection in the planetary boundary layer. A small domain was chosen because the goal was to  
267 produce high horizontal resolution QPF for the nowcasting range of 6 hours and also due to the  
268 use of a single radar that is strategically located in the Paraíba Valley close to the SPMR.  
269 Obtaining a QPF with high accuracy is very important in this area in order to give support in case  
270 of natural disasters (e.g., floods).

271         The NMC (National Meteorological Center) method (Parrish and Derber 1992) was used  
272 in this work to generate the background error statistics using the gen\_be utility from the WRFDA  
273 system. A dataset containing 3 months of cold-start 24-h forecasts over a domain covering the  
274 Southern Hemisphere summer was produced every day starting at 00 and 12 UTC. The  
275 differences between the 24- and 12-h forecasts valid at the same times were used to calculate the  
276 domain-averaged background error statistics.

277 For each of the 6 cases, the data assimilation and forecast experiment commenced at 12  
278 UTC with the GFS analysis as the initial conditions. A 3-hour forecast was conducted before the  
279 cycled 3D-Var analyses begin. Four continuously cycled analyses were performed at 15, 16, 17,  
280 and 18UTC and then a 6-hour forecast ensued. The assimilation process was divided into 2 steps:  
281 i) first the GTS data were assimilated and then ii) the radar data were assimilated. In the second  
282 step, if the switch for the LSAC was turned on then the large-scale analysis was assimilated  
283 together with the radar data. The cycling strategy is illustrated in Fig. 4.

284 This strategy allows for the separation of the large- and small-scale data and the use of  
285 different background variances and length scales (Ha and Lee 2012; Tong et al. 2014) at each  
286 step. In our experiments, the background error statistics were used without tuning for the GTS  
287 data assimilation in the first step, assuming that the background covariance matrix properly  
288 represents the background errors. For the radar data assimilation in the second step, the error  
289 statistics were tuned by halving the length scale and doubling the variance, following Tong et al.  
290 (2014).

291 Three numerical experiments (Table 3) were performed to evaluate the impact of the  
292 LSAC on the analysis and precipitation forecast when assimilating radar data. The CRT is the  
293 control run in which only conventional data from GTS are assimilated. In the other 2  
294 experiments, LSAC<sub>Off/On</sub>, both conventional and radar data are assimilated without/with the  
295 constraint term. The radar data are thinned to the 1-km resolution from the original 250-m and  
296 interpolated horizontally into a regular grid while the original polar coordinate is kept in the  
297 vertical.

#### 298 **4. Results**

299           The 6-h precipitation forecasts are evaluated for all 6 convective cases using the averaged  
300 FSS and LRMSE over the 6 cases and their results are shown by Fig. 5. The verification is  
301 against the hourly precipitation estimate obtained from the radar reflectivity. The experiment  
302 LSAC<sub>Off</sub> shows improvement over the control run (CTR) for both thresholds of 1 and 5 mm h<sup>-1</sup> ,  
303 which is expected and consistent with previous works. The radar data assimilation usually  
304 improves the forecast in the beginning and then it can lose skill depending on many factors like  
305 the forcing mechanism of the convective systems. If more weight is given to the radar data  
306 during the analysis process, the precipitation forecast can result in better skill in the first 1 or 2  
307 hours. However, due to the imbalance generated by forcing the model toward the radar data the  
308 skill can drop and become worse than the case without any radar data assimilation. Nevertheless,  
309 in the experiment LSAC<sub>Off</sub>, all 6-h forecasts are improved with the radar data assimilation on  
310 average over the 6 convective cases.

311           The LSAC<sub>On</sub> experiment results in improvement over the LSAC<sub>Off</sub> experiment for all 6-h  
312 forecasts in the averaged FSS. Figure 5 shows that the LSAC<sub>On</sub> experiment reduces the LRMSE,  
313 since it decreases the spurious convection (shown later). The improvement obtained in the  
314 LRMSE is due to the precipitation forecast with the intensity agreeing better with the radar  
315 observations. It is also shown in Figs. 5a and 5b the standard deviation of the mean over the 6  
316 cases for the 3 experiments and it shows that the variance of LRMSE among the cases is smaller  
317 in the experiment LSAC<sub>On</sub> compared to the experiments CTR and LSAC<sub>Off</sub>. This result suggests  
318 that in spite of the different behavior among the cases, which is expected for the different types  
319 of precipitation, the spread of the errors in the amount of precipitation is smaller in the LSAC<sub>On</sub>  
320 experiment. In other words, the positive impact of the radar data assimilation seems to be more  
321 homogeneous among all cases when compared to the experiment without the LSAC. For the FSS

322 the variance from the LSAC<sub>on</sub> is not always smaller compared to the CTR and LSAC<sub>off</sub>  
323 experiments, which means that the error in the position of the precipitation is not consistently  
324 improved.

325 The averaged FAR and (POD) over all cases are shown in Fig. 6 and it is observed that  
326 the inclusion of radar data (LSAC<sub>off</sub>) increases the POD. However, the FAR also increases  
327 compared with the control run. On the other hand, when the LSAC is turned on, not only is the  
328 POD slightly better than in the case of LSAC<sub>off</sub>, but also the FAR decreased. Therefore, this  
329 result corroborates the idea that using LSAC suppresses spurious convection. As shown in Fig. 5,  
330 the standard deviation of the mean was included in Fig. 6 and the result is quite similar, the  
331 standard deviation is smaller in almost all forecast time in the LSAC<sub>on</sub> experiment.

332 A detailed analysis is made for the case of January 22<sup>nd</sup>, 2012 to examine the physical  
333 reasons for the improvement. This event caused a lot of damage in many cities in southeastern  
334 Brazil with large amounts of precipitation and hail in some locations, including the SPMR and  
335 some cities in the Paraiba Valley. The system developed from scattered small convective cells  
336 due to the diurnal surface heating in a favorable large-scale environment and then those scattered  
337 storms merged together to become a well-organized convective band covering almost the entire  
338 Paraiba Valley in the São Paulo state.

339 The first question we investigate is whether the addition of the LSAC improves the  
340 balance in the analysis. The domain-averaged absolute surface pressure tendency  $N$  can be used  
341 to measure the initial imbalance characteristics in the forecast initialized by the WRF 3D-Var  
342 analysis (Stauffer and Seaman 1990; Lynch and Huang 1992; Chen and Huang 2006; Hsiao et al.  
343 2012),

344

$$N = \frac{1}{I \cdot J} \cdot \sum_{i=1}^I \sum_{j=1}^J \left| \frac{\partial p_s}{\partial t} \right|_{ij} \quad (7)$$

345 where  $p_s$  is the surface pressure. The summation denotes the calculation over the whole model  
 346 domain.

347 Figure 7 shows the domain-averaged absolute surface pressure adjustment in 30 and 60  
 348 minutes over the forecast interval. The results show that the LSAC<sub>Off</sub> experiment experiences  
 349 larger adjustment than the LSAC<sub>On</sub> experiment. In LSAC<sub>Off</sub> (LSAC<sub>On</sub>),  $N$  has an initial value of  
 350 38.4 (27.2) Pa h<sup>-1</sup> and reaches an asymptotic value around 15.5 (13.2) Pa h<sup>-1</sup> for the 60 min  
 351 tendency. This result shows that the LSAC can help produce an analysis containing less noise,  
 352 which is very helpful when the radar observations are assimilated.

353 The precipitation skill scores (FSS and LRMSE) are shown for the case of January 22<sup>nd</sup>,  
 354 2012 in Fig. 8. The behavior of these skill scores is similar to what is observed in Fig. 5a,b  
 355 except for the better performance of the CTR between the 1- and 2-h forecast for the 1mm  
 356 threshold when its FSS is very close to that of LSAC<sub>Off</sub>, but still worse than that of LSAC<sub>On</sub>. This  
 357 similar behavior is important for the following detailed analysis because it suggests that this  
 358 specific convective case can be representative of the overall results.

359 The hourly accumulated precipitation fields at  $t = 1h$  for LSAC<sub>Off</sub> and LSAC<sub>On</sub> are shown  
 360 in Fig. 9 and compared against the radar derived precipitation. It is evident that the precipitation  
 361 is over-estimated in the experiment LSAC<sub>Off</sub>. The left red circle in Fig. 9 points out a specific  
 362 area where much more precipitation is predicted when the constraint is turned off, while inside  
 363 the right red circle the opposite is observed, i.e., the LSAC<sub>Off</sub> underestimates the precipitation  
 364 amount and the LSAC<sub>On</sub> produces a better forecast for this region. Furthermore, the over-

365 prediction is not only in the amount of precipitation, but also in the spatial distribution. A closer  
366 examination of Fig. 9 reveals how the large-scale analysis suppresses some spurious convection  
367 resulting in improved precipitation forecast skill. For example, the over-estimated precipitation  
368 over the Minas Gerais state (Fig. 9, 22.50S and 45.25W) in LSAC<sub>Off</sub> is suppressed in LSAC<sub>On</sub>.  
369 Note that despite of the small differences in the skill scores presented in Fig. 5, the forecasted  
370 precipitation is highly sensitive to these errors and do impact the precipitation prediction. This  
371 result is in part due to the adjustment of the water vapor field when using the LSAC. The  
372 improvements in the dynamical and microphysical fields like wind, temperature and water vapor  
373 are detailed in the following discussion.

374         The analysis increments for wind components, temperature and water vapor mixing ratio  
375 are shown in Fig. 10. By comparing the increments between LSAC<sub>On</sub> and LSAC<sub>Off</sub>, we note the  
376 following important differences. First, for the wind components, the increments from the  
377 experiment LSAC<sub>Off</sub> are concentrated where the radar data are available (see Fig. 1 for the radar  
378 coverage), whereas in the experiment LSAC<sub>On</sub> the increments are distributed over the entire  
379 domain with larger-scale disturbances outside the radar data region. Second, the temperature and  
380 water vapor increments from the experiment LSAC<sub>Off</sub> are considerably smaller than those from  
381 LSAC<sub>On</sub>. Third, the magnitudes of the wind increments and the patterns in the region with radar  
382 data are similar between the two experiments, suggesting that the analysis with the LSAC is still  
383 able to achieve a close fit to the radar observations. This is confirmed by a comparison of the  
384 statistics of O-A (radar observation minus final analysis) between the two experiments, as shown  
385 in Fig. 11. However, it is expected that the LSAC not only maintains a close fit to the radar  
386 observations, but also improves the wind analysis in the neighboring region surrounding the  
387 radar data, which we will examine later in this section.

388           The 10-m surface analyses of the horizontal wind vectors and speed at 18 UTC are  
389 compared in Fig. 12 along with the Climate Forecast System Reanalysis 2 from NCEP (Saha et  
390 al. 2010). The analysis fields from both LSAC<sub>off</sub> and LSAC<sub>on</sub> have larger wind speeds over the  
391 land than the reanalysis because of the higher resolution and convection. However, large changes  
392 of wind speed and direction from LSAC<sub>off</sub> are identified in the northwest quarter of the domain,  
393 corresponding to the spurious convection in that region (see Fig. 9a). In contrast, the result from  
394 LSAC<sub>on</sub> shows the higher wind speed located mainly along the coastal region where the  
395 convection occurs (see Fig. 2 experiment 4) and the wind field is generally smoother and more  
396 comparable to the reanalysis. This pattern is much more coherent than the one shown in the  
397 LSAC<sub>off</sub> experiment. The winds in LSAC<sub>off</sub> clearly show downburst-like outflows, which is  
398 probably due to downdraft caused by raining out of unsustainable rainwater added by the data  
399 assimilation.

400           To verify the accuracy of the analyses with and without the LSAC, we resort to the  
401 independent surface dataset collected by the Brazilian National Institute of Meteorology  
402 (INMET). This surface network covers the entire country and some of the stations are located  
403 within the studied area, as shown in Fig. 1. In Fig. 13, the objective comparisons are made for  
404 surface wind, water vapor, and temperature at each station for the 16 stations in the domain. The  
405 wind direction and speed are much more coherent in the LSAC<sub>on</sub> than in the LSAC<sub>off</sub>  
406 experiments when compared with the observations. The same positive impact from the LSAC is  
407 observed on temperature and water vapor mixing ratio. The RMS error differences (RMSE) for  
408 the three fields are computed and their results are shown in the table on the lower left corner in  
409 Fig. 13. The decrease in the RMSE due to the LSAC is 62%, 42% and 25% for wind speed,

410 water vapor and temperature, respectively. These verification results clearly show that the LSAC  
411 yields improved analyses of these fields.

412 It is worth noting that the LSAC not only contributes to the improvement of the analysis in  
413 the region with radar data but also where no radar data are available. For example, the wind at  
414 the Sao Paulo surface weather station, which is within the radar coverage, is significantly  
415 reduced from 9.7 to 2.3 m s<sup>-1</sup> where the observed wind is from northeast at 3.4 m s<sup>-1</sup> (see Fig.  
416 13). At another station, Varginha, where no radar observations are available, the wind speed is  
417 reduced from 3.9 to 1.5 m s<sup>-1</sup>, which is in better agreement with the observed value of 1.9 m s<sup>-1</sup>.  
418 The use of the LSAC is clearly beneficial in constraining the analysis to prevent unrealistic large  
419 wind increments which is a common problem in radar data assimilation.

420 Since the water vapor plays an important role in convective initiation, next we provide a  
421 closer evaluation on how the water vapor field is improved by the LSAC. Figure 14 shows a  
422 comparison between the experiments LSAC<sub>Off</sub> and LSAC<sub>On</sub>, along with the GFS analysis and a  
423 surface analysis using all the 16 stations of INMET. The GFS analysis agrees relatively well with  
424 the surface analysis in the large-scale sense; the wetter west and the drier north/northeast  
425 distribution in the domain are well captured by the GFS analysis. The difference between the  
426 LSAC<sub>On</sub> and LSAC<sub>Off</sub> water vapor fields shows that considerable improvement is made by the  
427 large-scale analysis constraint. The west part from LSAC<sub>On</sub> is still relatively drier than the  
428 observations, even though the difference between LSAC<sub>On</sub> and LSAC<sub>Off</sub> is positive in that area,  
429 i.e., the LSAC<sub>Off</sub> is even drier. In Fig. 9 it is shown that the precipitation field over Minas Gerais  
430 state is improved by reducing the spurious precipitation and Fig. 14 shows that the water vapor is  
431 over-estimated (negative values in the difference field) in that state by LSAC<sub>Off</sub>, but corrected in

432 LSAC<sub>on</sub>. Another similar example is the overprediction of precipitation south of Sao Paulo (Fig.  
433 9, 23.80S and 46.65W) in the LSAC<sub>off</sub> experiment caused by the large amount of water vapor.

434 From the above evaluation, we found that the 3D-Var with the LSAC not only allows a  
435 close fit to observations within the convective region but also is capable of correcting the errors  
436 in the storm environment. It is known that due to the small-scale model error and the limited-area  
437 domain used in the regional model, the forecast error grows as the model advances in time (Xu  
438 and Zhong, 2009). Therefore, after a few cycles the background, that is actually the previous  
439 forecast, can be biased due to the rapid small-scale error growth. The LSAC can reduce the bias  
440 since the large-scale analysis (i.e., GFS analysis in the current study) can better represent the  
441 large-scale mean, where filtering out the small-scale disturbances resulting from the process of  
442 geostrophic adjustment has been one of the objectives in large-scale analysis. The fact that the  
443 LSAC can improve the dynamical and microphysical fields by using the information from the  
444 high-resolution radar data and from the large-scale analysis at the same time to achieve improved  
445 analyses, both in the convective region and in its environment, is the key for getting a better  
446 precipitation forecast.

447 One may wonder whether the large-scale increment in the region without radar can be  
448 achieved by using a larger length scale of the background error covariance. To answer that  
449 question, an extra experiment was performed in which the background error without any tuning  
450 was used in the LSAC<sub>off</sub> experiment. The increments from this experiment are shown in Fig. 15.  
451 The results reveal that when using a larger length scale the increments have a wider spread than  
452 those in Figs. 10a and 10b, just as expected. However, the small-scale increments where the  
453 convection occurs are smoothed (resulting in a poorer fit to observations), and the large  
454 increments in the nearby environment are questionable. Figure 16 compares the fit to radial

455 velocity and rainwater mixing ratio (converted from the reflectivity) between the two  
456 experiments with (LSAC<sub>off</sub>) and without tuning of the background error statistics. It shows that  
457 the fit to observations is improved when the length scale is tuned in LSAC<sub>off</sub>. In the case without  
458 tuning, the O-A clusters in the smaller class intervals of [-1, 1] m s<sup>-1</sup> for radial velocity and [-  
459 0.05, 0.05] g kg<sup>-1</sup> for rainwater is smaller and the mean absolute error (MAE) is greater for both  
460 radial velocity and rainwater mixing ratio. This result suggests that limiting the influence of the  
461 radar data to the radar coverage through the tuning of the length scale is a better choice and  
462 better results can be achieved. For those areas where radar information is not available, the  
463 LSAC can help to constrain and improve the final analysis.

## 464 **5. Conclusions**

465 This study presents a new methodology for constraining the cost function while  
466 assimilating the radar data. The implementation was evaluated by performing data assimilation  
467 and short-range forecasts of six cases of summer convection in Brazil. We first evaluated the  
468 precipitation forecast skill using four different indices over forecasts of six convective cases and  
469 then a detailed analysis of one case was conducted in order to provide a deeper understanding on  
470 why this methodology improved the QPF.

471 The results demonstrated that the LSAC introduced in the WRF 3D-Var improved the  
472 QPF by producing improved analyses of wind, humidity and temperature when verified by  
473 surface observations. These fields were not only improved in the convective region where radar  
474 observations were available but also in the surrounding region by reducing errors caused by  
475 spurious convection. The 3D-Var with the LSAC filtered out unreliable high increments of wind,  
476 producing a wind field much more reasonable and coherent with the observed convective system  
477 without losing the beneficial information from radar. The comparison of the analyzed water

478 vapor mixing ratio from both experiments against surface observations also showed that the  
479 LSAC was able to produce an analysis much closer to observations and this improvement was  
480 crucial for supporting convection in the right location and eliminating spurious convective  
481 activities. It is important to note that the improvement achieved in the precipitation prediction  
482 due to the inclusion of the LSAC was obtained despite the small differences in the averaged skill  
483 scores between LSAC<sub>on</sub> and LSAC<sub>off</sub> experiments, showing that the precipitation forecast is  
484 highly sensitive to these errors. Moreover, it was shown through the domain averaged pressure  
485 tendency that the noise in the initial analysis could be reduced using the constraint. The reason  
486 for this reduction is that the large-scale analysis is better balanced and it filters out the small-  
487 scale disturbances.

488         We also showed that it is not beneficial to spread too much of the radar information  
489 through a larger length scale to where the radar data are not available. The main reason for this is  
490 because convection is a small-scale phenomenon, and spreading the increments too far can  
491 contribute to spurious activity. The comparison made in this study between the analyses with and  
492 without tuning the length scale showed that a closer fit to observations was achieved when the  
493 length scale was reduced to better represent the radar observations. However, the use of the  
494 smaller length-scale resulted in very small increments outside the radar coverage region. By  
495 using the LSAC in the 3D-Var, the analyses in this region were substantially improved by adding  
496 information from the large-scale analysis, resulting in improved performance of the QPF.

497         This study has shown the importance of LSAC for radar data assimilation in a 3D-Var  
498 data assimilation system using a few convective cases over a specific region. Nevertheless, the  
499 technique is applicable in general for regional data assimilation involving high-resolution data  
500 using any 3D-Var data assimilation systems. For example, when assimilating large amount of

501 high resolution data in an operational cycle, the cycle cannot last too long since the errors grow  
502 quickly and the large scale pattern can be lost, being necessary to restart the cycle every few  
503 hours to alleviate the problem. In this case, the LSAC could be used to constrain the data  
504 assimilation process such that the large scale pattern can be continuously kept, without the need  
505 of restarting the process. Besides that, the method is relatively easy to implement and with only  
506 slight increase in computation cost, making it a practical approach to improve the performance of  
507 high-resolution regional data assimilation.

508         Although in the current study we employed the GFS analysis in the LSAC, other large-  
509 scale analyses can be used in place of the GFS analysis. In the future, we will explore the  
510 feasibility to use the analysis from a coarser resolution outer domain of WRF. The advantage of  
511 using the analysis from the outer domain of the same model is the easy accessibility of the data  
512 in operational applications.

513 *Acknowledgments.*

514         The authors thank the Brazil's National Council for the Improvement of Higher  
515 Education (CAPES) and the National Center for Atmospheric Research (NCAR) sponsored by  
516 the National Science Foundation for their support and the CHUVA Project (Foundation for  
517 Research Support of Sao Paulo - FAPES grant 2009/15235-8) for providing the radar data.

518  
519  
520  
521  
522

523 **References**

524 Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A Multicase Comparative Assessment of the  
525 Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses.  
526 *Mon. Wea. Rev.*, **137**, 1805–1824.

527 Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A multi-case comparative assessment of the  
528 ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble  
529 forecasts. *Mon. Wea. Rev.*, **138**, 1273-1292.

530 Barker, D. M., W. Huang, Y. R. GUO, and Q. XIAO, 2004: A three-dimensional variational  
531 (3dvar) data assimilation system for use with mm5: Implementation and initial results. *Mon.*  
532 *Wea. Rev.*, **132**, 897–914.

533 Bloom, S. C, L. L Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using  
534 Incremental Analysis Updates. *Mon. Wea. Rev.*, 124, 1256–1271.

535 Bouttier, F., 2009: Fine scale versus large scale data assimilation - a discussion. In: CAWCR  
536 Workshop on Ensemble Prediction and Data Assimilation, Melbourne, Australia.

537 Bryan, G. H. and H. Morrison, 2012: Sensitivity of a simulated squall line to horizontal  
538 resolution and parameterization of microphysics. *Mon. Wea. Rev.*, **140**, 202-225.

539 Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4D-Var and the ensemble  
540 Kalman filter techniques for radar data assimilation. *Mon. Weather Rev.*, **133**, 3081–3094.

541 Courtier, P., J. N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational  
542 implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**,  
543 1367–1387.

544 Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nature Climate Change*, **2**,  
545 491-496.

546 Chen, M., and X.-Y. Huang, 2006: Digital Filter Initialization for MM5. *Mon. Wea. Rev.*, **134**,  
547 1222-1236.

548 Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon  
549 experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, **46**, 3077–3107.

550 Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using  
551 Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

552 Gao, J., M. Xue, A. Shapiro, and K. K. Droegemeier, 1999: A variational analysis for the  
553 retrieval of three-dimensional mesoscale wind fields from two Doppler radars, *Mon. Wea. Rev.*,  
554 **127**, 2128-2142.

555 Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2007: A three-dimensional variational  
556 data analysis method with recursive filter for doppler radars. *J. Atmos. Oceanic Technol.*, **21**,  
557 457–469.

558 Guo, Y. R., S. Y. Fan, W. Wang, M. Chen, X. Y. Huang, Y. C. Wang, and Y. H. Kuo, 2007:  
559 Application of WRFVAR (3dvar) to a high resolution (3-km) model over Beijing area. In: 8th  
560 WRF Users' Workshop.

561 Ha, J.-H. and D.-K. Lee, 2012: Effect of length scale tuning of background Error in WRF-  
562 3DVAR system on assimilation of high-resolution surface data for heavy rainfall simulation.  
563 *Adv. Atmos. Sci.*, **29-6**, 1142-1158.

564 Hsiao, L.-F., S.-S. Chen, Y.-H. Kuo, Y.-R. Guo, T.-C. Yeh, J.-S. Hong, C.-T. Fong, and C.-S.  
565 Lee, 2012: Application of WRF 3DVAR to Operational Typhoon Prediction in Taiwan: Impact  
566 of Outer Loop and Partial Cycling Approaches. *Wea. Forecasting*, **27**, 1249-1263.

567 Hong, S. Y., and J. O. J. Lim, 2006: The WRF Single-Moment 6-Class Microphysics Scheme  
568 (WSM6), *J. Korean Meteor. Soc.*, **42**, 129–151.

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

571 Kawabata, T., H. Seko, K. Saito, T. Kuroda, K. Tamiya, T. Tsuyuki, Y. Honda, and Y.  
572 Wakazuki, 2007: An assimilation and forecasting experiment of the Nerima heavy rainfall with a  
573 cloud-resolving nonhydrostatic 4-dimensional variational data assimilation system. *J. Meteor.*  
574 *Soc. Japan*, **85**, 255–276.

575 Lee, M.-S, Y.-H. Kuo, D. M. Barker, and E. Lim, 2006: Incremental Analysis Updates  
576 Initialization Technique Applied to 10-km MM5 and MM5 3DVAR. *Mon. Wea. Rev.*, **134**,  
577 1389–1404.

578 Lorenc, A., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor.*  
579 *Soc.*, **112**, 1177-1194.

580 Lorenc, A. 1988: A practical approximation to optimal four-dimensional data assimilation. *Mon.*  
581 *Wea. Rev.*, **116**, 730-745.

582 Lynch, P., 1993: Digital Filters for Numerical Weather Prediction. HIRLAM Technical Report  
583 No. 10, FMI, PB 503, SF-00101, Helsinki, Finland,  
584 [http://mathsci.ucd.ie/~plynch/Publications/HIRLAM\\_Tech\\_Report\\_10.pdf](http://mathsci.ucd.ie/~plynch/Publications/HIRLAM_Tech_Report_10.pdf).

585 Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter.  
586 *Mon. Wea. Rev.*, **120**, 1019–1034.

587 Machado, L. A., and Coauthors, 2014: The Chuva Project: How Does Convection Vary across  
588 Brazil? *Bull. Amer. Meteor. Soc.*, **95**, 1365–1380.

589 Maity, R., S. Dey, and P. Varun, 2015: Alternative Approach for Estimation of Precipitation  
590 Using Doppler Weather Radar Data. *J. Hydrol. Eng.*, doi: 10.1061/(ASCE)HE.1943-  
591 5584.0001146.

592 Marshall, J. S. and W. M. Palmer, 1948: The distribution of raindrops with size. *J. Meteorol.*, **5**,  
593 165-166.

594 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative  
595 transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-  
596 wave. *J. Geophys. Res.*, **102** (D14), 16663–16682.

597 Ming, C., S. Y. Fan, J. Zhong, X. Y. Huang, Y. R. Guo, W. Wang, Y. Wang, and B. A. Kuo,  
598 2009: A WRF-based rapid updating cycling forecast system of BMB and its performance during  
599 the summer and Olympic games 2008. In: World Meteorological Organization Symposium on  
600 Nowcasting and Very Short Term Forecasting.

601 Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-  
602 interpolation analysis system. *Mon. Wea. Rev.*, **120**, 747–1763.

603 Reen, B. P., 2007: Data Assimilation Strategies and Land-Surface Heterogeneity Effects in the  
604 Planet Boundary Layer. Ph.D. thesis, The Pennsylvania State University, 246 pp.

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

607 Saha, S. and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer.*  
608 *Meteor. Soc.*, **91**, 1015-1057, doi:10.1175/2010BAMS3001.1.

609 Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wae. Rev.*, **98**,  
610 875-883.

611 Shimizu, S., M. Maki, T. Maesake, K. IWANAMI, and S. Shimada, 2011: Short-range forecast  
612 using MPradar network and 3DVAR assimilation for the heavy rain- fall in North Tokyo on July  
613 5th 2010. In: Japan Geoscience Union Meeting 2011. Chiba, Japan.

614 Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang,  
615 W. Wang, and J. G. Powers, 2008: A description of the Advanced Research WRF version 3.  
616 NCAR Technical Note 475, NCAR/TN-4751STR.

617 Snyder, C., and F. Zhang, 2003: Assimilation of simulated doppler radar observations with an  
618 ensemble kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.

619 Stauffer, D. R., and Nelson L. Seaman, 1990: Use of Four-Dimensional Data Assimilation in a  
620 Limited-Area Mesoscale Model. Part I: Experiments with Synoptic-Scale Data. *Mon. Wea. Rev.*,  
621 **118**, 1250-1277.

622 Sugimoto, S., N. A. Crook, J. Sun, Q. Xiao, and D. M. Barker, 2009: An Examination of WRF  
623 3DVAR Radar Data Assimilation on Its Capability in Retrieving Unobserved Variables and  
624 Forecasting Precipitation through Observing System Simulation Experiments, *Mon. Wea. Rev.*,  
625 **137**, 4011-4029.

626 Sun, J.-H., X.-L. Zhang, J. Wei, and S.-X. Zhao, 2005: A study on severe heavy rainfall in north  
627 China during the 1990s (in Chinese). *Climatic Environ. Res.*, **10**, 492–506.

628 Sun, J., and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar  
629 observations using a cloud model and its adjoint: Part I. model development and simulated data  
630 experiments. *J. Atmos. Sci.*, **54**, 1642-1661.

631 Sun, J., S. B. Trier, Q. Xiao, M. L. Weisman, H. Wang, Z. Ying, M. Xu, and Y. Zhang, 2012:  
632 Sensitivity of 0–12-h Warm-Season Precipitation Forecasts over the Central United States to  
633 Model Initialization. *Wea. Forecasting*, **27**, 832–855.

634 Sun, J., and H. Wang, 2013: Radar data assimilation with WRF 4d-var: Part II. Comparison with  
635 3d-var for a squall line case. *Mon. Wea. Rev.*, **141**, 2245–2264.

636 Sun, J., W. Tong, H. Wang, Y. Zhang, C.-Y. Lin, and D. Xu, 2014: Comparison of the impacts  
637 of momentum control variables on limited-area high-resolution variational data assimilation and  
638 convective forecasting. Submitted to *Mon. Wea. Rev.*.

639 Tong, M., and M. Xue, 2005: Ensemble kalman filter assimilation of doppler radar data with a  
640 compressible nonhydrostatic model: Oss experiments. *Mon. Wea. Rev.*, **133**, 1789–1807.

641 Tong, M., and M. Xue, 2008: Simultaneous estimation of microphysical parameters and  
642 atmospheric state with simulated radar data and ensemble square-root Kalman filter. Part II:  
643 Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1649–1668.

644 Tong, W., J. Sun, G. Li, and H. Wang, 2014: A study on the assimilation cycling configuration  
645 for convective precipitation forecast using WRF 3DVAR. In: 15th WRF Users' Workshop -  
646 <http://www2.mmm.ucar.edu/wrf/users/workshops/WS2014/WorkshopPapers.php>.

647 Wang, H., J. Sun, S. Y. Fan, and X. Y. Huang, 2013a: Indirect assimilation of radar reflectivity  
648 with wrf 3d-var and its impact on prediction of four summertime convective events. *J. Appl.*  
649 *Meteor. Climatol.*, **52**, 889–902.

650 Wang H., J. Sun, X. Zhang, X. Huang, and T. Auligne, 2013b: Radar data assimilation with  
651 WRF 4D-Var: Part I. System development and preliminary testing. *Mon. Wea. Rev.*, **141**, 2224-  
652 2244.

653 Xiao, Q., Y. Kuo, J. Sun, W. Lee, D. M. Barker, and L. Eunha, 2007: An approach of radar  
654 reflectivity data assimilation and its assessment with the inland qpf of typhoon rusa (2002) at  
655 landfall. *J. Appl. Meteor. and Climatol.*, **46**, 14-22.

656 Xu, Jian-Yu and Q. Zhong, 2009: The Effect of Error Growth and Propagation on the  
657 Predictability of Quantitative Precipitation in a Cloud-Resolving Model. *Atmos. Oceanic Sci.*  
658 *Lett.*, **2**, 79-84.

659 Zou, X., Y.-H. Kuo, and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a  
660 nonhydrostatical adjoint model. *Mon. Wea. Rev.*, **123**, 2229-2249.

661 Zupanski, M, 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*,  
662 **133**, 1710–1726.

663

664

665

666

667

668 **List of Figures**

669 Figure 1: Topography, model domain and INMET (Brazil's National Institute of Meteorology)  
670 weather stations. The radar position and its maximum range is also shown. The Sao Paulo and  
671 Minas Gerais states, the Paraiba Valley and the São Paulo Metropolitan Region are labeled in  
672 blue.

673

674 Figure 2: 3-km radar reflectivity in dBZ for all 6 convective cases at 18 UTC.

675

676 Figure 3: Location of radar and positions of airep, metar, synop, ship and sound observations.

677

678 Figure 4: Illustration of the cycling strategy.  $GFS_i$  and  $WRF_i$  indicate initial conditions from GFS  
679 and WRF, respectively; DA means data assimilation and RA and GTS stand for data from radar  
680 and GTS, respectively.

681

682 Figure 5: Averaged FSS of the 6 cases for thresholds of (a)  $1\text{mm h}^{-1}$  and radius of 20 km and (b)  
683  $5\text{mm h}^{-1}$  and radius of 10 km. The LRMSEs are shown on the lower panels. The vertical bars are  
684 the standard deviation of the average.

685

686 Figure 6: Averaged False Alarm and Probability of Detection over the 6 cases. The vertical bars  
687 are the standard deviation of the average

688

689 Figure 7: Domain-averaged absolute surface pressure tendency  $N$ . The curves for  $LSAC_{\text{Off}}$   
690 (closed circle) and  $LSAC_{\text{On}}$  (open circle) are shown. The colors represent the different time

691 windows used to calculate the tendency, i.e., black for 1 hour [ $\text{Pa } 60\text{m}^{-1}$ ], blue for 30 minutes [ $\text{Pa}$   
692  $30\text{m}^{-1}$ ].

693

694 Figure 8: Similar to Fig. 5, but only for the case of January 22<sup>nd</sup>, 2012.

695

696 Figure 9: 1-h forecast of precipitation for the case of January 22<sup>nd</sup>, 2012 for (a) experiment  
697 LSAC<sub>Off</sub>, (b) experiment LSAC<sub>On</sub> and (c) that retrieved from radar reflectivity.

698

699 Figure 10: Analysis increments at the third model level (around 150m) for meridional (upper-  
700 left) and zonal (upper-right) wind components, temperature (low-left) and water vapor mixing  
701 ratio (low right) for (a) experiment LSAC<sub>Off</sub> and (b) experiment LSAC<sub>On</sub> at the analysis time 18  
702 UTC, January 22<sup>nd</sup>, 2012. Wind are in [ $\text{m s}^{-1}$ ], temperature in [K] and water vapor mixing ratio in  
703 [ $\text{g kg}^{-1}$ ].

704

705 Figure: 11 - Ratio between observation minus analysis (O-A) and observation minus background  
706 (O-B) for radial velocity for both experiments LSAC<sub>Off</sub> and LSAC<sub>On</sub> for the case of January 22<sup>nd</sup>,  
707 2012.

708

709 Figure 12: 10-m wind vectors and speed (shaded) in [ $\text{m s}^{-1}$ ] from LSAC<sub>Off</sub> (upper), LSAC<sub>On</sub>  
710 (middle) and NCEP/CSFR 2 reanalysis with  $0.5^\circ$  resolution (low), valid at 18 UTC, January 22<sup>nd</sup>,  
711 2012.

712

713 Figure 13: Evaluation of the wind, water mixing ratio and temperature of the LSAC<sub>Off</sub> and  
714 LSAC<sub>On</sub> experiments using the observations from stations shown in Fig. 1 for the case of January  
715 22<sup>nd</sup>, 2012 at 18 UTC.

716  
717 Figure 14: Water vapor mixing ratio near the surface for LSAC<sub>Off</sub> (upper-left), LSAC<sub>On</sub> (upper-  
718 right), GFS (lower-left) and observation (lower-right) in [ $\text{g kg}^{-1}$ ] all for the case of January 22<sup>nd</sup>,  
719 2012. The observation is an interpolation from surface weather stations (blue dots).

720  
721 Figure 15: Same as Fig. 10, but for the experiment without tuning the length scale.

722  
723 Figure 16: Difference between observation and analysis (O-A) of radial velocity (upper panel)  
724 and rainwater mixing ratio (lower panel) for the experiment LSAC<sub>Off</sub> (shaded) and the  
725 experiment without tuning the length scale (crosses). The Mean Absolute Error for each  
726 experiment is also shown.

727  
728  
729  
730  
731  
732  
733  
734  
735

736 Table 1: Contingency table.

		Observed	
		Yes	No
Forecast	Yes	hits	false alarms
	No	misses	correct negatives

737

738

739 Table 2: Convective cases and their characteristics.

Case	Date	Description
1	Nov 29 <sup>th</sup> , 2011	Local convection triggered by surface heating and large scale forcing
2	Dec 01 <sup>st</sup> , 2011	Local convection triggered by surface heating and large scale forcing
3	Jan 07 <sup>th</sup> , 2012	Local convection with weak forcing
4	Jan 22 <sup>nd</sup> , 2012	From scattered local convection to a well-organized convective system
5	Feb 11 <sup>th</sup> , 2012	Large convective system associated to a frontal system
6	Feb 12 <sup>nd</sup> , 2012	Large convective system associated to a frontal system

740

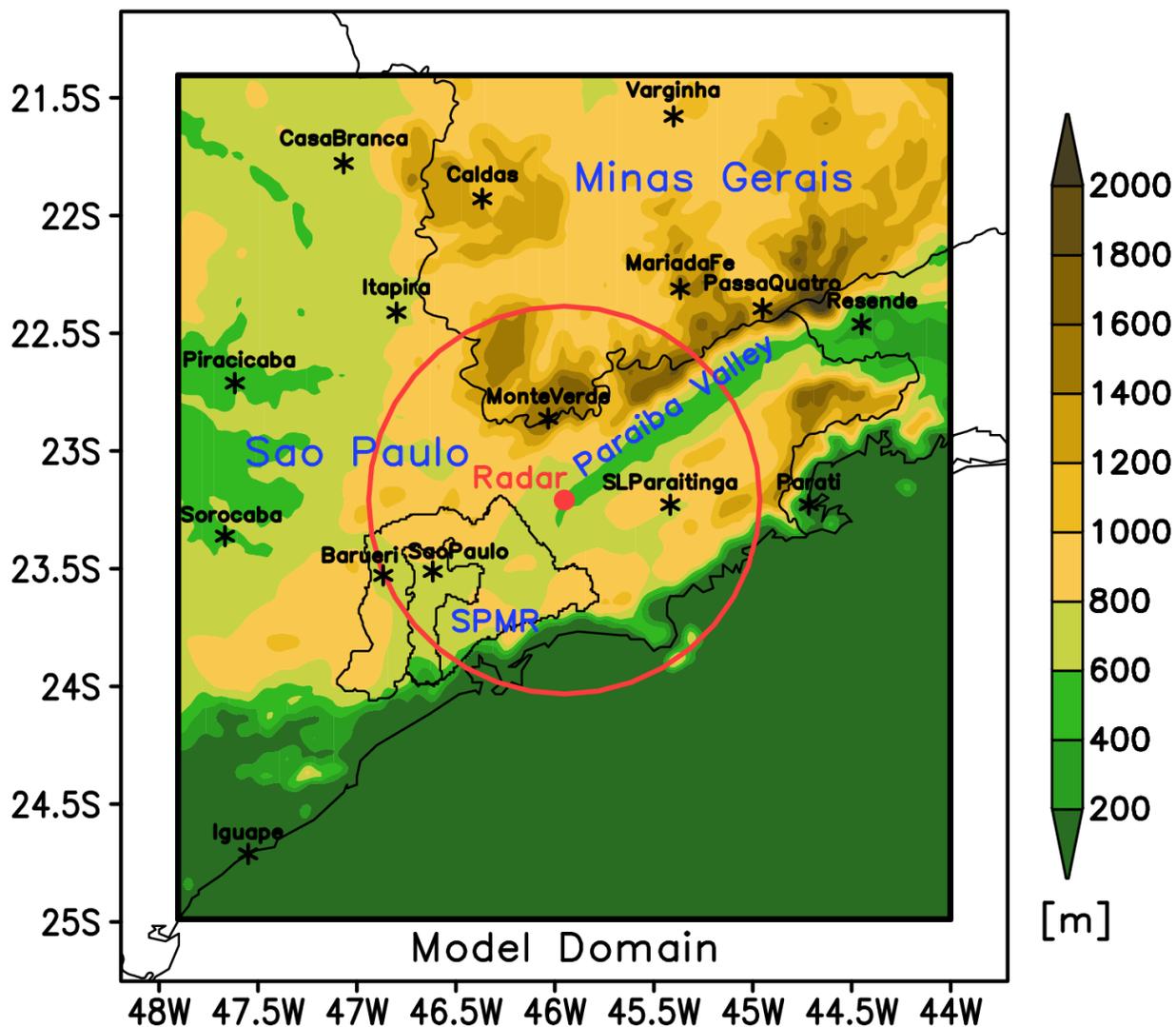
741

742 Table 3: Experiments descriptions.

Experiment	Assimilated Data	Constraint Term
CTR	Conventional (GTS)	Off
LSAC <sub>Off</sub>	Conventional + Radar (reflectivity and radial velocity)	Off
LSAC <sub>On</sub>	Conventional + Radar (reflectivity and radial velocity)	On

743

# Topography



744

745 Figure 1: Topography, model domain and INMET (Brazil's National Institute of Meteorology)  
746 weather stations. The radar position and its maximum range is also shown. The Sao Paulo and  
747 Minas Gerais states, the Paraiba Valley and the São Paulo Metropolitan Region are labeled in  
748 blue.

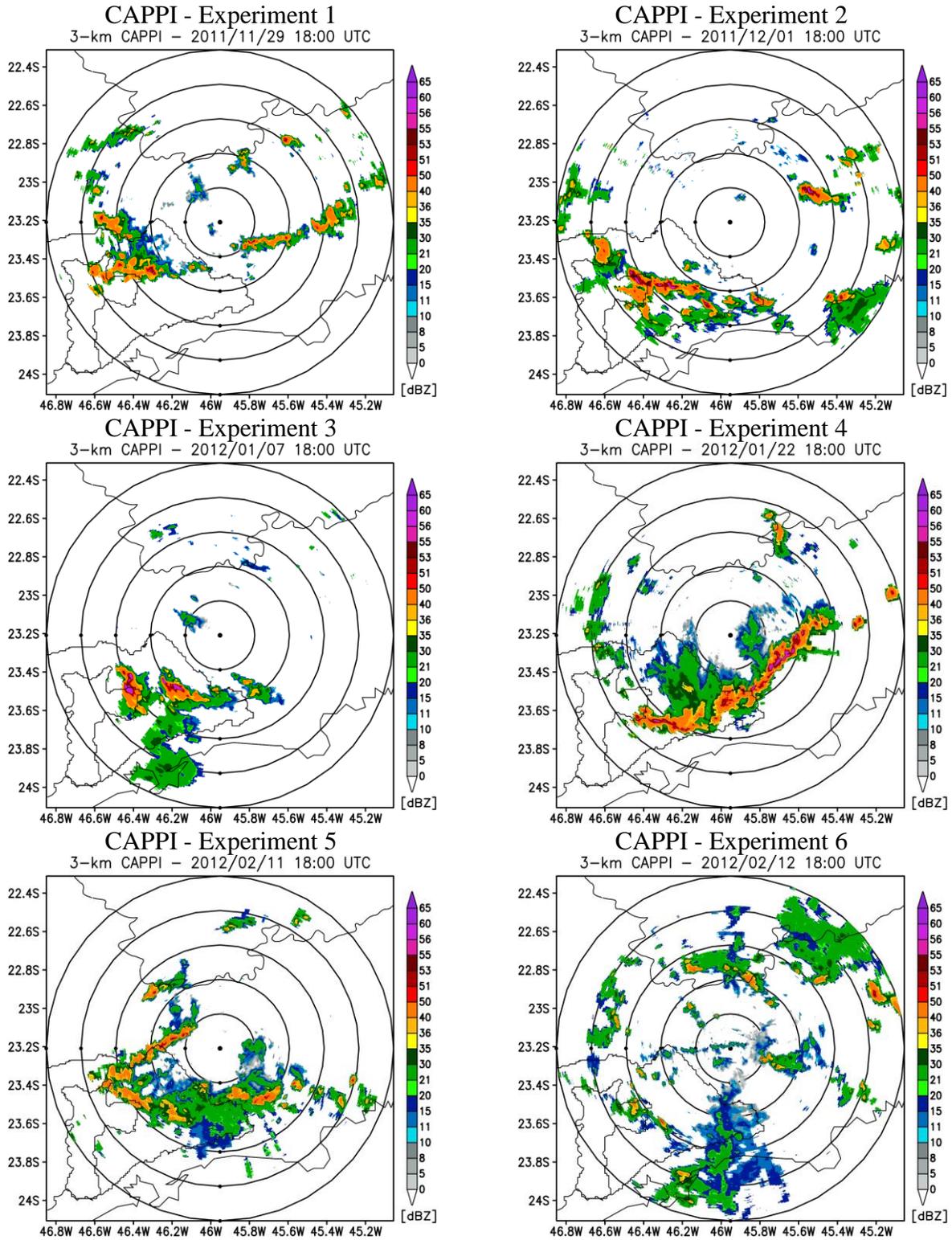
749

750

751

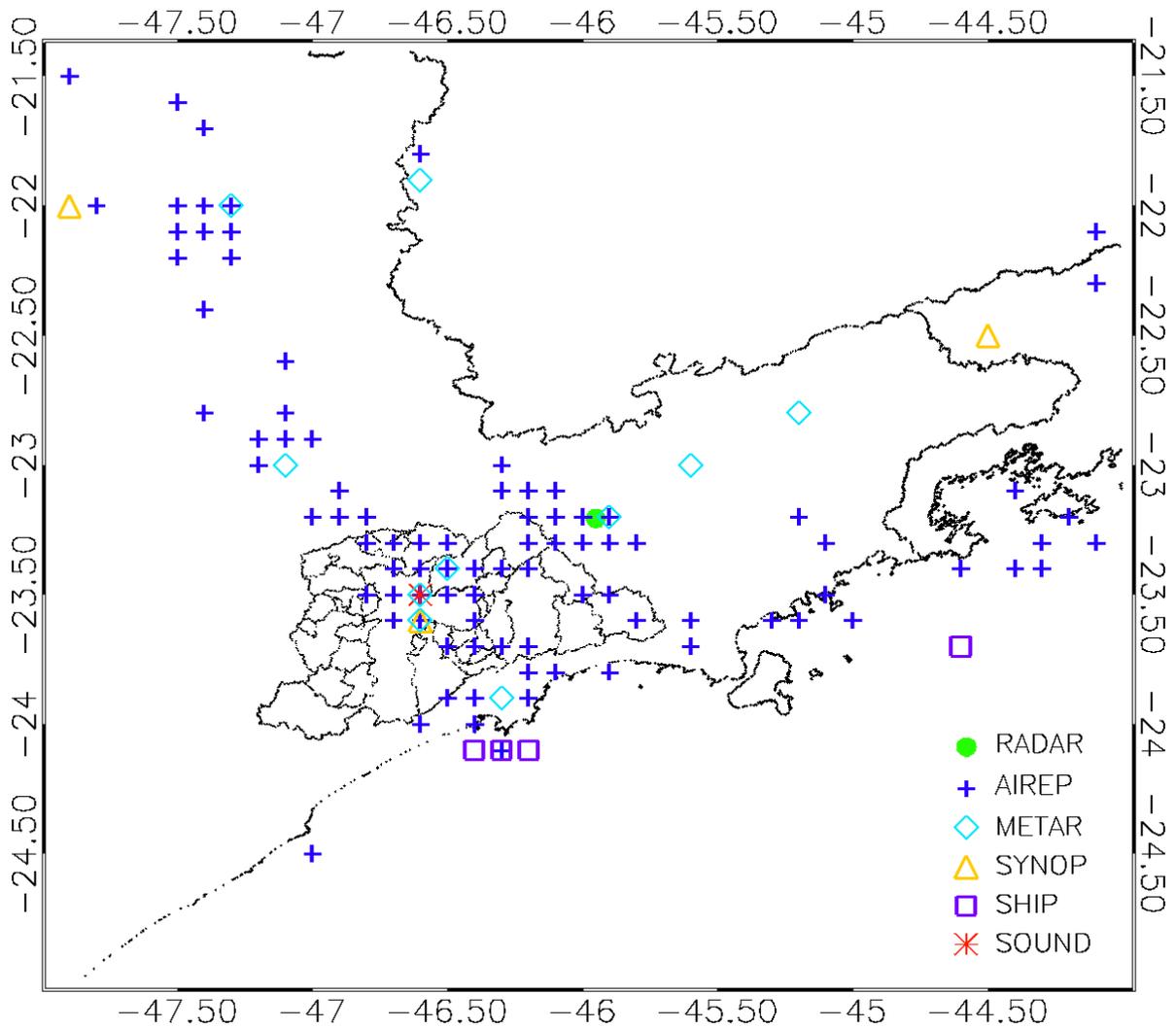
752

753



754 Figure 2: 3-km radar reflectivity in dBZ for all 6 convective cases at 18 UTC.

755



756

757 Figure 3: Location of radar and positions of airep, metar, synop, ship and sound observations.

758

759

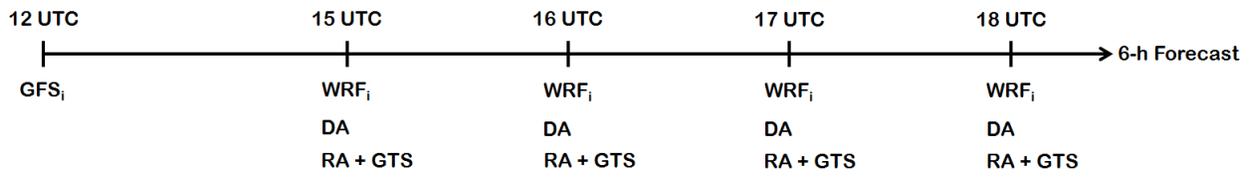
760

761

762

763

764



765

766 Figure 4: Illustration of the cycling strategy.  $GFS_i$  and  $WRF_i$  mean initial conditions from GFS

767 and WRF, respectively; DA means data assimilation and RA and GTS mean data from radar and

768 GTS, respectively.

769

770

771

772

773

774

775

776

777

778

779

780

781

782

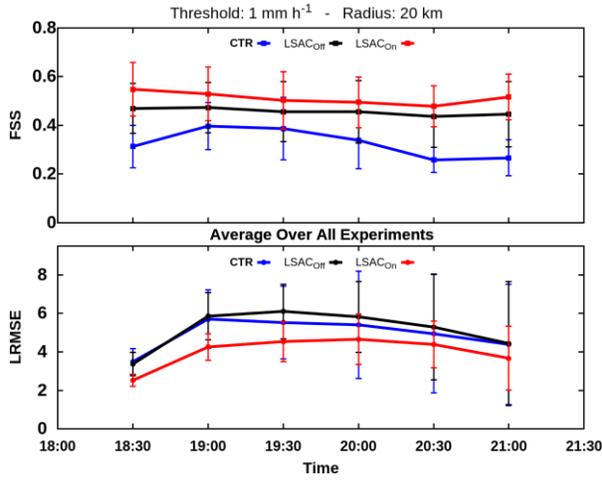
783

784

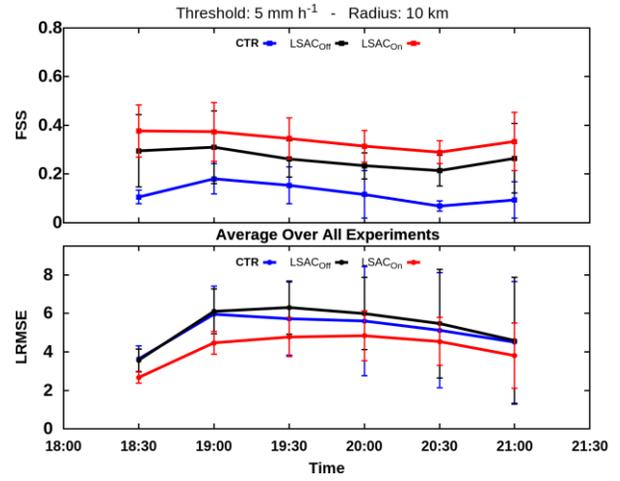
785

786

787



(a)



(b)

788 Figure 5: Averaged FSS of the 6 cases for thresholds of (a) 1mm h<sup>-1</sup> and radius of 20 km and  
 789 5mm h<sup>-1</sup> and radius of 10 km. The LRMSEs are shown on the lower panels. The vertical bars are  
 790 the standard deviation of the average.

791

792

793

794

795

796

797

798

799

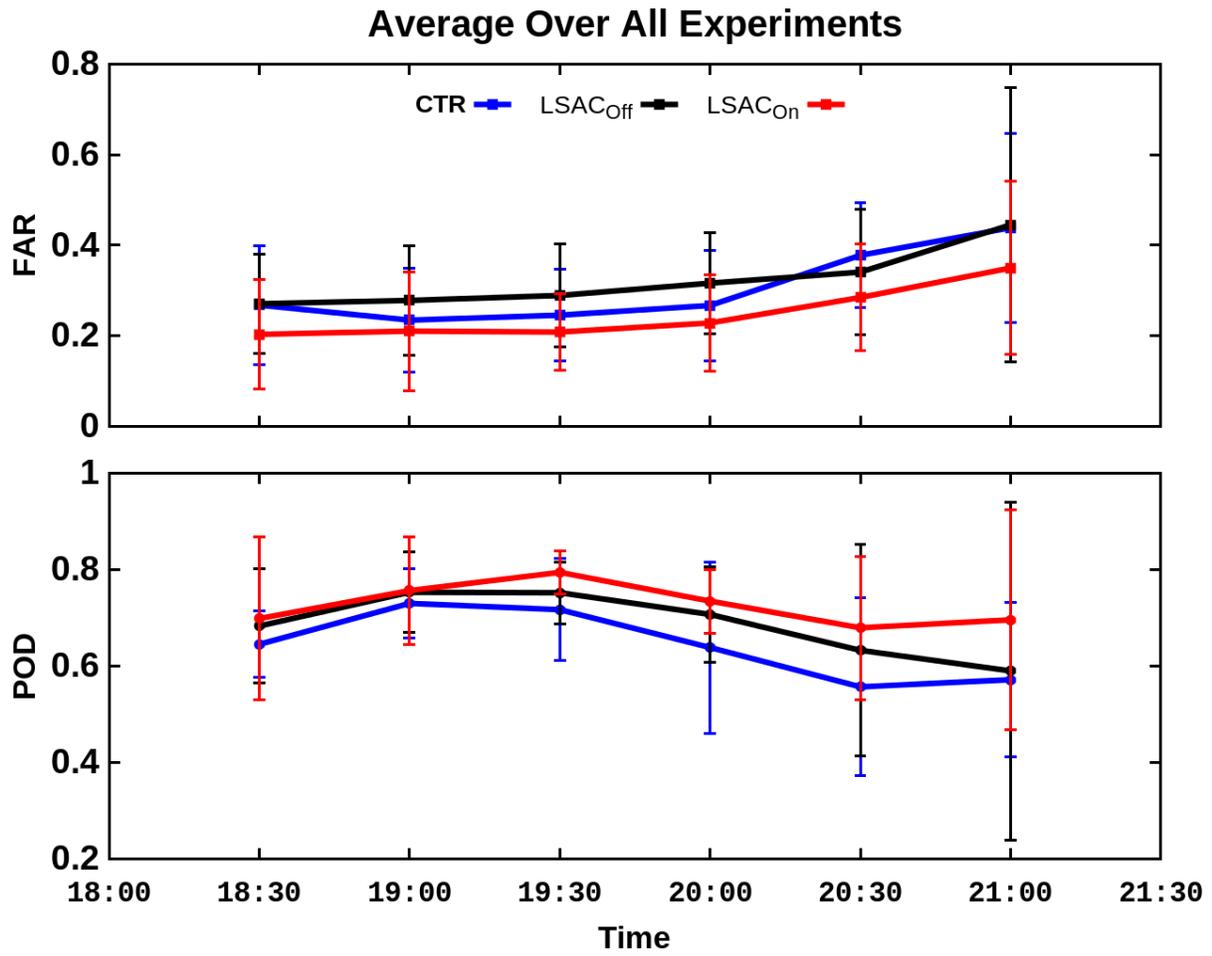
800

801

802

803

804



805

806 Figure 6: Averaged False Alarm and Probability of Detection over the 6 cases. The vertical bars

807 are the standard deviation of the average.

808

809

810

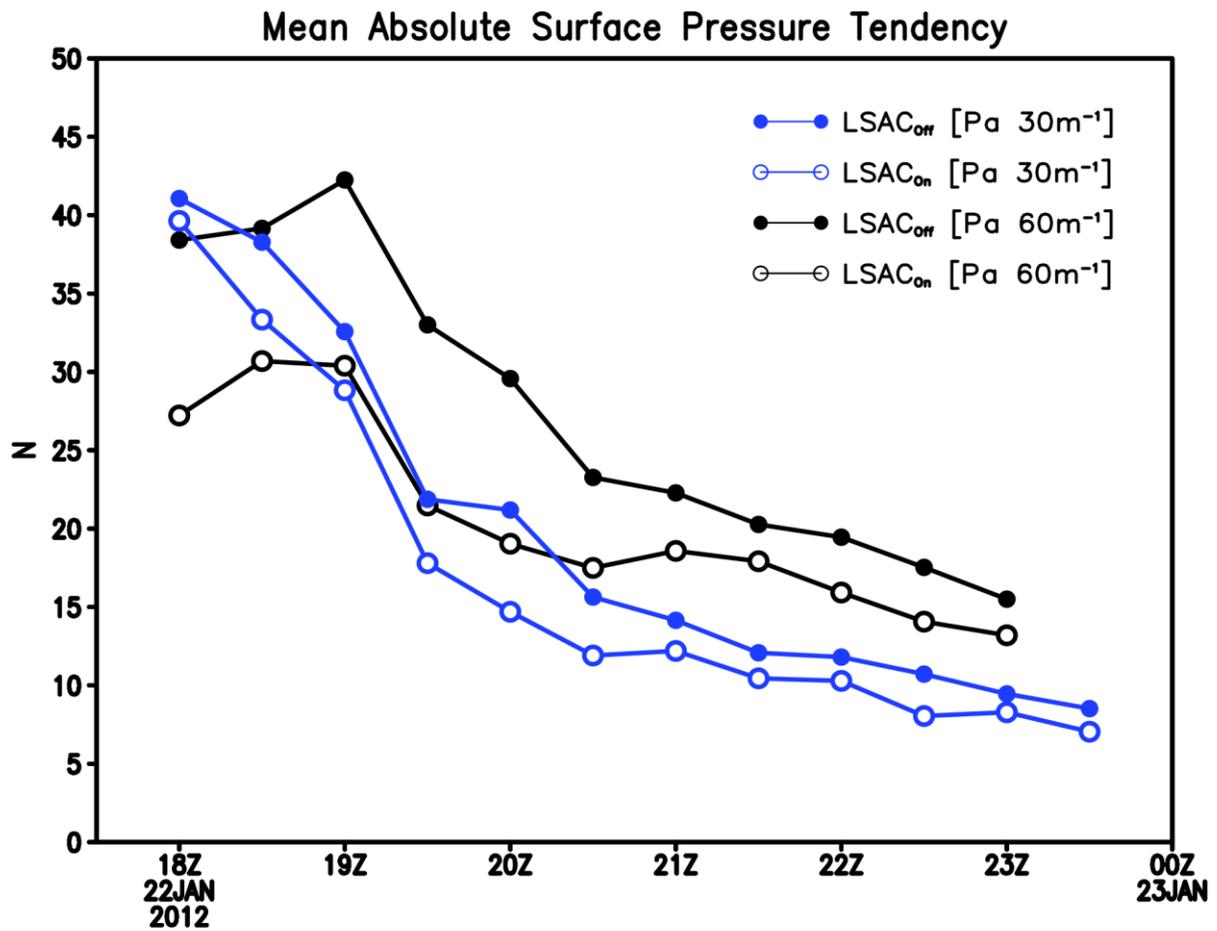
811

812

813

814

815  
816  
817  
818



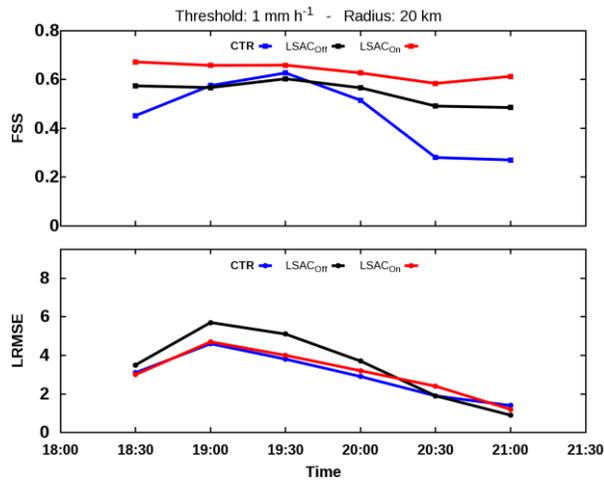
819  
820  
821  
822  
823  
824  
825

Figure 7: Domain-averaged absolute surface pressure tendency  $N$ . The curves for LSAC<sub>Off</sub> (closed circle) and LSAC<sub>On</sub> (open circle) are shown. The colors represent the different time windows used to calculate the tendency, i.e., black for 1 hour [Pa 60m<sup>-1</sup>], blue for 30 minutes [Pa 30m<sup>-1</sup>].

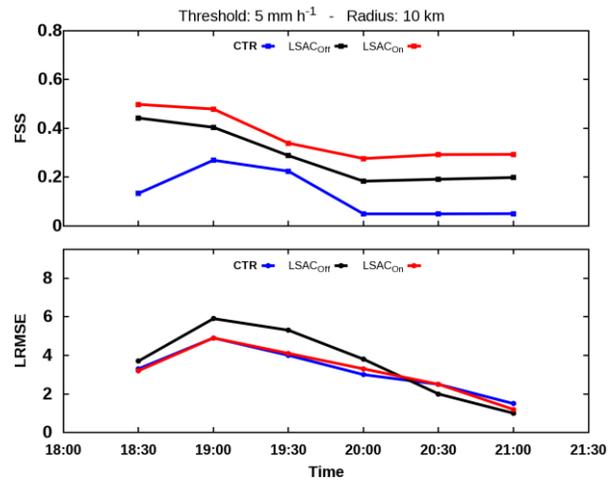
826

827

828



(a)



(b)

829 Figure 8: Similar to Fig. 5, but only for the case of January 22<sup>nd</sup>, 2012.

830

831

832

833

834

835

836

837

838

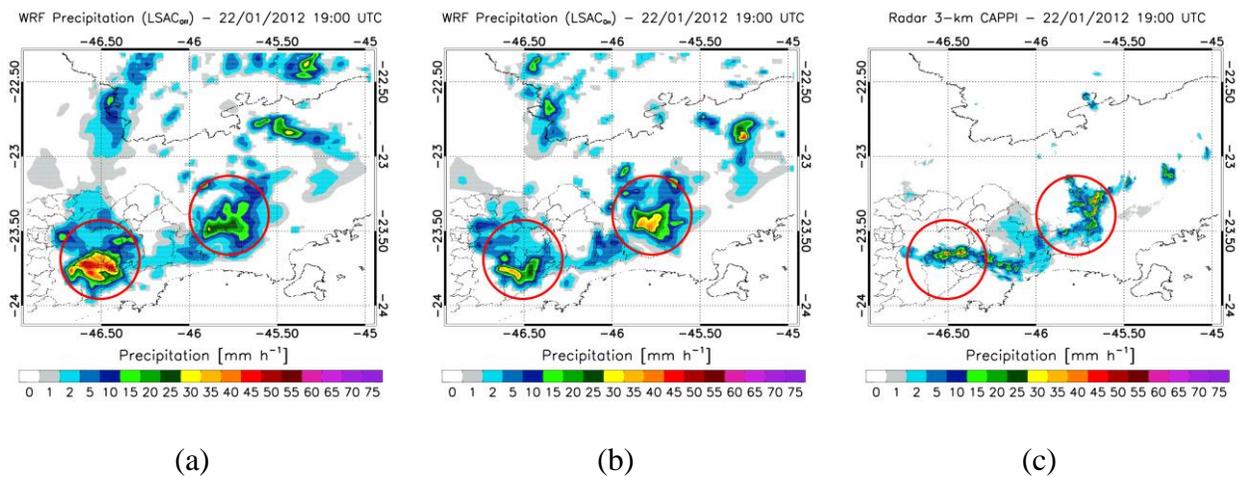
839

840

841

842

843



844 Figure 9: 1-h forecast of precipitation for the case of January 22<sup>nd</sup>, 2012 for (a) experiment

845 LSAC<sub>Off</sub>, (b) experiment LSAC<sub>On</sub> and (c) that retrieved from radar reflectivity.

846

847

848

849

850

851

852

853

854

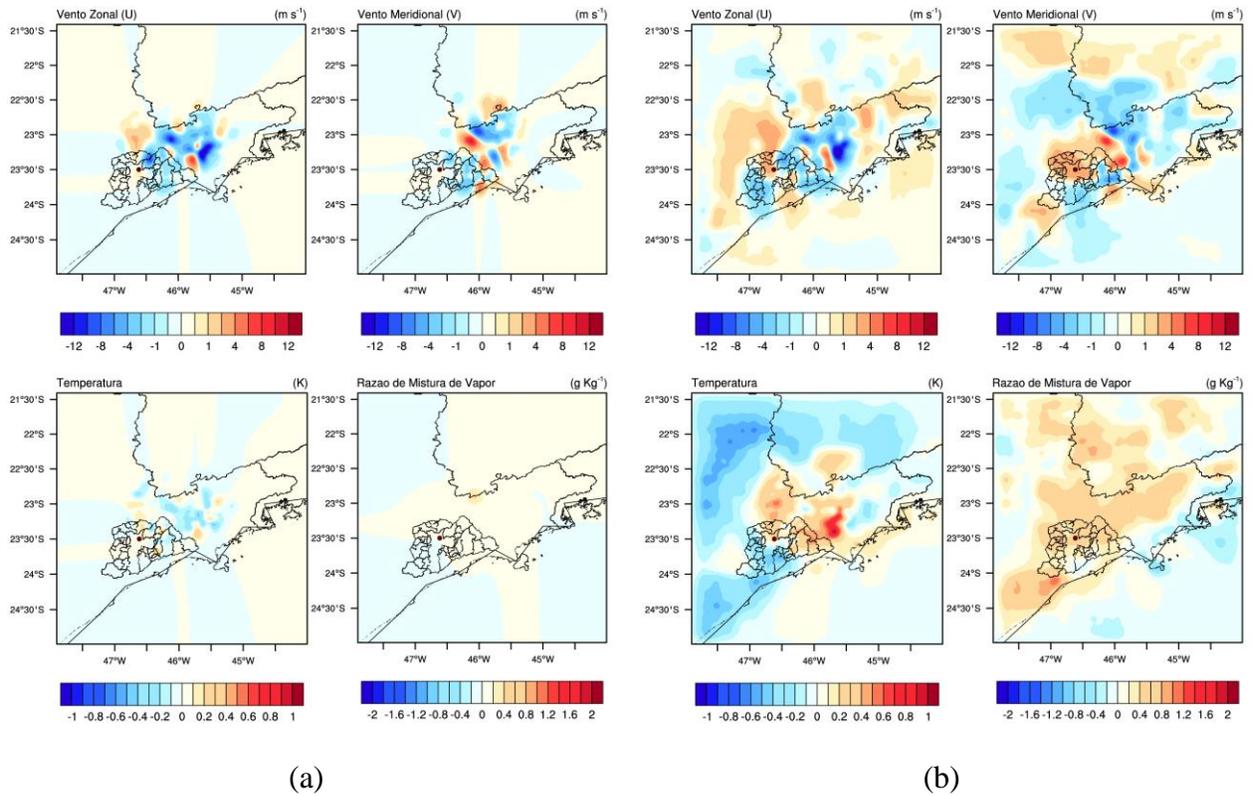
855

856

857

858

859



860 Figure 10: Analysis increments at the third model level (around 150m) for meridional (upper-  
861 left) and zonal (upper-right) wind components, temperature (low-left) and water vapor mixing  
862 ratio (low right) for (a) experiment LSAC<sub>Off</sub> and (b) experiment LSAC<sub>On</sub> at the analysis time 18  
863 UTC, January 22<sup>nd</sup>, 2012. Wind are in [m s<sup>-1</sup>], temperature in [K] and water vapor mixing ratio in  
864 [g kg<sup>-1</sup>].

865

866

867

868

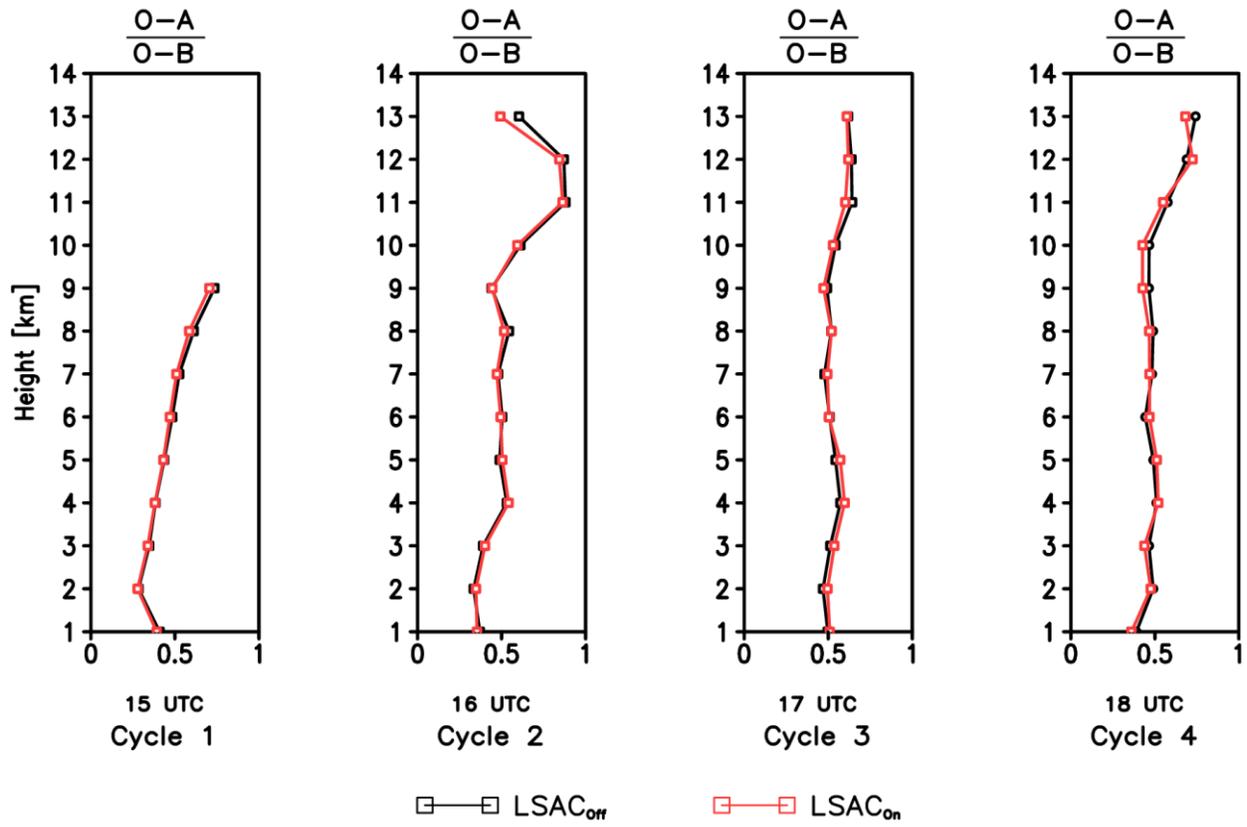
869

870

871

872

## Radial Velocity – Experiment 4



873

874 Figure: 11 - Ratio between observation minus analysis (O-A) and observation minus background

875 (O-B) for radial velocity for both experiments LSAC<sub>Off</sub> and LSAC<sub>On</sub> for the case of January 22<sup>nd</sup>,

876 2012.

877

878

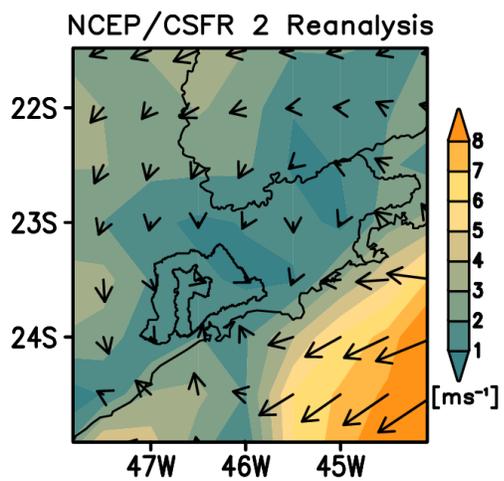
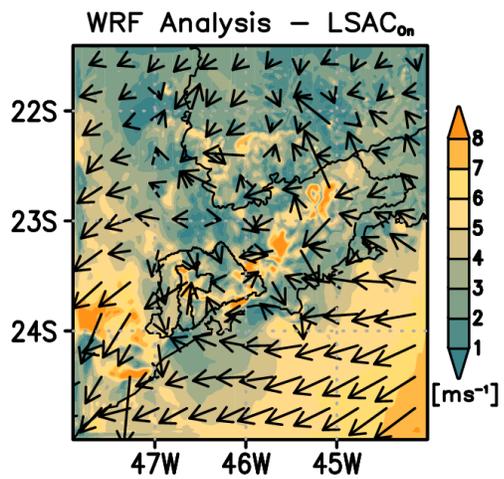
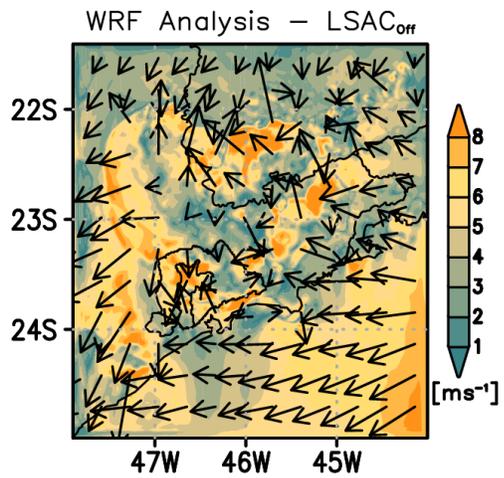
879

880

881

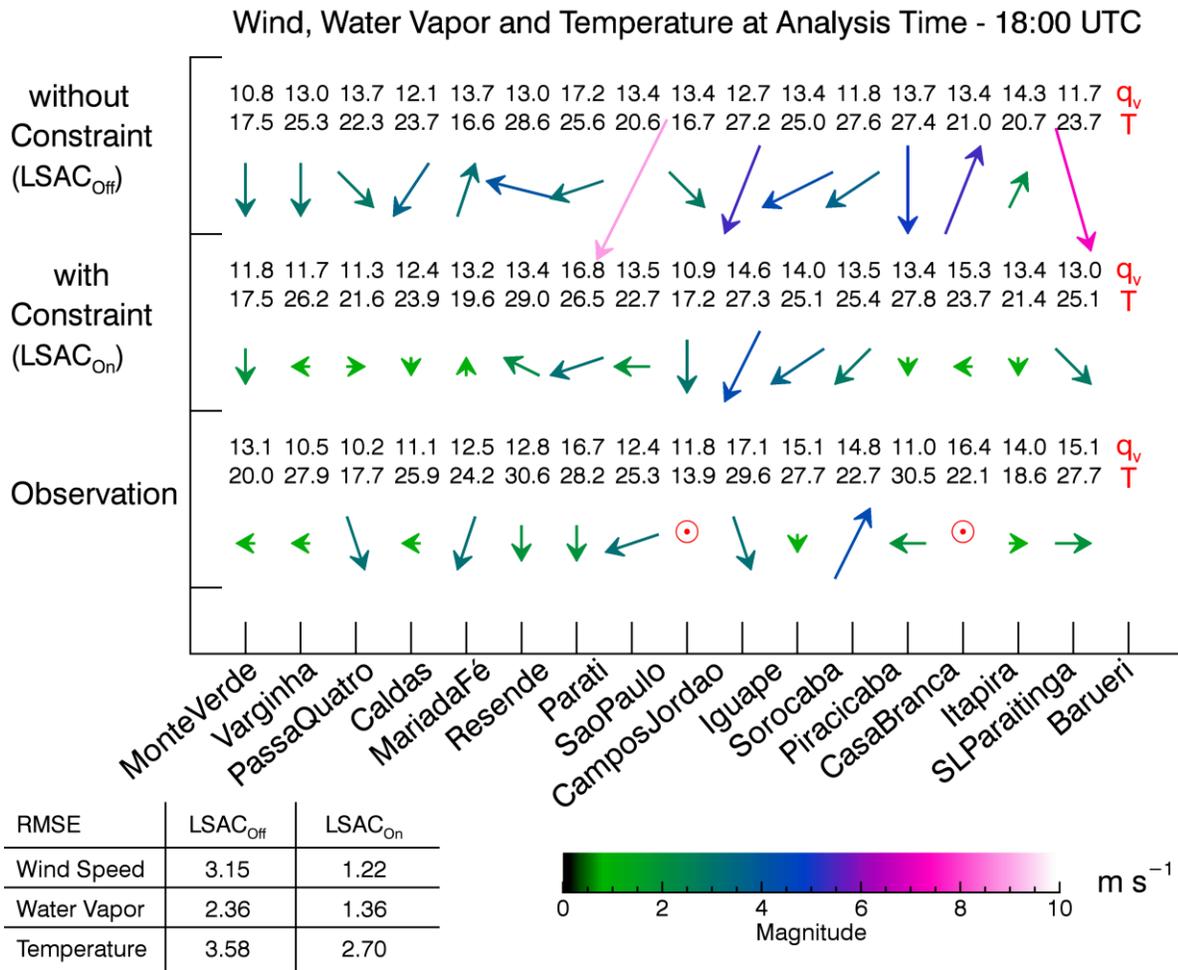
882

883



884

885 Figure 12: 10-m wind vectors and speed (shaded) in [ $\text{m s}^{-1}$ ] from LSAC<sub>Off</sub> (upper), LSAC<sub>On</sub>  
 886 (middle) and NCEP/CSFR 2 reanalysis with  $0.5^\circ$  resolution (low), valid at 18 UTC, January 22<sup>nd</sup>,  
 887 2012.



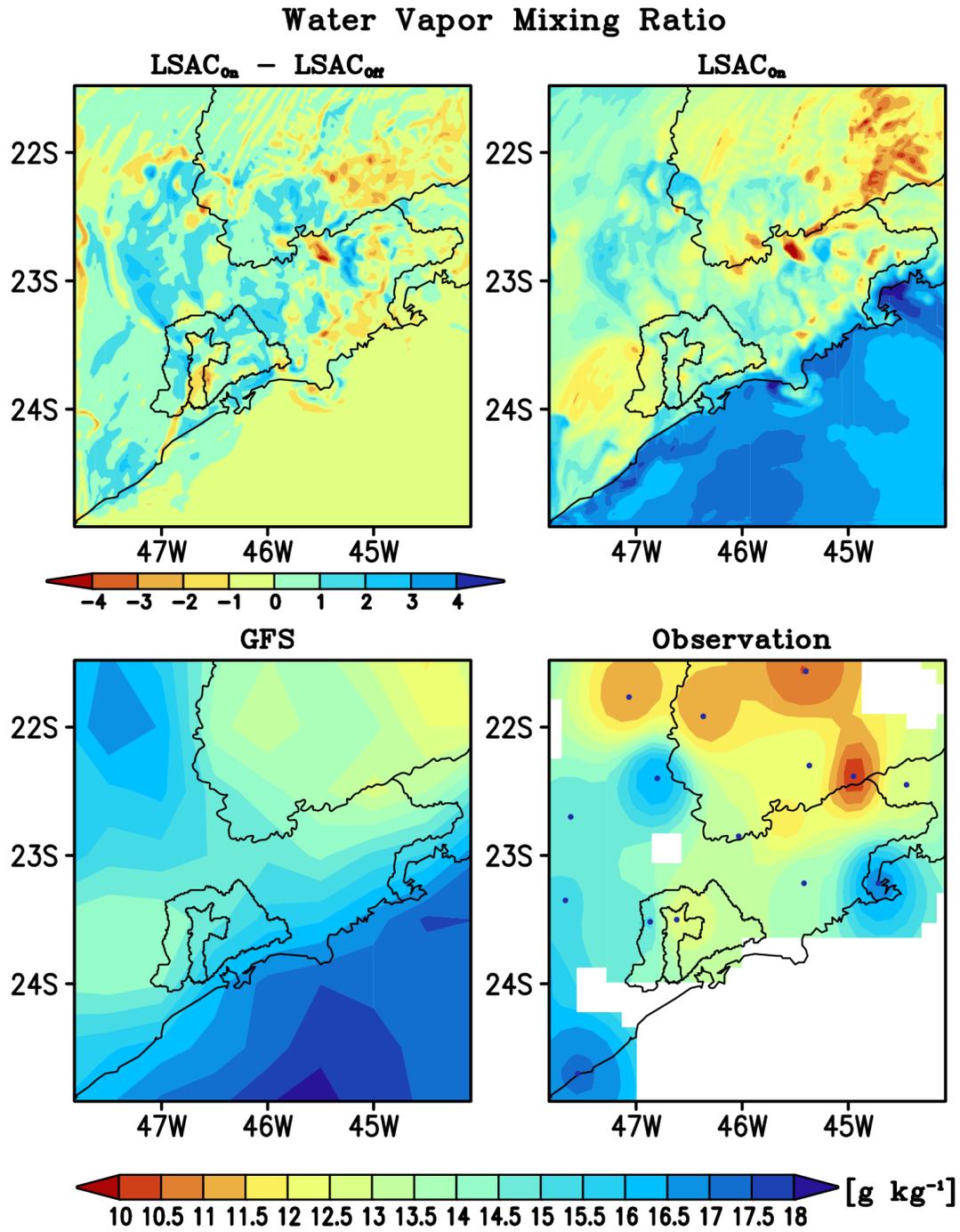
888  
 889 Figure 13: Evaluation of the wind, water mixing ratio and temperature of the LSAC<sub>Off</sub> and  
 890 LSAC<sub>On</sub> experiments using the observations from stations shown in Fig. 1 for the case of January  
 891 22<sup>nd</sup>, 2012 at 18 UTC.

892  
 893  
 894

895

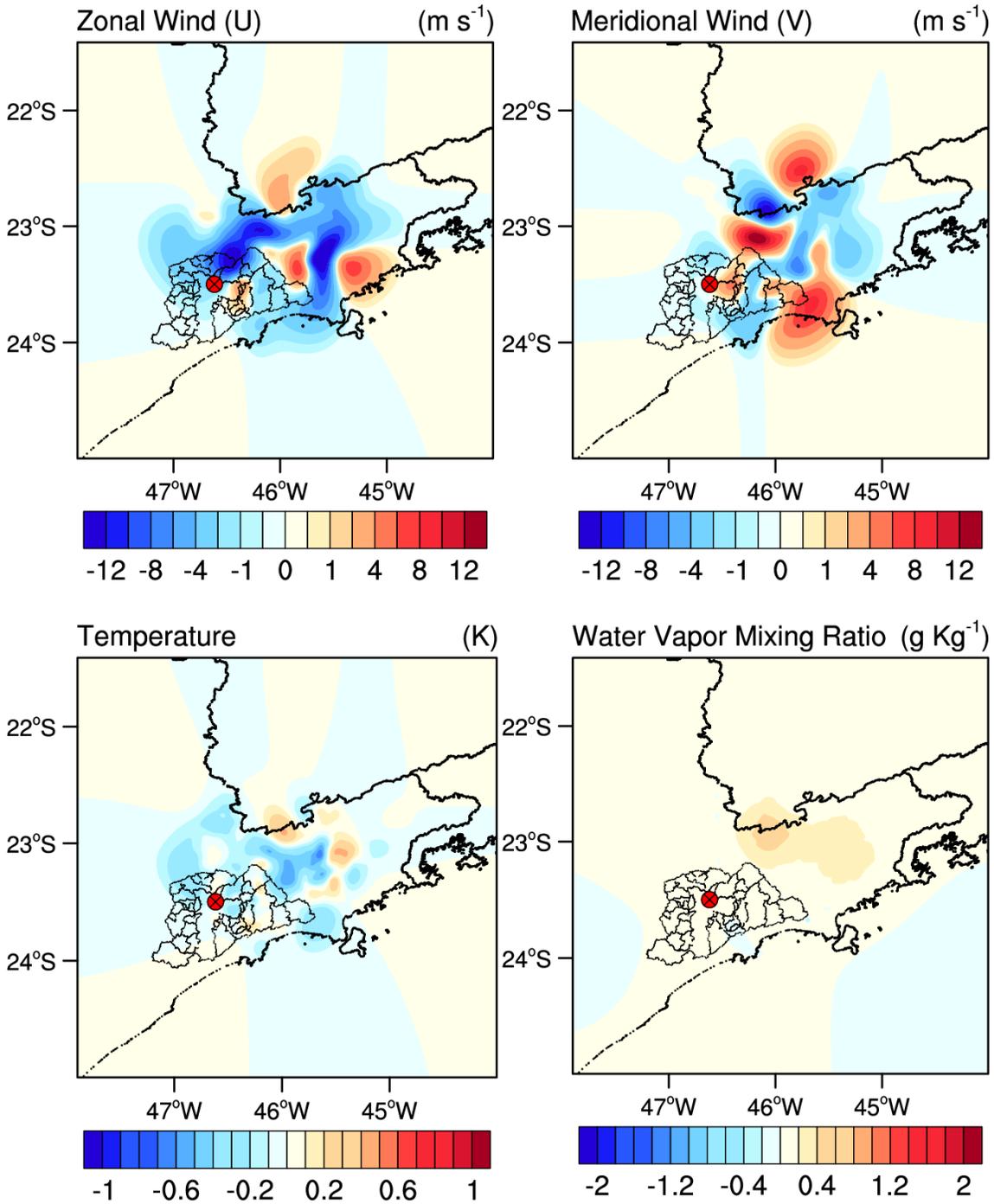
896

897



898

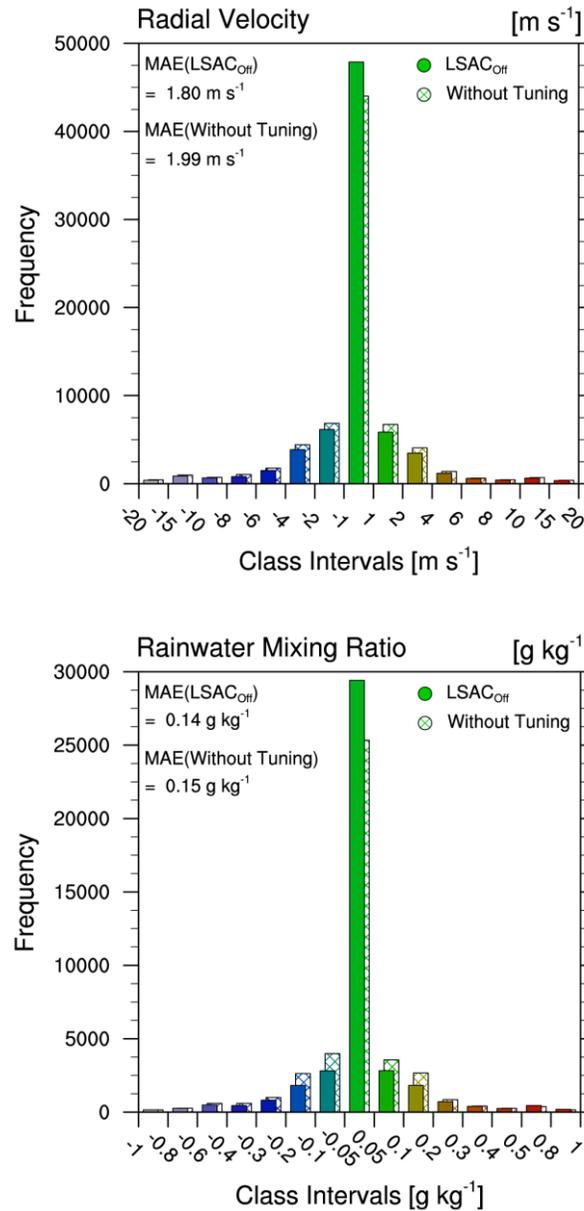
899 Figure 14: Water vapor mixing ratio near the surface for LSAC<sub>Off</sub> (upper-left), LSAC<sub>On</sub> (upper-  
 900 right), GFS (low-left) and observation (low-right) in [ $\text{g kg}^{-1}$ ] all for the case of January 22<sup>nd</sup>,  
 901 2012. The observation is an interpolation from surface weather stations (blue dots).



903 Figure 15: Same as Fig. 10, but for the experiment without tuning the length scale.

904

905



906 Figure 16: Difference between observation and analysis (O-A) of radial velocity (upper panel)  
907 and rainwater mixing ratio (lower panel) for the experiment LSAC<sub>Off</sub> (shaded) and the  
908 experiment without tuning the length scale (crosses). The Mean Absolute Error for each  
909 experiment is also shown.