Sensitivity of WRF Forecasts for South Florida to Initial Conditions

BRIAN ETHERTON

Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina

PABLO SANTOS

National Weather Service, Miami, Florida

(Manuscript received 20 December 2006, in final form 14 December 2007)

ABSTRACT

This study presents results from an experiment conducted to measure the impact of locally initializing a numerical weather prediction model on that model's ability to predict precipitation and other surface parameters. The study consisted of quantifying the impact of initializing the Weather and Research Forecast (WRF) model with the Advanced Weather Interactive Processing System (AWIPS) Local Analysis and Prediction System (LAPS) diagnostic analyses. In the experiment, WRF was run for two different initial times: 0600 and 1800 UTC. For each initial time, the model was run twice, once using LAPS for the initial conditions, and once using the North American Mesoscale model (NAM; also known as the Eta Model at the time of the experiment). The impact of the local LAPS initialization on the model forecast of surface parameters is presented. Additionally, the model's quantitative precipitation forecast (QPF) skill is compared for three different model configurations: 1) WRF initialized with LAPS, 2) WRF initialized with NAM, and 3) the standard NAM/Eta Model. The experiment ran from 1 June 2005 to 31 July 2005.

Results show that WRF forecasts initialized by LAPS have a more accurate representation of convection in the short range. LAPS-initialized forecasts also offer more accurate forecasts of 2-m temperature and dewpoint, 10-m wind, and sea level pressure, particularly in the short range. Most significantly, precipitation forecasts from WRF runs initialized by LAPS are more accurate than WRF runs initialized by NAM. WRF initialized with LAPS also demonstrates higher QPF skill than does the NAM/Eta Model, particularly in the short range when the precipitation thresholds are higher (0.25 in. in 3 h versus 0.10 in. in 3 h), and when forecasts are initialized at 0600 UTC rather than initialized at 1800 UTC.

1. Introduction

In south Florida, particularly during the summer, mesoscale weather features (e.g., land–sea breezes, thermal troughs, outflow boundaries, etc.) have a significant impact on day to day weather forecasting, as they frequently represent the primary forcing for convection. These mesoscale features necessitate the use of high-resolution forecast tools in order to provide the detailed information needed to improve local forecasts and warnings. The advent of the Local Analysis and Prediction System (LAPS) at National Weather Service (NWS) forecast offices (WFOs) has made it possible to ingest local high-resolution datasets to support local high-resolution analyses that more accurately resolve some of these features.

This study examines the impact of initializing a numerical weather prediction model, the Weather and Research Forecast model (WRF; detailed in section 2b), with high-resolution data, namely Advanced Weather Interactive Processing System (AWIPS) LAPS diagnostic analyses. In addition to evaluating the impact of the LAPS initialization on the WRF forecasts, comparisons of these forecasts to WRF forecasts initialized from the National Centers for Environmental Prediction (NCEP) North American Mesoscale model (NAM) and to forecasts from the NAM model are also presented. Intercomparisons of WRF forecasts using either NAM or LAPS for initial conditions were evaluated using 10-m winds, 2-m temperature, moisture information, and sea level pressure. Model performance for both LAPS- and NAM-initialized WRF precipitation forecasts as well as NAM precipitation fore-

Corresponding author address: Brian Etherton, Dept. of Geography and Earth Sciences, University of North Carolina at Charlotte, McEniry Bldg., Rm. 237, Charlotte, NC 28223.
E-mail: betherto@uncc.edu

DOI: 10.1175/2007WAF2006115.1

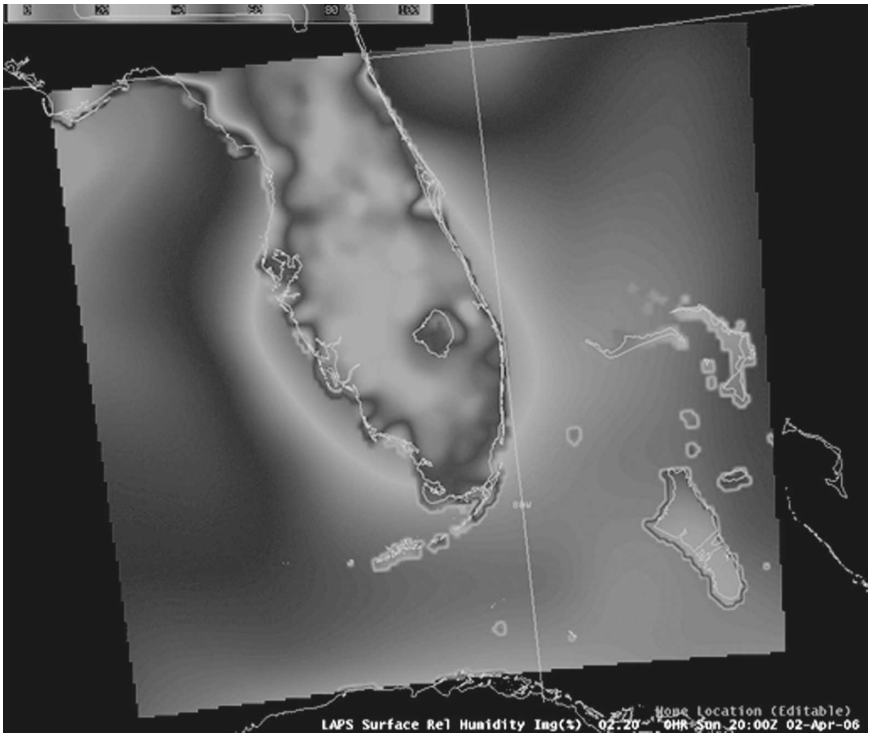

FIG. 1. Sample surface relative humidity analysis from LAPS illustrating the WFO Miami domain of the LAPS analyses.

casts was evaluated using grid-based threat scores, bias scores, the probability of detection, and false alarm ratios for different precipitation thresholds. The study ran from 1 June 2005 to 31 July 2005. This work is the result of a Cooperative Program for Operational Meteorology, Education and Training (COMET) Partners Project (information online at http://comet.ucar.edu/outreach/) between the NWS Forecast Office in Miami and the University of North Carolina at Charlotte.

Section 2 will describe the data and methods we used: LAPS, WRF, and our verification data. Results are presented in section 3, consisting of comparisons of forecasts of surface parameters (temperature, wind, pressure, and moisture) and of forecasts of precipitation. Conclusions are summarized in section 4.

2. Data and methods

a. Local Analysis and Prediction System

LAPS became available to the WFOs with the advent of AWIPS. As delivered in AWIPS, LAPS is a diagnostic tool only. LAPS produces high-resolution three-dimensional analyses of the atmosphere centered on a domain of the users choosing. Our analysis domain, centered on the WFO Miami County Warning Area (CWA), is shown in Fig. 1. The analysis is made by combining a first guess of the state of the atmosphere (the "background field") with locally and centrally available data from a wide variety of meteorological observation systems. The background field for the LAPS analyses is obtained from the AWIPS Rapid Update Cycle (RUC) 40-km 1-h forecast. The data used in LAPS analyses comes from local networks of surface observing systems, Doppler radars, satellites, and wind and temperature profilers [e.g., from the Radio Acoustic Sounding System (RASS) at 404 MHz and for the boundary layer at 915 MHz], as well as aircraft are incorporated into the analysis (Hiemstra et al. 2006; Albers 1995; Albers et al. 1996; Birkenheuer 1999; McGinley 2001; Schultz and Albers 2001). Figure 2 represents a summary of all of the data sources LAPS is capable of assimilating into its three-dimensional analyses, as well as those datasets used in the AWIPS LAPS running at WFO Miami. During the experimental period, analyses were produced twice an hour on a three-dimensional grid covering an area, as shown in Fig. 1, nearly 830 km east–west by 775 km north–south. The horizontal resolution of the hourly LAPS surface analyses produced at WFO Miami is 5 km, and there are 39

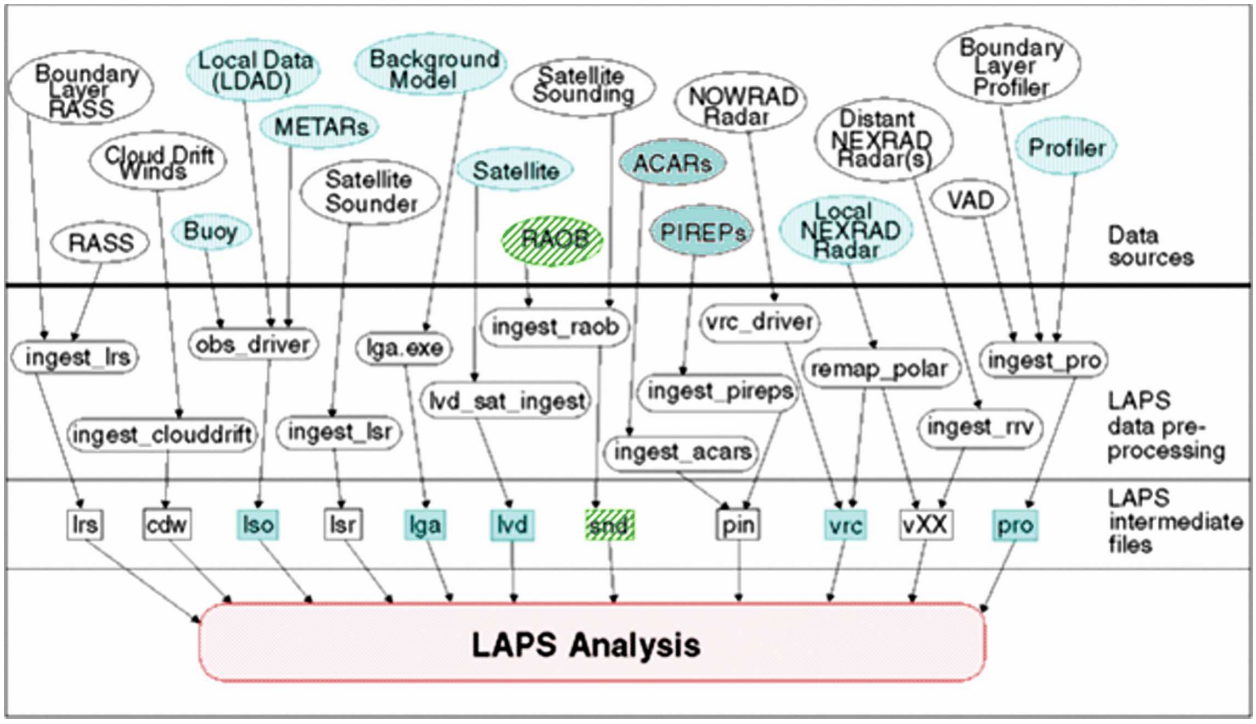

FIG. 2. Schematic of LAPS data sources. Although LAPS is capable of ingesting many different data streams, only those highlighted in blue and green are used in the operational LAPS analyses at a typical WFO running AWIPS operational build 5, which was available at the time the experiment was run (summer 2005).

vertical levels, at 25-mb intervals, from 1000 up to 50 mb.

As is evident in Fig. 2, not all data that LAPS is capable of ingesting are actually used operationally at the local WFO level. In addition, a Kalman filter for quality control (McGinley 2001) is not used in the WFO AWIPS version of LAPS due to hardware limitations. In an attempt to improve the accuracy of the local analyses, the WFO in Miami has worked on incorporating additional local data networks into the analysis via the Local Data Acquisition and Distribution (LDAD) system, which is a component of AWIPS. This effort has led to a substantial increase in the amount of surface data going into the analyses and an increase in the quality of the analyses, as documented in Etherton and Santos (2006). Bad input data are identified and removed from the analysis by either objective analysis routines built into LAPS or by blacklisting bad observations as identified by the forecasters. The inclusion of data from local mesonets enhances the LAPS analyses of both inland and coastal gradients, as well as its depiction of the effect of Lake Okeechobee on surface fields (Etherton and Santos 2006). Figure 3 illustrates these mesonets, showing the standard observing network and all of the surface observing sites whose data are fed into LAPS. These findings are also consistent with those of Hiemstra et al. (2006).

As illustrated in Fig. 2, LAPS ingests satellite data that it uses to create a three-dimensional cloud analysis (Albers et al. 1996; Birkenheuer 1999; Schultz and Albers 2001; Shaw et al. 2001). This cloud analysis is used in combination with quasigeostrophic balance constraints to create diabatic initialization grids used for the initialization/hot start of the WRF (Shaw et al. 2001). The local radar data used for the analysis are level III reflectivity data from both the WFO Miami and WFO Key West radars.

b. Weather and Research Forecast model

The Advanced Research WRF (ARW) model, version 2, was used for this project. The ARW system consists of the ARW dynamics solver together with physics schemes, initialization routines, and a data assimilation package [though for this study, the ARW data assimilation package, WRF three-dimensional variational data assimilation (3DVAR), was not used]. Skamarock et al. (2005) highlight the major features of version 2 of the ARW system. As used in this experiment, WRF is a fully compressible, Euler nonhydrostatic model using a terrain-following hydrostatic-

FIG. 3. (left) The std surface observing network (approximately 65 stations at any given time) and (right) the additional nonstd surface observations used in creating the LAPS analysis at WFO Miami (approximately 170 additional stations at any given time).

pressure vertical coordinate, with vertical grid stretching such that the vertical levels are closer together near the surface and more spread out aloft. The horizontal grid is an Arakawa C grid. For integrating the equations, a third-order Runge–Kutta scheme with a smaller time step for acoustic and gravity wave modes is used.

For the computation of radiation within WRF, the Rapid and Accurate Radiative Transfer Model scheme (RRTM; Mlawer et al. 1997) was used for longwave radiation, and the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5; Dudhia 1989) for shortwave radiation. For surface layer physics, the Monin–Obukhov similarity theory scheme taken from the Hong and Pan (1996) PBL scheme [commonly referred to as the Medium Range Forecast model (MRF) scheme] was used. The Noah land surface scheme (Ek et al. 2003) was used, as was the Yonsei University (YSU) PBL scheme (Hong et al. 2006). For resolvable moist physics processes, we used the Lin et al. microphysics scheme (Lin et al. 1983). For representing the effects of subgrid-scale convective clouds, we used the Kain–Fritsch convective scheme (Kain and Frisch 1993; Kain 2004). We chose to use Kain–Fritsch (as opposed to not using any convective scheme) based on observing that without any convective scheme in place, convection was too extreme. We observed that with KF in place, convection triggers slightly earlier than if it were not in place. With convection triggering earlier, surface parcels do not become as unstable, and thus updraft velocities, rainfall rates, and surface pressure perturbations are not as large as if no convective scheme was in place. This is based on observations during the experimental phase, and no formal study of the benefits or drawbacks of including the Kain–Fritsch scheme in our 5-km resolution model. Such a study is beyond the scope of this work.

The WRF model was run twice daily, once for 0600 UTC initial conditions and the second time for 1800 UTC initial conditions. Forecasts were made out to 18 h after initialization time. The model resolution was 5 km, with a domain consisting of 121 × 121 grid boxes, centered at 26.8°N, 81.2°W. Figure 4 shows the area covered by our model domain, which is not the same as the area of our LAPS analyses: The WRF domain was within the LAPS domain. For all WRF model runs, the NAM 12-km model [at the time of the experiment, the NAM was known as the Eta Model; Black (1994); Janjić (1996)] was used to provide land or sea initial conditions and was used as the boundary conditions. Boundary conditions were updated every 3 h, using data from the NAM "tile" files obtained from the NCEP FTP site. The resolution of the NAM data was 12 km. The tiles used for this study are shown in Fig. 4. WRF runs in which the NAM (Eta) model was used for initial conditions will henceforth be referred to as WRF/NAM. WRF runs that instead used LAPS for initial conditions will be called WRF/LAPS. The title NAM/ETA will be used to denote the operational runs of the NAM model (which was, as stated earlier, the Eta Model during the summer of 2005).

c. Validation data

One means of evaluating model skill was to quantify the accuracy of precipitation forecasts. The River Forecast Centers (RFC) stage III precipitation data [a blend

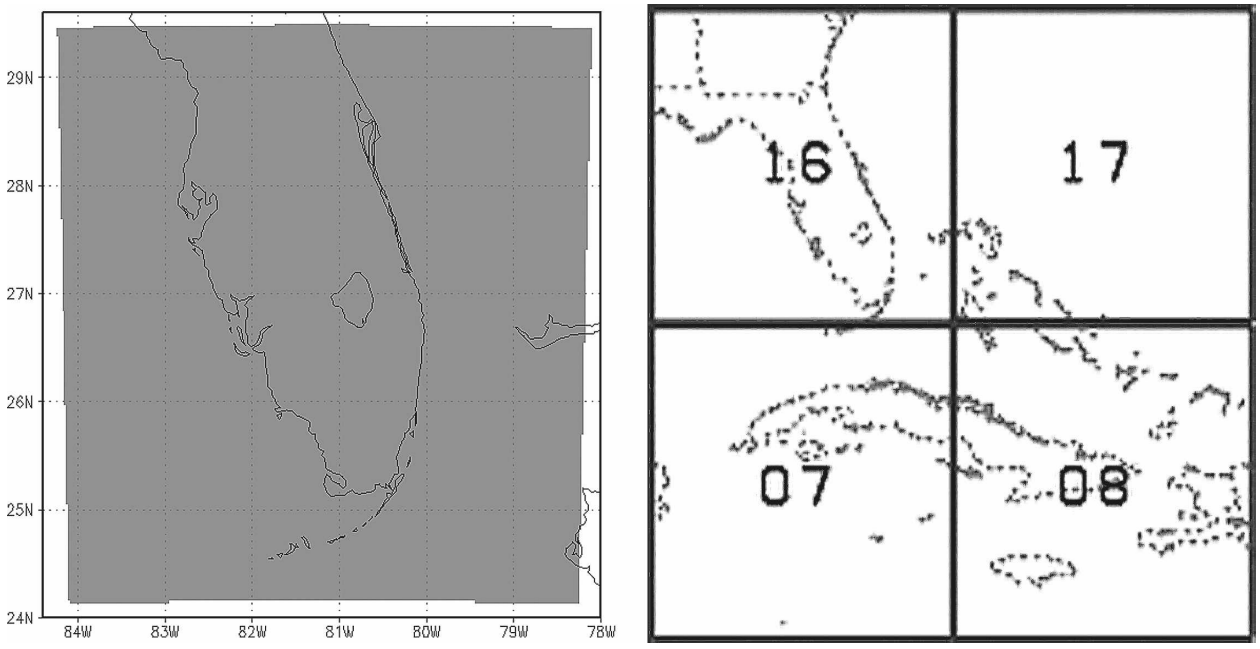

FIG. 4. (left) The area domain of the WRF model and (right) the NAM 12-km tile files (from NCEP grid 218) used as boundary and/or initial conditions.

of quality controlled gauge, radar, and satellite data; Breidenbach et al. (1998); McDonald and Graziano (2001)] were used as ground truth. These gridded data are retrievable by request from the National Precipitation Verification Unit (NPVU) Web site (information online at http://www.hpc.ncep.noaa.gov/npvu/) in gridded binary (GRIB) format. These rainfall totals were obtained for the study period at 3-hourly intervals. All precipitation forecasts and verifications were interpolated onto the grid of the 12-km NAM forecasts. These projections were made by averaging both the 4-km stage III data or the 5-km WRF forecast grid boxes over the area covered by the 12-km NAM/ETA forecast grid boxes. This averaging was done to ensure that the same amount of total precipitation makes it to the ground regardless of whether or not the higher-resolution precipitation fields are interpolated to the 12-km domain.

Model skill was also measured by comparing model output to surface observations. The value of a surface parameter, such as temperature, humidity, wind speed, or surface pressure, from an aviation routine weather report (METAR) observation is compared to the model value from the model grid box that contains the location of the observation. METAR observations used in this study had to be far enough from the water such that the model grid box was not a "water surface" grid box, as surface fluxes of temperature and moisture are much different over water than land, leading to significant biases in temperature and dewpoint forecasts. For this study, these 25 METAR observation sites were used for the evaluation of the WRF forecasts, and these sites are shown in Fig. 5.

3. Model evaluation

a. Qualitative case studies

Examples of the improvements possible from using LAPS analyses, rather than NAM, to initialize WRF are given in Fig. 6. In Fig. 6, radar indicates convection is on going at 1800 UTC on 6 June 2005 over the Florida Everglades. The 700-hPa relative humidity field from the WRF/NAM 1800 UTC analysis shows no such features. However, the WRF/LAPS 1800 UTC analysis does show high relative humidity levels (>90%) in the location of the convection. The LAPS analysis, having used satellite and radar data, had a more accurate representation of the convection, and thus the moisture fields, at 1800 UTC. The 1-h precipitation forecast made from both sets of initial conditions also shows that the WRF/LAPS initialization produced a more accurate forecast.

These sorts of improvements were common during our 2-month simulation. A more complete and objective evaluation of model performance, with respect to surface parameters, and with respect to precipitation, follows.

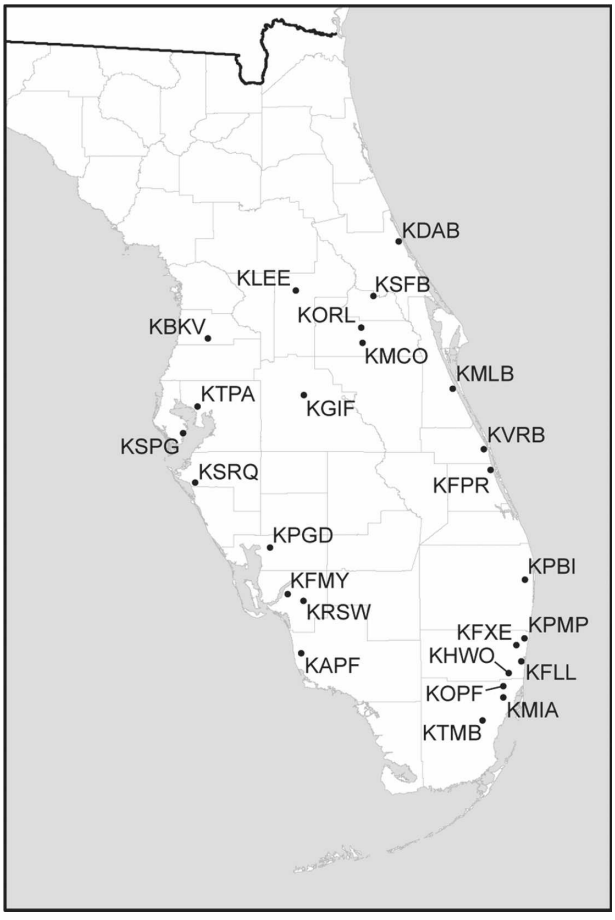

FIG. 5. The METAR observation sites used for comparison to model forecasts.

b. Comparison of surface parameters

As stated earlier, it is mesoscale phenomena, most notably the sea breeze, which drive convection in south Florida during the summertime. In an effort to better understand the impact of LAPS analyses on WRF forecasts, the differences in surface wind, moisture, temperature, and pressure fields from the different initializations were explored.

Figure 7 shows the composite difference between the 1-h forecasts made from WRF initialized with LAPS and WRF forecasts initialized by NAM, for forecasts initialized at 1800 UTC. This time of day, 1900 UTC, is when the sea breeze is typically quite active in Florida. The first notable difference between LAPS-initialized and NAM-initialized WRF forecasts is that those initialized by LAPS have lower 2-m temperatures. The differences, as shown in Fig. 7, are negative for every grid box. In addition, there is a pattern to these differences. Temperatures are cooler in all places, but while they are about 1°F cooler over the water, they are as much as 3°F cooler over the land. Given that this is 1900 UTC, 2-m air temperatures over the land will be warmer (around 90°F) than over the water (near 80°F). As a result, the difference in temperature between the air over the Florida Peninsula and the air over the surrounding water is not as great in LAPS-initialized WRF forecasts. With a weaker land–sea temperature gradient, LAPS-initialized WRF forecasts will have a weaker sea breeze than NAM-initialized forecasts.

Figure 7 also shows the differences in sea level pressure, and these differences also support a weakened sea breeze in WRF forecasts initialized using LAPS. Sea level pressure values over the peninsula are about 1 mb higher in LAPS-initialized forecasts than in NAM-initialized forecasts, but they are about the same as NAM-initialized forecasts over the water. This pressure gradient supports less (more) onshore (offshore) surface winds. The wind field differences, shown in Fig. 7, are consistent with the temperature and pressure field differences. The sea breeze (as measured by the zonal component of the wind, u) is weaker in WRF forecasts initialized from LAPS analyses.

The last panel in Fig. 7 shows the 2-m mixing ratio. WRF forecasts initialized with LAPS show higher values of q over the land (but smaller values of q over the water) than do the forecasts initialized with NAM. Different low-level moisture fields could impact upon the precipitation forecasts—the higher q values of the LAPS-initialized WRF forecasts resulting in higher values of the quantitative precipitation forecasts (QPFs).

Having noted these differences between 1-h forecasts, it is important to also know how the forecasts compare to actual observations. Figure 8 shows the average difference (the model bias) between either the WRF/LAPS or WRF/NAM forecasts and the METAR observations for the sites shown in Fig. 5. We evaluate the accuracy of the WRF forecasts for the 2-m temperature and dewpoint, the 10-m u component of the wind, and the sea level pressure. WRF forecast values for a site were obtained by using the values from the 5 km × 5 km grid box that contained the METAR site. No downscaling was done, as raw model output was used. As the WRF runs were all on the same domain, comparisons between different WRF forecasts have consistency. However, the NAM model, being on a different grid, would present inconsistencies. As a result, only the WRF hourly forecasts were compared to the METAR observations.

The biases (forecast − observations) of forecasts were calculated from fifty 0600 UTC runs and for fifty-six 1800 UTC runs. A jackknife bootstrap (Efron 1982), in which the average errors are estimated from all possible subsamples of the total dataset, was used to esti-

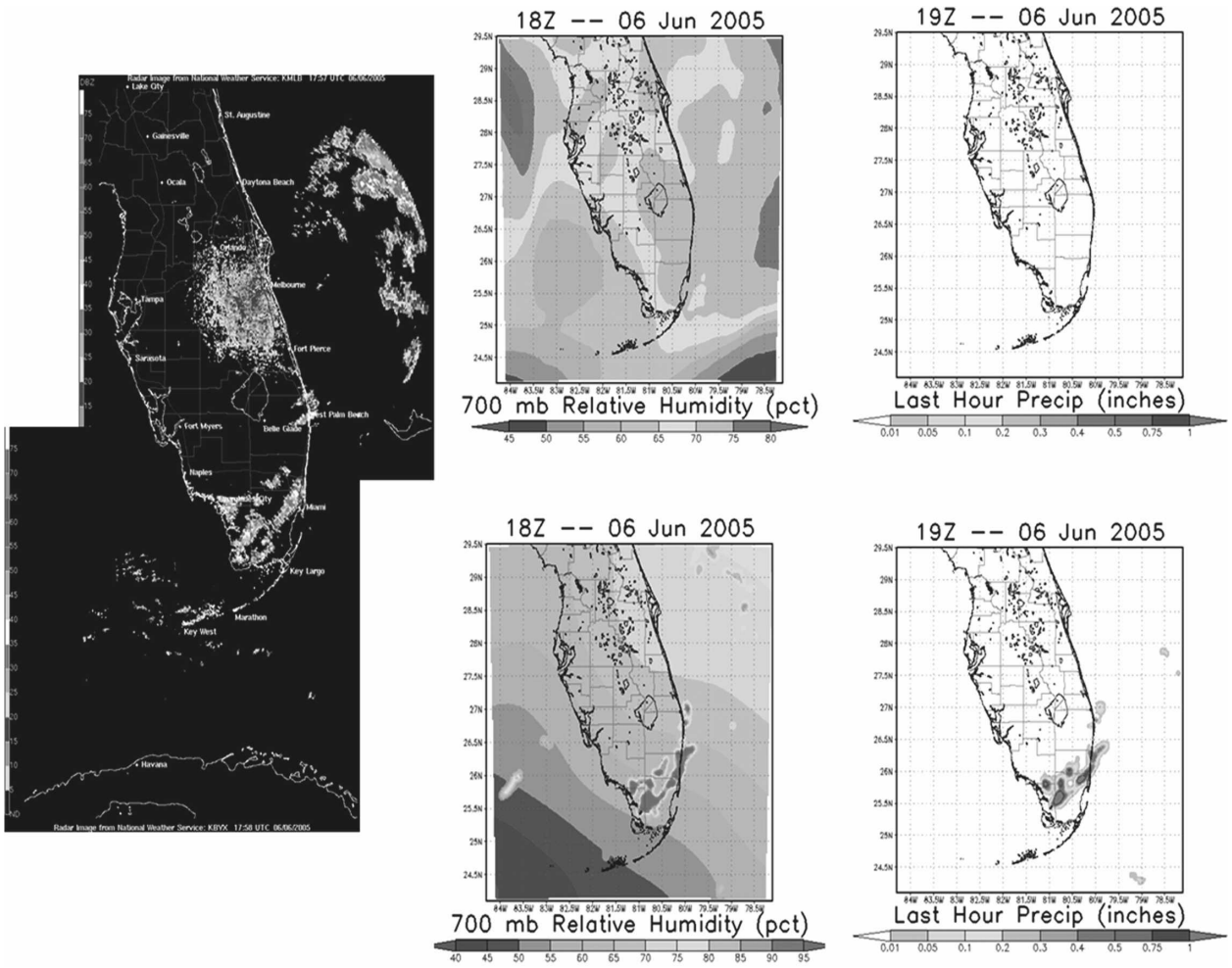

FIG. 6. (left) Radar mosaic (Key West and Melbourne radar sites; Miami was not available at this time frame) for 1800 UTC 6 Jun 2005. (right) Initial conditions for WRF simulations (top, WRF/NAM; bottom, WRF/LAPS). Also shown are the 1-h precipitation forecasts from (top right) WRF/NAM and (bottom right) WRF/LAPS.

mate uncertainty. Subsamples consisted of using all but one forecast; thus, each subsample consists of 49 cases for the 0600 UTC runs and 55 cases for the 1800 UTC runs. From the jackknife estimate of the average error, 95% confidence intervals of the average errors were constructed. Note that in Figs. 8 and 9, the error bars are often smaller than the icon (dot, box, etc.) representing each value. Thus, some of our error bars are beneath the icons and are not visible.

For both the 0600 and 1800 UTC initializations, WRF forecasts initialized by LAPS are generally more accurate in the first 6 h of the forecast than are the NAM-initialized WRF forecasts. For longer lead times, the difference between the two sets of forecasts is small. The average errors of both sets of forecasts converge with time. Given that both forecasts use the same boundary conditions (lateral and surface), which have an ever greater impact on longer lead time forecasts, it is not surprising that as lead time increases, the mean forecast errors become more similar.

While LAPS-initialized WRF forecasts in general have smaller biases, the exception is forecasts of 2-m temperature between 1200 and 1800 UTC for the 0600 UTC initialization runs. The cause of these errors is not clear to the authors—though it appears that the rate of warming in the model is not fast enough during the morning hours. However, despite this, it is clear from the results in Fig. 8 that the LAPS initialization results in more accurate forecasts in the first 6 h of the forecast.

Analyses of the mean squared errors, rather than the biases, are shown in Fig. 9. In general, LAPS-initialized WRF forecasts are more accurate than the NAM-initialized WRF forecasts. The magnitude of the differences in accuracy varies considerably. Wind forecast

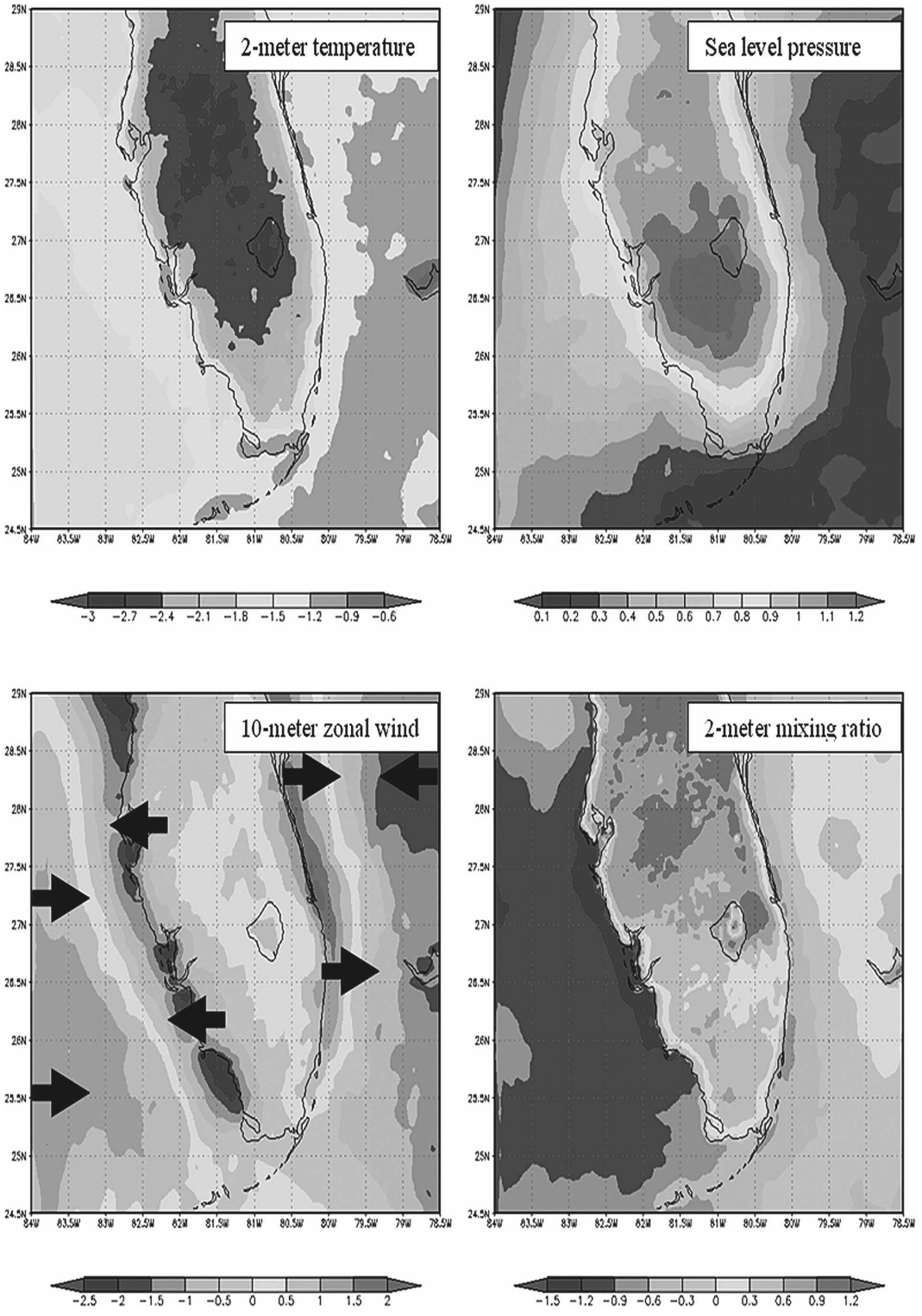

FIG. 7. Composite of the differences (WRF/LAPS − WRF/NAM) between the 1-h forecasts (valid at 1900 UTC) of 2-m temperature (°F), sea level pressure (mb), 10-m zonal wind (m s^{-1}), and mixing ratio (g kg^{-1}). Composites are from 50 different 1-h forecasts. All forecasts are initialized at 1800 UTC.

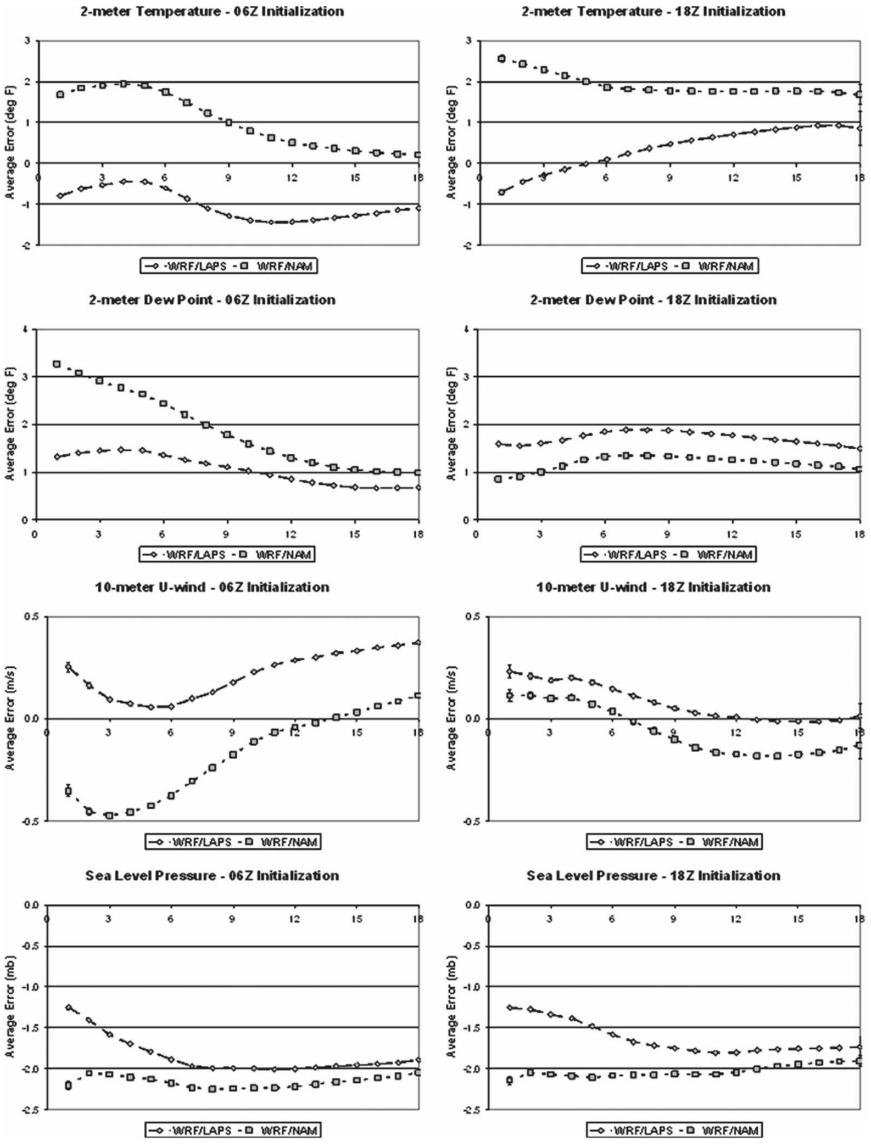

FIG. 8. Comparison of the grid box values of (top) 2-m temperature and (second row) dewpoint, (third row) 10-m u component of the wind, and (bottom) sea level pressure for the (left) 0600 and (right) 1800 UTC cycles to METAR observations for 25 sites in FL within the model domain, as shown in Fig. 6. The x axis (h) since initialization and the y-axis errors for temperature and dewpoint (F°), wind (m s^{-1}), and sea level pressure (mb). Error bars represent 95% confidence intervals, with uncertainty determined from a jackknife estimate (Efron 1982).

MSEs are about the same at all times, whereas there are significant differences in the MSEs of the temperature and dewpoint forecasts. Most of the improvements in the LAPS-initialized WRF forecasts occur during the first 6 h of a model forecast. For longer lead times, the two sets of forecasts are of nearly equal accuracy.

The accuracy of the sea-breeze forecasts can, to some extent, be quantified using the mean squared errors of wind and pressure. In the midday to afternoon hours when the sea breeze is starting, that is, 1500–1800 UTC (9–15-h forecasts of the 0600 UTC WRF runs), the WRF forecasts initialized from LAPS have more accurate pressure and wind forecasts, as shown in Fig. 9. Additionally, Fig. 8 shows that in addition to smaller MSE, the bias of the WRF/LAPS forecasts is smaller.

c. *Precipitation forecasts*

Summertime rain in south Florida is convective and cellular in nature, and locally heavy rain is possible with

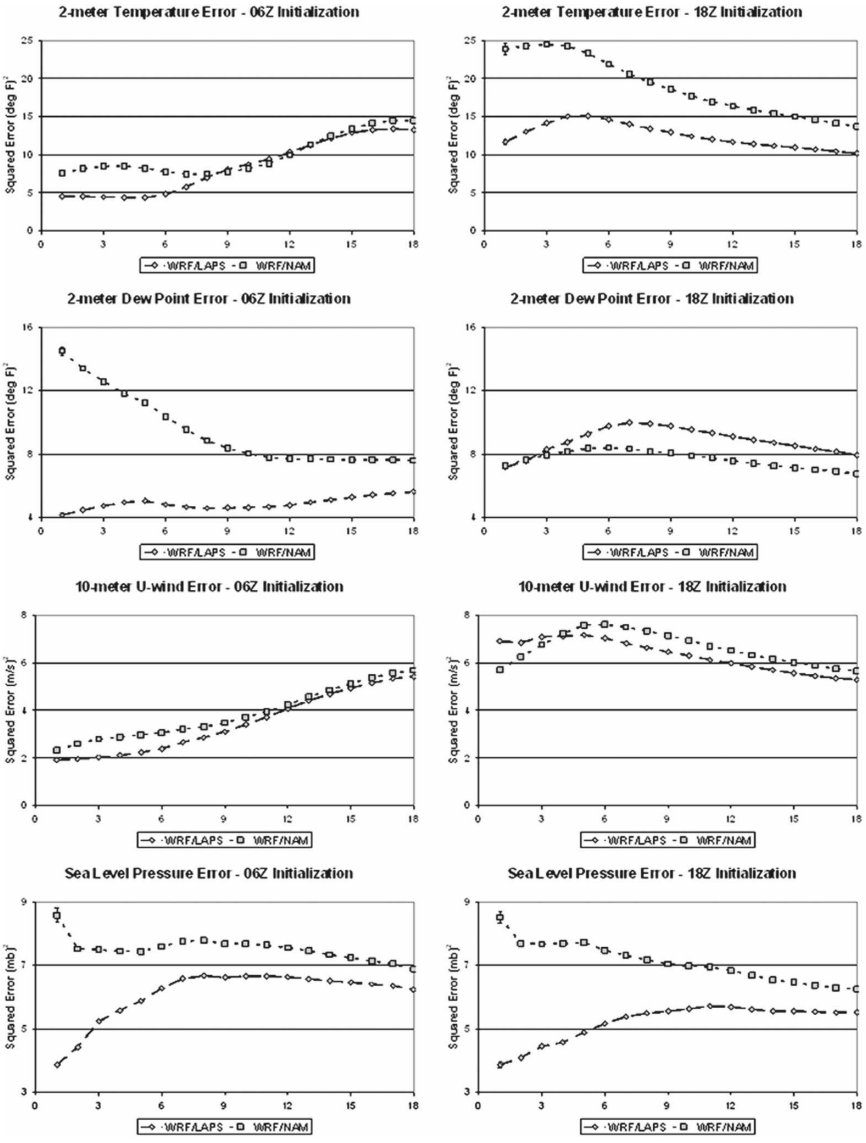

FIG. 9. Comparison of the grid-box values of (top) 2-m temperature and (second row) dewpoint, (third row) 10-m u component of the wind, and (bottom) sea level pressure for the (left) 0600 and (right) 1800 UTC cycles to METAR observations for 25 sites in Florida within the model domain, as shown in Fig. 6. The x axis (h) since initialization and the y-axis squared errors for temperature and dewpoint ($^{\circ}F^2$) , wind ($m^2\ s^{-2}$), and sea level pressure (mb^2). Error bars represent 95% confidence intervals, with uncertainty determined from a jackknife estimate (Efron 1982).

any cell that develops. We chose to evaluate our WRF forecast by looking at small scales and moderate to heavy precipitation amounts. Our model evaluation consists of grid-box-based calculations of the threat scores (TS), probability of detection (POD), false alarm ratios (FAR), and mean error (BIAS). These scores were computed for two precipitation thresholds: 0.10 and 0.25 in. of precipitation in a 3-h period. We do not intend to suggest these thresholds to be the standard in assessing model skill, but for our experiment, these thresholds were roughly in the middle of our distribution of precipitation amounts. Below are the equations we used to produce our metrics, all of which are discussed in greater detail in Wilks (1995).

Given an area forecast (A_f) of precipitation, an area observed (A_o) of precipitation, and the area over which both of these intersect, referred to as the area correct (A_c), the threat score (TS) is defined as

$$\mathrm{TS} = \frac{A_c}{A_f + A_o - A_c}. \quad (1)$$

The smaller the threat scores, the less skill in the forecast. If the area forecast and area observed are identical, then $A_c = A_f = A_o$, and the threat score is 1. If the forecast and observed areas are the same size, and half overlap, then $A_f = A_o = 1$, $A_c = 0.5$, and TS = 1/3.

The probability of detection (POD) is defined as

$$\mathrm{POD} = \frac{A_c}{A_o}, \quad (2)$$

where A_c is the number of observed rainy grids that were forecast and A_o is the total number of observed rainy grids. Ideally, one would like a high POD. A POD of 1.0 would mean that every grid box that received rain was accurately forecast. The primary difference between POD and TS is that POD does not penalize for overforecasted precipitation. Finally, false alarm ratio (FAR) is defined as

$$\mathrm{FAR} = \frac{A_f}{A_f + A_c}. \quad (3)$$

The bias score is simply the average of the difference between the model forecasts and the observed values over all grid boxes. In mathematical form, the bias score for N number of grid boxes is

$$\mathrm{BIAS} = \frac{1}{N}\sum_{i=1}^{N}(M_i - R_i), \quad (4)$$

where M_i and R_i are the model forecasts and observed precipitation amounts at each grid box, respectively. For the mean square error, the formula is

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(M_i - R_i)^2. \quad (5)$$

These quantities, TS, POD, FAR, BIAS, and MSE, were calculated for each of the model configurations described in the previous section: WRF initialized with LAPS (WRF/LAPS), WRF initialized with NAM (WRF/NAM), and the NAM/ETA model forecasts. All precipitation quantities (verification, WRF, and NAM) were evaluated on the coarsest grid, that of the NAM/ETA model. For the WRF output and the verification data, we interpolated onto the coarser grid. While this choice does undo some of the benefits of running WRF at high resolution, evaluating precipitation forecasts on this coarse grid does give the most fair comparison.

Statistics were calculated for both the 0600 and 1800 UTC runs, separately, and averaged over the study time. The scores were calculated from 106 model runs: fifty 0600 UTC runs and fifty-six 1800 UTC runs. For each model cycle, the statistics were stratified into periods for the 0.10- and 0.25-in. precipitation thresholds. In the 0600 UTC cycle, the periods are 0600–0900, 0900–1200, 1200–1500, 1500–1800, 1800–2100, and 2100–0000 UTC. For the 1800 UTC cycle, there periods were 1800–2100, 2100–0000, 0000–0300, 0300–0600, 0600–0900, and 0900–1200 UTC. The uncertainties of all of the metrics were estimated using a jackknife approach, as was done in the previous section.

Figure 10 shows the TS for the experimental period broken down by model cycle (rows) and thresholds (columns). In all cases, the WRF/LAPS forecasts have larger threat scores than do the WRF/NAM forecasts. For the 0600 UTC runs, with a 0.10-in. threshold, the TSs in the first 6 h for the WRF/LAPS forecasts are about double those for the WRF/NAM and NAM/ETA forecasts. For all initialization times and thresholds, the WRF/LAPS forecasts are most accurate during the early part of the integration. However, the greater accuracy relative to WRF/NAM and NAM/ETA seem to last longer in the 0600 UTC runs than the 1800 UTC runs.

While WRF/LAPS fares better than WRF/NAM for both precipitation thresholds at both initialization times, it does not consistently outperform the NAM/ETA forecasts. For the larger threshold (0.25 in. of rain in 3 h) the WRF/LAPS outperformed the NAM/ETA at all times for the 0600 UTC initialization (with the exception of the 1200–1500 UTC time period, where the error bars overlap), but only for shorter lead times (1800–0000 UTC, the first 6 h) for the 1800 UTC initialization. For the smaller threshold (0.10 in. of rain in 3 h) NAM/ETA always has the highest TS for the 1800 UTC initialization, and is better at the longer lead times for the 0600 UTC initialization. The NAM/ETA threat scores were always highest during the 1800–0000 UTC time periods, the time when convection is most active, and thus the times when precipitation forecasts are most important. Compared to NAM/ETA, WRF/LAPS did well with larger precipitation amounts, and for the 0600 UTC initialization time forecasts. The LAPS hot start was not as beneficial when the initialization time was 1800 UTC, and WRF was not as helpful for QPFs of lighter amounts.

Figure 11 shows the POD scores for the experimental period broken down by model cycle (rows) and thresholds (columns). WRF/LAPS consistently outperforms WRF/NAM. The better performance of the WRF/LAPS over the other configurations is most dramatic in the 0600 UTC model cycle, and early on in the integration. The PODs from WRF/LAPS are always greater than for NAM/ETA for all lead times for the 0.25-in. in

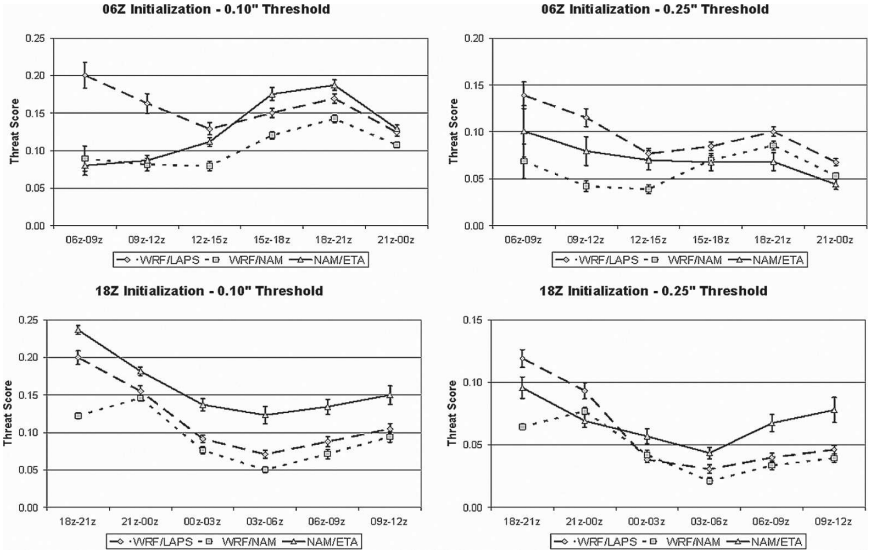

FIG. 10. Threat scores for the three model configurations evaluated for the (top) 0600 and (bottom) 1800 UTC model cycles and the (left) 0.10- and (right) 0.25-in. precipitation thresholds. WRF/LAPS denotes the LAPS-initialized runs, WRF/NAM the NAM12-initialized runs, and NAM/ETA the NCEP NAM12 runs. Error bars represent 95% confidence intervals, with uncertainty determined from a jackknife estimate (Efron 1982).

3-h threshold, but the reverse is nearly true for the 0.10-in. in 3-h threshold. In comparison to NAM/ETA, WRF/LAPS is more accurate in forecasting heavy rather than light precipitation.

The FAR scores are shown in Fig. 12. With the exception of the first 3 h of the forecast, for lighter precipitation amounts, WRF/LAPS forecasts consistently have fewer false alarms than do WRF/NAM forecasts. However, NAM/ETA forecasts almost always have the fewest false alarms of the three models. This means that WRF tends to create more false rainy grid boxes than NAM/ETA with the exception of the 0.10-in. precipitation threshold early on in the integration. This implies that, in general, WRF has greater area coverage of

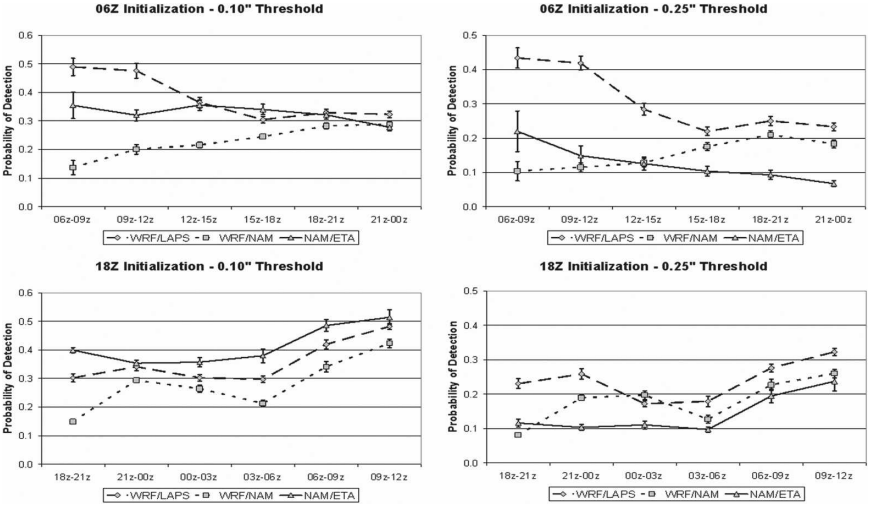

FIG. 11. POD scores for the three model configurations evaluated for the (top) 0600 and (bottom) 1800 UTC model cycles and the (left) 0.10- and (right) 0.25-in. precipitation thresholds. WRF/LAPS denotes the LAPS-initialized runs, WRF/NAM the NAM12-initialized runs, and NAM/ETA the NCEP NAM12 runs. Error bars represent 95% confidence intervals, with uncertainty determined from a jackknife estimate (Efron 1982).

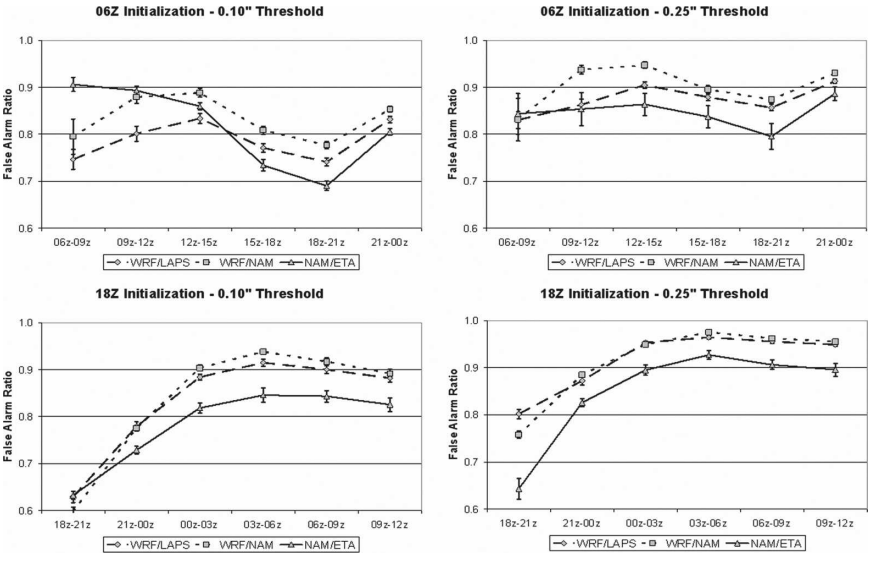

FIG. 12. FAR scores for the three model configurations evaluated for the (top) 0600 and (bottom) 1800 UTC model cycles and the (left) 0.10- and (right) 0.25-in. precipitation thresholds. WRF/LAPS denotes the LAPS-initialized runs, WRF/NAM the NAM12-initialized runs, and NAM/ETA the NCEP NAM12 runs. Error bars represent 95% confidence intervals, with uncertainty determined from a jackknife estimate (Efron 1982).

rainy grid boxes than does NAM/ETA, especially for larger precipitation amounts.

Figure 13 shows the bias scores for all three model configurations for both the 0600 and 1800 UTC cycles. Overall, NAM/ETA has the smallest bias, and the WRF forecasts have a higher bias. This is consistent with the high FAR of the WRF forecasts: More false rainy grid boxes implies a greater total precipitation amount. Note that in the first 3 h of the forecast WRF/NAM has a low bias, whereas WRF/LAPS has a high bias. It is likely that the smaller humidity values of the NAM analyses result in less convection taking place during the first 3 h, where as the more moist LAPS analyses are more likely to produce more rainfall. Also, LAPS analyses have convection active at the initial time; NAM initial conditions require WRF to spin up the convection. As with most measures of forecast difference and accuracy used in this study, the differences in bias between WRF/NAM and WRF/LAPS become nearly negligible after the first 6 h of the forecast.

Figure 13 shows low bias for NAM/ETA throughout the forecast period. This raises questions about the biases in the precipitation forecast distribution. Figure 14 shows both a histogram analysis of the precipitation ranges for all model configurations tested plus the NPVU rain dataset (which is assumed to be ground truth—interpolated to the 12-km NAM/ETA domain—in this study) and the ratio of the area forecast to the area observed. For light precipitation (less than 0.25 in. in 3 h) amounts, the NAM/ETA forecasts overestimate

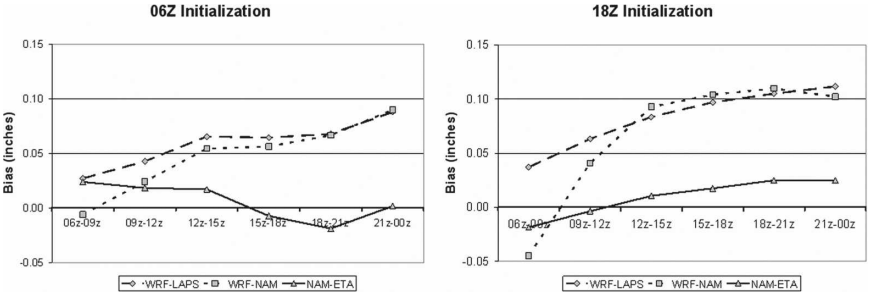

FIG. 13. Bias scores for the three model configurations evaluated for the (left) 0600 and (right) 1800 UTC model cycles. WRF/LAPS denotes the LAPS-initialized runs, WRF/NAM the NAM12-initialized runs, and NAM/ETA the NCEP NAM12 runs.

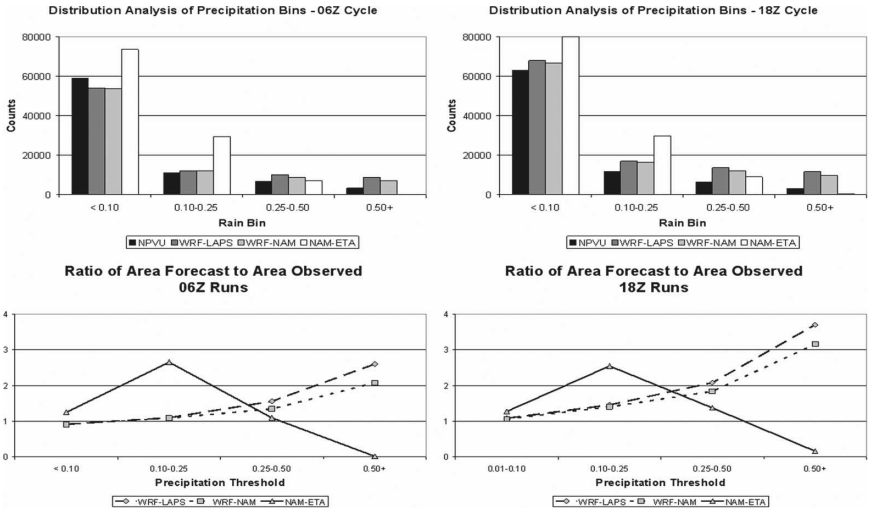

FIG. 14. Histograms of precipitation in 3-h period ranges (bins) on the x axis for the three model configurations tested plus the precipitation data used as ground truth (NPVU) along the top, ratios of area forecast to area observed for the same precipitation ranges are shown along the bottom. The (left) 0600 and (right) 1800 UTC initialized forecasts. The <0.10 bin does not include 0 rain bins. WRF/LAPS denotes the LAPS-initialized runs, WRF/NAM the NAM12-initialized runs, and NAM/ETA the NCEP NAM12 runs.

the area coverage, but WRF forecasts have nearly the correct area coverage. For heavier rainfall amounts (greater than 0.25 in. in 3 h), WRF forecasts overestimate the area coverage, and NAM/ETA rarely predicts 3-h rainfall amounts greater than 0.5 in.

WRF forecasts, regardless of initialization, vastly overpredict the areas of heavy precipitation. This overprediction of heavy rains is a likely cause for the higher bias of the WRF forecasts. The higher bias of the WRF runs appears to be a product of the combination of convective parameterization (KF) and moist physics schemes (Lin et al. 1983) used, which produces excessive rainfall amounts. This overprediction likely leads to better TS forecasts for WRF, as WRF has more chances for a "hit." The authors believe that the smaller bias of the NAM/ETA is related the Betts–Miller–Janjić (BMJ; Betts and Miller 1986; Janjić 1994) convective parameterization scheme. BMJ, used in NAM/ETA, tends to create large areas of light to moderate precipitation, but does not produce heavy precipitation amounts. This results in greater area forecasts of lesser precipitation amounts, but a smaller total amount of precipitation forecast. Given that NAM/ETA rarely forecasts heavy precipitation, there is less opportunity to overforecast and, hence, have a high bias. However, large areas of light to moderate precipitation are not necessarily representative of the cellular characteristics of south Florida summertime convection. Thus, there is a tendency to overforecast the area coverage of precipitation amounts between 0.10 and 0.25 in. in 3 h, as indicated in Fig. 14 (NAM/ETA has about 2.5 times the number of rainy grid boxes than does the NPVU verification for 0.10–0.25 in. in 3 h). This feature of the BMJ convective parameterization has also been documented in recent studies (Jankov et al. 2005; Etherton and Santos 2006).

4. Summary and conclusions

This study presents results of an experiment conducted during the summer of 2005 across south Florida to measure the impact of locally initializing the WRF (ARW core) model with LAPS. The impact of the LAPS initialization was measured by comparing forecasts of surface parameters such as temperature, specific humidity, wind, and sea level pressure between LAPS-initialized WRF forecasts (WRF/LAPS) and NAM-initialized forecasts (WRF/NAM) to surface observations from METARS. Additionally, precipitation forecast skill levels were compared for three different model configurations: WRF/LAPS, WRF/NAM, and NAM/ETA using threat scores, the probability of detection, the false alarm ratio, and bias scores. A total of 106 model runs (50 from the 0600 UTC model cycle and 56 from the 1800 UTC model cycle) were used as part of the analysis. The model ran out to 18 h with precipitation skill statistics broken down by 3-hourly intervals, by model cycles (0600 and 1800 UTC), and by precipitation thresholds (namely 0.10 and 0.25 in.).

Results indicate that in general the LAPS initializa-

tions result in more accurate WRF forecasts of surface parameters, particularly during the first 6 h of the integration. WRF/LAPS precipitation forecasts were also more accurate than those of WRF/NAM. Our results support those of Jankov et al. (2007), in which WRF initialized from LAPS had higher equitable threat scores than did WRF initialized by NAM/ETA. The benefits of using LAPS to hot start WRF were most significant early in the forecast period, and, for longer lead times, the impacts of initialization waned. Jankov et al. (2007) have similar results.

WRF/LAPS precipitation forecasts are more accurate than those of WRF/NAM. WRF/LAPS precipitation forecasts are also consistently more accurate than NAM/ETA during the short range (first 6 h of the integration or so) with the exception of low precipitation thresholds with the 1800 UTC cycle where NAM/ETA exhibited more accuracy. WRF/LAPS forecasts are more accurate than NAM/ETA when both are initialized at 0600 UTC, but less accurate than NAM/ETA when both are initialized at 1800 UTC with the exception of early on in the forecast (within 6 h or so) for the higher precipitation thresholds (greater than 0.25 in.). Thus, WRF/LAPS, in general, fares better than NAM/ETA when predicting higher amounts of precipitation and when forecasts are initialized at night rather than in midafternoon.

These results suggest that agencies that produce high resolution guidance should consider incorporating local high-resolution initialization. The NAM forecasts–analyses are designed to represent as best as possible the state of the atmosphere across the continental United States and into the Atlantic and Pacific Oceans. As shown earlier, there are features present in LAPS analyses that were not present in NAM fields in 2005. Analyses at a much higher resolution were not feasible across the entire continental United States at the time of our experiment. However, NCEP now produces a real-time mesoscale analysis (RTMA; DePondeca et al. 2007) at 5-km grid spacing, for some surface variables. Perhaps, in time, this analysis will be performed for the entire 3D volume of the atmosphere, and could then be used to initialize a mesoscale model. For now, the local analyses and local forecasts complement the analyses and forecasts produced at NCEP. Our results are in general agreement with those of Bogenschutz et al. (2004), who noted that NAM/ETA did do well for forecasts of lesser amounts of precipitation, but that WRF/LAPS did better for higher precipitation thresholds. Additionally, they also found WRF/LAPS's skill decreased with time, with a peak in skill at a lead time of 6 h. Szoke et al. (2004) also had early forecast improvements resulting from a hot-start scheme. However, they found that after the first hour of the forecast the improvements waned, as the model "lost" the convection—particularly of elevated convection. They suggest possible modifications to the balances used by LAPS in making analyses that would allow WRF to hold onto convection in the LAPS analysis longer. Casual observations during the experimental phase of this study support this. The authors observed that often the WRF/LAPS initial wind field was not representative of the raw LAPS wind analysis, resulting in circulations that were not in balance with the depicted convection in the moisture field. The authors believe this is the result of the quasigeostrophic balancing imposed on the LAPS analysis prior to creating the diabatic initialization grids. This is a topic worth investigating in future studies.

The differences between forecasts initialized using LAPS and those initialized from NAM tiles waned over time. Further improvements to local high-resolution forecasts will require improvements in local boundary conditions. While the atmospheric boundary conditions are not likely to be improved at the local level, the surface boundary conditions may be. Improvements in the surface characteristics (sea surface temperature, land surface temperature, and soil moisture, etc.) can all contribute to improving the accuracy of locally produced forecasts beyond the first 12 h. Future work should concentrate on testing and/or exploiting this potential in order to develop more accurate high-resolution short-term forecast guidance locally.

Acknowledgments. The authors thank the Cooperative Program for Operational Meteorology, Education and Training (COMET), which funded this research (COMET Outreach Project S04-44694). We thank the Cartography Laboratory of the University of North Carolina at Charlotte for their assistance in preparing figures for this manuscript. We appreciate Steve Lazarus providing help with our experimental design. Last, but not least, we thanks the three anonymous reviewers who helped make this paper clearer and more scientifically rigorous.

REFERENCES

Albers, S., 1995: The LAPS wind analysis. *Wea. Forecasting,* **10,** 342–352.

——, J. McGinley, D. Birkenheuer, and J. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analyses of clouds, precipitation, and temperature. *Wea. Forecasting,* **11,** 273–287.

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.,* **112,** 693–709.

Birkenheuer, D., 1999: The effect of using digital satellite imagery in the LAPS moisture analysis. *Wea. Forecasting,* **14,** 782–788.

Black, T. L., 1994: The New NMC mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Bogenschutz, P., and Coauthors, 2004: Summer season verification of the first NWS operational WRF Model forecasts from the NOAA Coastal Storms Initiative Project in northeast Florida. Preprints, *17th Conf. on Probability and Statistics/20th Conf. on Weather and Forecasting/16th Conf. on Numerical Weather Prediction,* Seattle WA, Amer. Meteor. Soc., J12.1. [Available online at http://ams.confex.com/ams/pdfpapers/70493.pdf.]

Breidenbach, J. P., D.-J. Seo, and R. Fulton, 1998: Stage II and III post processing of NEXRAD precipitation estimates in the modernized Weather Service. Preprints, *14th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* Phoenix, AZ, Amer. Meteor. Soc., 263–266.

De Pondeca, M. S. F. V., and Coauthors, 2007: The development of the real time mesoscale analysis system at NCEP. Preprints, *23rd Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* San Antonio, TX, Amer. Meteor. Soc., P1.10. [Available online at http://ams.confex.com/ams/87ANNUAL/techprogram/paper_118813.htm.]

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

Efron, B., 1982: *The Jackknife, the Bootstrap, and other Resampling Plans.* Society for Industrial and Applied Mathematics, 92 pp.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Etherton, B., and P. Santos, 2006: The effects of using AWIPS LAPS to locally initialize the Workstation Eta. *Natl. Wea. Assoc. Dig.,* **30,** 49–60.

Hiemstra, C. A., G. E. Liston, R. A. Pielke, D. L. Birkenheuer, and S. C. Albers, 2006: Comparing Local Analysis and Prediction System (LAPS) assimilations with independent observations. *Wea. Forecasting,* **21,** 1024–1040.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.,* **124,** 2322–2339.

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

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

——, 1996: The Mellor–Yamada level 2.5 scheme in the NCEP Eta Model. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 333–334.

Jankov, I., W. A. Gallus Jr., M. Segal, B. Shaw, and S. E. Koch, 2005: The impact of different WRF Model physical parameterizations and their interactions on warm season MCS rainfall. *Wea. Forecasting,* **20,** 1048–1060.

——, ——, ——, and S. E. Koch, 2007: Influence of initial conditions on the WRF–ARW Model QPF response to physical parameterization changes. *Wea. Forecasting,* **22,** 501–519.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.,* **43,** 170–181.

——, and J. M. Frisch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.,* No. 46, Amer. Meteor. Soc., 165–170.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.,* **22,** 1065–1092.

McDonald, B. E., and T. M. Graziano, 2001: The National Precipitation Verification Unit (NPVU): Operational implementation. Preprints, *Conf. on Precipitation Extremes: Prediction, Impacts, and Responses,* Albuquerque, NM, Amer. Meteor. Soc., P1.32. [Available online at http://ams.confex.com/ams/annual2001/techprogram/paper_17535.htm.]

McGinley, J. A., 2001: Toward a surface data continuum: Use of the Kalman filter to create a continuous, quality controlled surface data set. Preprints, *18th Conf. on Weather Analysis and Forecasting,* Fort Lauderdale, FL, Amer. Meteor. Soc., P2.18. [Available online at http://ams.confex.com/ams/pdfpapers/23059.pdf.]

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.,* **102D,** 16 663–16 682.

Schultz, P., and S. Albers, 2001: The use of three-dimensional analyses of cloud attributes for diabatic initialization of mesoscale models. Preprints, *18th Conf. on Weather Analysis and Forecasting/14th Conf. on Numerical Weather Prediction/9th Conf. on Mesoscale Processes,* Fort Lauderdale, FL, Amer. Meteor. Soc., JP2.5a. [Available online at http://ams.confex.com/ams/WAF-NWP-MESO/techprogram/paper_23324.htm.]

Shaw, B. L., J. A. McGinley, and P. Schultz, 2001: Explicit initialization of clouds and precipitation in mesoscale forecast models. Preprints, *14th Conf. on Numerical Weather Prediction,* Fort Lauderdale, FL, Amer. Meteor. Soc., J2.6a. [Available online at http://ams.confex.com/ams/pdfpapers/23438.pdf.]

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF version 2. NCAR Tech Note NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307; also available online at http://box.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf.]

Szoke, E. J., J. M. Brown, and B. Shaw, 2004: Examination of the performance of several mesoscale models for convective forecasting during IHOP. Preprints, *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction/17th Conf. on Probability and Statistics in the Atmospheric Sciences,* Seattle, WA, Amer. Meteor. Soc., J13.6. [Available online at http://ams.confex.com/ams/pdfpapers/68930.pdf.]

Wilks, D., 1995: *Statistical Methods in Atmospheric Sciences.* 2nd ed. Elsevier Press, 648 pp.