Constraining a 3DVAR Radar Data Assimilation System with Large-Scale Analysis to Improve Short-Range Precipitation Forecasts

EDER PAULO VENDRASCO
National Institute for Space Research (INPE), Cachoeira Paulista, Sao Paulo, Brazil

JUANZHEN SUN
National Center for Atmospheric Research,* Boulder, Colorado

DIRCEU LUIS HERDIES
National Institute for Space Research (INPE), Cachoeira Paulista, Sao Paulo, Brazil

CARLOS FREDERICO DE ANGELIS
National Center for Monitoring and Early Warning of Natural Disaster, Cachoeira Paulista, Sao Paulo, Brazil

(Manuscript received 24 December 2014, in final form 3 August 2015)

ABSTRACT

It is known from previous studies that radar data assimilation can improve short-range forecasts of precipitation, mainly when radial wind and reflectivity are available. However, from the authors’ experience radar data assimilation, when using the three-dimensional variational data assimilation (3DVAR) technique, can produce spurious precipitation results and large errors in 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 3DVAR 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 3DVAR analysis with the constraint is more accurate when verified against an independent dataset. The performance of this new constraint on improving precipitation forecasts is tested using six convective cases and verified against radar-derived precipitation by employing four skill indices. All of the skill indices show improved forecasts when using the methodology presented in this paper.

1. Introduction

In recent years, an increase in natural disasters such as tornados and extreme precipitation events has been globally observed (Coumou and Rahmstorf 2012). Therefore, there has been an increased demand for improved forecasting of severe convective weather and its associated hazards. An improvement can be achieved by either increasing the grid resolution, developing physical parameterization schemes suitable for high-resolution models, or improving the initial conditions. Roberts and Lean (2008), for example, conducted experiments using 12-, 4-, and 1-km horizontal resolutions for 10 days of summer convection cases, and their results showed that the 1-km model grid was more skillful than the 12-km model grid at all horizontal scales greater than 15 km for heavy precipitation. Besides the model resolution, the initial conditions also play an important role in the skill of convective forecasting, and, therefore, it is crucial to initialize the model using observations that describe not only the large scale, but also the convective scale. The use of convective-scale

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Eder Paulo Vendrasco, Rodovia Presidente Dutra, Km 39, Cachoeira Paulista, Sao Paulo CEP 16300-000, Brazil.
E-mail: eder.vendrasco@ftp.cptec.inpe.br

DOI: 10.1175/JAMC-D-15-0010.1

© 2016 American Meteorological Society
observations such as radar data through data assimilation can potentially improve the initial conditions and lead to better convective weather prediction. Data assimilation is a technique for generating an accurate image of the true state of the atmosphere at a given time in which the observed information is accumulated into the model state by taking advantage of consistency constraints with laws of time evolution and physical properties.

According to Bouttier (2009), the most challenging aspects of finescale data assimilation stem from the occurrence of complex, intermittent convective structures in the atmosphere (and in numerical models) that have a strong sensitivity to the initial conditions and the potential for rapid error growth. On the one hand, kilometer-scale numerical weather prediction (NWP) systems can better provide important background information needed for convective-scale data assimilation, including low-level winds, convergence lines, tropospheric humidity, and preexisting convective systems that may influence subsequent convective triggering. In addition, the location and magnitude of these features can have significant differences from the reality of the atmosphere.

Doppler radar observations (reflectivity and radial velocity) have been used in complex data assimilation systems in order to improve the initial conditions for convection-permitting models (e.g., Caya et al. 2005; Sun et al. 2005; Tong and Xue 2005; Gao et al. 2004; Xiao et al. 2007; Ming et al. 2009; Aksoy et al. 2009, 2010; Wang et al. 2013a) since they are almost the only source of three-dimensional information at this scale (Aksoy et al. 2009). These data assimilation systems—for example, three- and four-dimensional variational data assimilation methods [3DVAR (Sasaki 1970; Lorenc 1986) and 4DVAR (Lorenc 1988)] and ensemble Kalman filter (EnKF; Evensen 1994) and hybrid methods (Zupanski 2005)—have been developed for both research and operational purposes, and previous works using those systems have shown improvements in their precipitation forecasts (e.g., Snyder and Zhang 2003; Kawabata et al. 2007; Tong and Xue 2008; Shimizu et al. 2011; Sun et al. 2012; Sun and Wang 2013; Wang et al. 2013b).

Although the above-mentioned studies have demonstrated that radar data assimilation can improve the forecasting of precipitation, the positive impact tends to last only for a few hours. Using a 3DVAR radar data assimilation system, Sun et al. (2012) found that the impact of assimilating Doppler radars lasted only for 3 h for forecasts initialized in the early mornings when the convective systems are mostly scattered and radar coverage is small. From our experience with 3DVAR, we have found that radar data assimilation can introduce large wind increments and spurious precipitation, which can potentially make the positive impact of radar data assimilation short lived and even cause adverse effects on longer forecasts. We believe one of the reasons for this behavior could be due to the lack of proper balance in the dynamical and microphysical fields of the initial analysis. According to Reen (2007), the 3DVAR method tends to produce noise, and a digital filter (Lynch 1993) can be used in an attempt to diminish it. However, besides the fact that the digital filter may also remove real atmospheric features, it can only be implemented after the data assimilation process. Another methodology that tries to deal with the noise produced in a 3DVAR system is an incremental analysis update (IAU; Bloom et al. 1996). Lee et al. (2006) performed experiments using the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5) 3DVAR with the IAU methodology. The authors found that IAU could reduce the initial noise, and the moisture (i.e., the model variables related with the water cycle) spun up efficiently because of the gradual incorporation of the analysis increment. However, the IAU method is unable to eliminate the noise in the analysis; rather, it reduces the impact of the noise by the gradual addition of the increment during the forecast. Although 4DVAR and hybrid (4DVAR + EnKF) systems may have advantages in producing better-balanced analyses and hence a greater potential than the 3DVAR, their computational costs are still much higher than what many operational centers can afford.

Sugimoto et al. (2009) conducted experiments assimilating radar data using the Weather Research and Forecasting (WRF) Model’s (Skamarock et al. 2008) three-dimensional variational data assimilation system (WRFDA 3DVAR; Barker et al. 2004) with observation simulation system experiments. Although encouraging results were shown in their study, they also found that the 3DVAR had limited ability in retrieving the unobserved variables by radar. This implies that the wind increment resulting from the assimilation of radar radial velocity is not balanced by other fields at the convective scales and can contain significant noise. The noise can even contaminate the large-scale balance. Guo et al. (2007) and Ming et al. (2009) addressed the issue of the final analysis balance when using the WRFDA 3DVAR scheme. The authors subjectively reduced by half these background error variances in order to give more weight to the background during the assimilation process for the purpose of keeping the background dynamical balance.

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

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

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

2. Methodology

a. WRFDA 3DVAR

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

\[
J = J_b + J_o + J_c = \frac{1}{2} v^T v + \frac{1}{2} (d - H' U v)^T R^{-1} (d - H' U v),
\]

where \( J_b \) and \( J_o \) stand for the background and observation terms, respectively. The term \( v \) is the control variable (CV) defined by \( v = U^{-1}(x - x_b) \), where \( U \) is the decomposition of the background error covariance \( B \) via \( B = U U^T \); \( x \) is the full analysis variable; and \( x_b \) is the background variable. The innovation vector that measures the departure of the observation \( y_o \) from its counterpart computed from the background \( x_b \) is given by \( d = y_o - H(x_b) \). Here, \( H' \) is the linearization of the nonlinear observation operator \( H \), and \( R \) is the observation error covariance matrix.

The control variables used in this study are velocity components \( u \) and \( v \), temperature \( T \), surface pressure \( P_s \), and pseudo–relative humidity (RH, where the humidity is divided by its background). These variables are the same as in previous radar data assimilation studies using WRFDA 3DVAR except for the absence of momentum variables. A recent study by Sun et al. (2016) compared the original momentum control variables of the streamfunction \( \psi \) and velocity potential \( \chi \) with that of velocity components \( u \) and \( v \) and concluded that the \( u-v \) pair outperformed \( \psi - \chi \) for limited-area convective-scale data assimilation. They found that the use of the \( u-v \) control variables allowed closer fits to dense observations such as radar radial velocity and resulted in improved precipitation forecasts. The \( u-v \) momentum control variables have been used in other variational convective-scale data assimilation systems in previous studies (i.e., Sun and Crook 1997; Zou et al. 1995; Gao et al. 1999). We tested both control variable options and obtained conclusions consistent with the results of Sun et al. (2016). For reflectivity data assimilation, we follow the indirect assimilation scheme described by Wang et al. (2013a) in which the retrieved rainwater instead of the reflectivity itself is assimilated. The authors argued that assimilating the rainwater mixing ratio avoided nonlinearity issues caused by the linearized observation operator for reflectivity required in the incremental formulation.

b. Large-scale analysis constraint

The LSAC is introduced into WRFDA 3DVAR by adding a new term \( J_c \) to Eq. (1) that measures the deviation of the 3DVAR analysis from a coarser-resolution large-scale analysis for horizontal velocity components, temperature, and humidity:

\[
J = J_b + J_o + J_c = J_b + J_o + \frac{1}{2} (d_c - H' U v)^T R_c^{-1} (d_c - H' U v),
\]

where \( d_c \), given by \( d_c = y_c - H(x_b) \), is the innovation vector that measures the departure of the LSAC \( y_c \) from
The term \( \mathbf{R}_x \) is the large-scale analysis error covariance matrix for \( u, v, T, \) and \( q_v \). The \( \mathbf{R}_x \) matrix is constructed by considering constant uncorrelated errors for each variable: 2.5 m s\(^{-1}\) for wind components, 2°C for temperature, and 3 g kg\(^{-1}\) for the water vapor mixing ratio. These values are chosen based on the diagnostics of the performance of the GFS available from the Environmental Modeling Center’s web page (http://www.emc.ncep.noaa.gov/GFS). The sensitivity of the analysis and forecast with respect to the error variance was evaluated by performing (not shown) experiments that increase or decrease the errors by 10% and no significant sensitivity was found. The large-scale analysis data used in LSAC are mapped to the closest 3DVAR analysis grid point by trilinear interpolation.

c. Methods for precipitation verification

Our primary goal in implementing radar data assimilation is to improve precipitation forecasts. Therefore, besides some diagnoses intended to evaluate the quality of the 3DVAR analysis, the skill of the precipitation forecasts is also used to measure the performance of the data assimilation experiments. The fractional skill score (FSS; Roberts and Lean 2008) and the local root-mean-square error (LRMSE) are applied for the precipitation verification. The FSS is defined by

\[
FSS = 1 - \frac{\text{FBS}}{\text{FBS}_w} = 1 - \frac{1}{N} \sum_{k=1}^{N} [P_{M(k)} - P_{O(k)}]^2, \tag{3}
\]

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

\[
\text{LRMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (\overline{M}_k - \overline{O}_k)^2}, \tag{4}
\]

where \( \overline{M} \) and \( \overline{O} \) are the forecast and observed precipitation amounts averaged over all grid points within the same radius of influence used for calculating the FSS. The subscript \( k \) represents the \( k \)th grid point and \( N \) the total number of grid points in the verification domain. In addition to the above two scores, the traditional skill scores of the false-alarm ratio (FAR) and probability of detection (POD) are also used to verify the precipitation forecast. The FAR and the POD are defined by

\[
\text{FAR} = \frac{\text{false alarms}}{\text{false alarms} + \text{hits}}, \quad \text{and} \quad \text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}}. \tag{5, 6}
\]

and the contingency table is defined in Table 1.

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

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False alarms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct negatives</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1. Contingency table.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
</tr>
<tr>
<td>Hits</td>
</tr>
<tr>
<td>Misses</td>
</tr>
<tr>
<td>False alarms</td>
</tr>
<tr>
<td>Correct negatives</td>
</tr>
</tbody>
</table>

3. Observations and experimental setup

Assimilated observations in this study include radar data from the Cloud Processes of the Main Precipitation...
Systems in Brazil: A Contribution to Cloud Resolving Modeling and to the Global Precipitation Measurement (GPM), known as the CHUVA project (Machado et al. 2014), and the surface and upper-air conventional data from the Global Telecommunication System (GTS). The CHUVA project aims to map the main precipitating systems across Brazil. So far, experiments have been performed at seven sites with various measurements including those with dual-polarization radar, lidar, microwave radiometers, disdrometers, radiosondes, rain gauges, and various other instruments. Only data from the Paraiba Valley Experiment were used and six convective cases were chosen to examine the proposed new LSAC methodology. Figure 1 shows the topography of this region, the X-band radar site used in this study, and the surface network used for verification. Table 2 summarizes the six convective cases studied in this paper. Figure 2 shows the 3-km constant-altitude plan position indicator (CAPPI) for each case at 1800 UTC observed by the X-band radar. These convective systems all produced localized heavy rainfall. Because of the transient and small-scale nature of storms, it is challenging for operational NWP models to accurately forecast them.

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

The conventional observations are obtained from GTS, and they include pilot reports (AIREPs), METARs, surface synoptic observations (SYNOPs), ship reports, and soundings in this region. Figure 3 shows the locations of the conventional data. As can be seen, the amount of data is quite small and sparse, even though this region is very important because of the factors explained before. Particularly in this area, because of the limited amount of GTS data and the small domain, the impact of the radar data is very strong and can dominate over the GTS observations. The parameterizations used in the WRF forecast include the WRF single-moment 6-class microphysics (WSM6; Hong and Lim 2006), the Rapid Radiative Transfer Model longwave radiation (Mlawer et al. 1997), the Dudhia shortwave radiation scheme (Dudhia 1989), and the Yonsei University planetary boundary layer (Hong et al. 2006). The initial and boundary conditions are derived from the GFS 0.5° × 0.5° analysis, and the model domain is shown in Fig. 1. The horizontal grid resolution is 2 km, and the vertical resolution is about 60 m near the surface and 1 km in higher levels with a total of 45 sigma levels. Previous studies have shown (e.g., Bryan and Morrison 2012) that 60-m vertical resolution is sufficient for modeling convection in the planetary boundary layer. A small domain was chosen because the goal was to produce high horizontal resolution QPFs for the nowcasting range of 6 h and also because of the use of a single radar that is strategically located in the Paraiba valley close to
FIG. 2. The 3-km radar reflectivity (dBZ) for all six convective cases at 1800 UTC.
the SPMR. Obtaining a QPF with high accuracy is very important in this area in order to give support in case of natural disasters (e.g., floods).

The National Meteorological Center (NMC, now known as NCEP) method (Parrish and Derber 1992) was used in this work to generate the background error statistics using the gen_be utility from the WRFDA system. A dataset containing 3 months of cold-start 24-h forecasts over a domain covering the Southern Hemisphere summer was produced every day starting at 0000 and 1200 UTC. The differences between the 24- and 12-h forecasts valid at the same times were used to calculate the domain-averaged background error statistics.

For each of the six cases, the data assimilation and forecast experiment commenced at 1200 UTC with the GFS analysis as the initial conditions. A 3-h forecast was conducted before the cycled 3DVAR analyses began. Four continuously cycled analyses were performed at 1500, 1600, 1700, and 1800 UTC and then a 6-h forecast ensued. The assimilation process was divided into two steps: first, the GTS data were assimilated and, second, the radar data were assimilated. In the second step, if the switch for the LSAC was turned on, then the large-scale analysis was assimilated together with the radar data. The cycling strategy is illustrated in Fig. 4.

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

Three numerical experiments (Table 3) were performed to evaluate the impact of the LSAC on the analysis and precipitation forecast when assimilating radar data. The control run (CTR) is the run in which only conventional data from GTS are assimilated. In the other two experiments, LSACOff and LSACOn, both conventional and radar data are assimilated without or with the constraint term. The radar data are thinned to the 1-km resolution from the original 250-m level and interpolated horizontally into a regular grid while the original polar coordinate is kept in the vertical.

4. Results

The 6-h precipitation forecasts are evaluated for all six convective cases using the averaged FSS and LRMSE over the six cases, and their results are shown by Fig. 5. The verification is against the hourly precipitation estimate obtained from the radar reflectivity. The experiment LSACOff shows improvement over the CTR for both thresholds of 1 and 5 mm h$^{-1}$, which is expected and consistent with previous works. The radar data assimilation usually improves the forecast in the beginning, and then it can lose skill depending on many factors like the forcing mechanism of the convective systems. If more weight is given to the radar data during the analysis process, the precipitation forecast can result in better skill in the first 1 or 2 h. However, because of the imbalance generated by forcing the model toward the radar data the skill can drop and become worse than the case without any radar data assimilation. Nevertheless, in
experiment LSAC_{Off}, all 6-h forecasts are improved with the radar data assimilation on average over the six convective cases.

The LSAC_{On} experiment results in improvement over the LSAC_{Off} experiment for all 6-h forecasts in the averaged FSS. Figure 5 shows that the LSAC_{On} experiment reduces the LRMSE, since it decreases the spurious convection (shown later). The improvement obtained in the LRMSE is due to the precipitation forecast, with the intensity agreeing better with the radar observations. The improvement is also shown in Figs. 5a and 5b, where the standard deviation of the mean over the six cases for the three experiments and the variance of LRMSE among the cases is smaller in experiment LSAC_{On} compared with experiments CTR and LSAC_{Off}. This result suggests that in spite of the different types of behavior among the cases, which is expected considering the different types of precipitation, the spread of the error in the amounts of precipitation is smaller in the LSAC_{On} experiment. In other words, the positive impact of the radar data assimilation seems to be more homogeneous among all cases when compared to the experiment without the LSAC. For the FSS the variance from the LSAC_{On} is not always smaller relative to the CTR and LSAC_{Off} experiments, which means that the error in the position of the precipitation is not consistently improved.

The averaged FAR and POD over all cases are shown in Fig. 6, and it is observed that the inclusion of radar data (LSAC_{Off}) increases the POD. However, the FAR also increases compared with the control run. On the other hand, when the LSAC is turned on, not only is the POD slightly better than in the case of LSAC_{Off}, but also the FAR is decreased. Therefore, this result corroborates the idea that using LSAC suppresses spurious convection. As shown in Fig. 5, the standard deviation of the mean was included in Fig. 6 and the result is quite similar, as the standard deviation is smaller at almost all forecast times in the LSAC_{on} experiment.

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

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

### Table 3. Experiment descriptions.

<table>
<thead>
<tr>
<th>Expt</th>
<th>Assimilated data</th>
<th>Constraint term</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTR</td>
<td>Conventional (GTS)</td>
<td>Off</td>
</tr>
<tr>
<td>LSAC_{Off}</td>
<td>Conventional + radar (reflectivity and radial</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>velocity)</td>
<td></td>
</tr>
<tr>
<td>LSAC_{On}</td>
<td>Conventional + radar (reflectivity and radial</td>
<td>On</td>
</tr>
<tr>
<td></td>
<td>velocity)</td>
<td></td>
</tr>
</tbody>
</table>
\[
N = \frac{1}{I \cdot J} \sum_{i=1}^{I} \sum_{j=1}^{J} | \frac{\partial P_s}{\partial t} |_{ij}
\]

where \( P_s \) is the surface pressure. The summation denotes the calculation over the whole model domain.

Figure 7 shows the domain-averaged absolute surface pressure tendency in 30 and 60 min over the forecast interval. The results show that the LSAC_{Off} experiment experiences a larger adjustment than does the LSAC_{On} experiment. In LSAC_{Off} (LSAC_{On}), \( N \) has an initial value of 38.4 (27.2) Pa h\(^{-1}\) and reaches an asymptotic value around 15.5 (13.2) Pa h\(^{-1}\) for the 60-min tendency. This result shows that the LSAC can help produce an analysis containing less noise, which is very helpful when the radar observations are assimilated.

The precipitation skill scores (FSS and LRMSE) are shown for the case of 22 January 2012 in Fig. 8. The behavior of these skill scores is similar to what is observed in Figs. 5a and 5b except for the better performance of the CTR between the 1- and 2-h forecasts for the 1-mm threshold when its FSS is very close to that of LSAC_{Off} but still worse than that of LSAC_{On}. This similar behavior is important for the following detailed analysis because it suggests that this specific convective case can be representative of the overall results.

The hourly accumulated precipitation fields at \( t = 1 \) h for LSAC_{Off} and LSAC_{On} are shown in Fig. 9 and compared against the radar-derived precipitation. It is evident that the precipitation is overestimated in experiment...
LSAC\textsubscript{Off}. The left red circle in Fig. 9 points out a specific area where much more precipitation is predicted when the constraint is turned off, while inside the right red circle the opposite is observed; that is, the LSAC\textsubscript{Off} underestimates the precipitation amount and the LSAC\textsubscript{On} produces a better forecast for this region. Furthermore, the overprediction is not only in the amount of precipitation, but also in the spatial distribution. A closer examination of Fig. 9 reveals how the large-scale analysis suppresses some spurious convection resulting in improved precipitation forecast skill. For example, the overestimated precipitation over the Minas Gerais state (Fig. 9; 22.50°S and 45.25°W) in LSAC\textsubscript{Off} is suppressed in LSAC\textsubscript{On}. Note that despite the small differences in the skill scores presented in Fig. 5, the forecast precipitation is highly sensitive to these errors and does impact the

![Fig. 9](image-url)  
**FIG. 9.** The 1-h forecast of precipitation for the case of 22 Jan 2012 for (a) experiment LSAC\textsubscript{Off} and (b) experiment LSAC\textsubscript{On}, and (c) the result retrieved from radar reflectivity.

![Fig. 10](image-url)  
**FIG. 10.** For experiments (a) LSAC\textsubscript{Off} and (b) LSAC\textsubscript{On} at the analysis time 1800 UTC 22 Jan 2012, analysis increments at the third model level (around 150 m) for (clockwise from top left) zonal and meridional wind components (m s\textsuperscript{-1}), water vapor mixing ratio (g kg\textsuperscript{-1}), and temperature (K).
precipitation prediction. This result is in part due to the
adjustment of the water vapor field when using LSAC.
The improvements in the dynamical and microphysical
fields like wind, temperature, and water vapor are de-
tailed in the following discussion.

The analysis increments for wind components,
temperature, and water vapor mixing ratio are shown
in Fig. 10. By comparing the increments between
LSACon and LSACoff, we note the following impor-
tant differences. First, for the wind components, the
increments from experiment LSACOff are concen-
trated where the radar data are available (see Fig. 1
for the radar coverage), whereas in experiment
LSACOn the increments are distributed over the en-
tire domain with larger-scale disturbances outside the
radar data region. Second, the temperature and water
vapor increments from experiment LSACoff are
considerably smaller than those from LSACon. Thir-
d, the magnitudes of the wind increments and the pat-
terns in the region with radar data are similar between
the two experiments, suggesting that the analysis with
LSAC is still able to achieve a close fit to the radar
observations. This is confirmed by a comparison of
the statistics of the radar observation minus the final
analysis \((O - A)\) between the two experiments, as
shown in Fig. 11. However, it is expected that the
LSAC not only maintains a close fit to the radar ob-
servations, but also improves the wind analysis in the
neighboring region surrounding the radar data, which
we will examine later in this section.

The 10-m surface analyses of the horizontal wind
vectors and speed at 1800 UTC are compared in Fig. 12
along with the Climate Forecast System Reanalysis 2
from NCEP (Saha et al. 2010). The analysis fields from
both LSACoff and LSACon have larger wind speeds over
the land than the reanalysis because of the higher res-
olution and convection. However, large changes in wind
speed and direction from LSACoff are identified in the
northwest quarter of the domain, corresponding to the
spurious convection in that region (see Fig. 9a). In
contrast, the result from LSACOn shows the higher wind
speed located mainly along the coastal region where the
convection occurs (see Fig. 2; experiment 4), and the
wind field is generally smoother and more comparable
to the reanalysis. This pattern is much more coherent
than the one shown in the LSACOff experiment. The
winds in LSACOff clearly show downburstlike outflows,
which are probably due to downdrafts caused by the
raining out of unsustainable rainwater added by the data
assimilation.
To verify the accuracy of the analyses with and without the LSAC, we resort to the independent surface dataset collected by the Brazilian National Institute of Meteorology [Instituto Nacional de Meteorologia (INMET)]. This surface network covers the entire country and some of the stations are located within the studied area, as shown in Fig. 1. In Fig. 13, objective comparisons are made for surface wind, water vapor, and temperature at each station for the 16 stations in the domain. The wind direction and speed are much more coherent in LSACOn than in LSACOff when compared with the observations. The same positive impact from LSAC is observed on temperature and water vapor mixing ratio. The RMS error (RMSE) differences for the three fields are computed, and their results are shown in the table in the lower-left corner of Fig. 13. The decreases in the RMSE due to the LSAC are 62%, 42%, and 25% for wind speed, water vapor, and temperature, respectively. These verification results clearly show that the LSAC yields improved analyses of these fields.

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

Since water vapor plays an important role in convective initiation, next we provide a closer evaluation on how the water vapor field is improved by the LSAC. Figure 14 shows a comparison between experiments LSACOff and LSACOn, along with the GFS analysis and a surface analysis using all 16 stations of INMET. The GFS analysis agrees relatively well with the surface analysis in the large-scale sense; the wetter west and the drier north-northeast distribution in the domain are well captured by the GFS analysis. The difference between the LSACOn and LSACOff water vapor fields shows that considerable improvement is made by the large-scale analysis constraint. The west part from LSACOn is still relatively drier than the observations, even though the difference between LSACOn and LSACOff is positive in that area (i.e., LSACOff is even drier). In Fig. 9 it is shown that the precipitation field over the Minas Gerais state is improved by reducing the spurious precipitation, and Fig. 14 shows that the water vapor is overestimated.

![Fig. 12. The 10-m wind vectors and speed (shaded; m s$^{-1}$) from (top) LSACOff, (middle) LSACOn, and (bottom) NCEP Climate Forecast System Reanalysis 2 with 0.5° resolution, valid at 1800 UTC 22 Jan 2012.](image-url)
(negative values in the difference field) in that state by LSAC$\text{Off}$ but is corrected in LSAC$\text{On}$. Another similar example is the overprediction of precipitation south of São Paulo (Fig. 9; $23.80^\circ$S and $46.65^\circ$W) in the LSAC$\text{Off}$ experiment caused by the large amount of water vapor.

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

One may wonder whether the large-scale increment in the region without radar can be achieved by using a larger length scale of the background error covariance. To answer that question, an extra experiment was performed in which the background error without any tuning was used in the LSAC$\text{Off}$ experiment. The increments from this experiment are shown in Fig. 15. The results reveal that when using a larger length scale the increments have a wider spread than those in Figs. 10a and 10b, just as expected. However, the small-scale increments where the convection occurs are smoothed (resulting in a poorer fit to the observations), and the large increments in the nearby environment are questionable. Figure 16 compares the fit to the radial velocity and rainwater mixing ratio (converted from the reflectivity) between the two experiments with (LSAC$\text{Off}$) and without tuning of the background error statistics. It shows that the fit to the observations is improved when the length scale is tuned in LSAC$\text{Off}$. In the case without tuning, the $O - A$ clusters in the smaller class intervals of $[-1, 1] \text{m s}^{-1}$ for radial velocity and $[-0.05, 0.05] \text{g kg}^{-1}$ for rainwater are smaller, and the mean absolute error (MAE) is greater for both radial velocity and rainwater mixing ratio. This result suggests that limiting the influence of the radar data to the radar coverage through the tuning of the length scale is a better choice and better results can be achieved. For those areas where radar information is not available, the LSAC can help to constrain and improve the final analysis.
5. Conclusions

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

![Water Vapor Mixing Ratio](image)

**FIG. 14.** Water vapor mixing ratio (g kg\(^{-1}\)) near the surface for (top left) LSAC\(_{\text{off}}\), (top right) LSAC\(_{\text{on}}\), (bottom left) GFS, and (bottom right) the observations for the cases on 22 Jan 2012. The observation is an interpolation from surface weather stations (blue dots).
The results demonstrated that the LSAC introduced in the WRF 3DVAR improved the QPF by producing improved analyses of wind, humidity, and temperature when verified by surface observations. These fields were not only improved in the convective region where radar observations were available but also in the surrounding region by reducing errors caused by spurious convection. The 3DVAR with LSAC filtered out unreliably high increments of wind, producing a wind field that was much more reasonable and coherent with the observed convective system without losing the beneficial information from the radar. The comparison of the analyzed water vapor mixing ratio from both experiments against surface observations also showed that the LSAC was able to produce an analysis much closer to the observations, and this improvement was crucial for supporting convection in the right location and eliminating spurious convective activities. It is important to note...
that the improvement achieved in the precipitation prediction due to the inclusion of the LSAC was obtained despite the small differences in the averaged skill scores between the LSACon and LSACoff experiments, showing that the precipitation forecast is highly sensitive to these errors. Moreover, it was shown through the domain-averaged pressure tendency that the noise in the initial analysis could be reduced using the constraint. The reason for this reduction is that the large-scale analysis is better balanced and it filters out the small-scale disturbances.

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

This study has shown the importance of LSAC for radar data assimilation in a 3DVAR data assimilation system using a few convective cases over a specific region. Nevertheless, the technique is applicable in general for regional data assimilation involving high-resolution data using any 3DVAR data assimilation system. For example, when assimilating large amounts of high-resolution data in an operational cycle, the cycle cannot last too long since the errors grow quickly and the large-scale pattern can be lost, making it necessary to restart the cycle every few hours to alleviate the problem. In this case, the LSAC could be used to constrain the data assimilation process such that the large-scale pattern can be continuously kept, without the need to restart the process. Additionally, the method is relatively easy to implement and with only slight increases in computation cost, making it a practical approach to improving the performance of high-resolution regional data assimilation.

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

Acknowledgments. The authors thank Brazil’s National Council for the Improvement of Higher Education (CAPES) and the National Center for Atmospheric Research (NCAR) sponsored by the National Science Foundation for their support and the CHUVA project (Foundation for Research Support of São Paulo—FAPES Grant 2009/15235-8) for providing the radar data.
REFERENCES


Unauthenticated | Downloaded 01/29/24 08:43 AM UTC