Model Bias in a Continuously Cycled Assimilation System and Its Influence on Convection-Permitting Forecasts

GLEN S. ROMINE, CRAIG S. SCHWARTZ, CHRIS SNYDER, JEFF L. ANDERSON, AND MORRIS L. WEISMAN
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 10 April 2012, in final form 10 October 2012)

ABSTRACT

During the spring 2011 season, a real-time continuously cycled ensemble data assimilation system using the Advanced Research version of the Weather Research and Forecasting Model (WRF) coupled with the Data Assimilation Research Testbed toolkit provided initial and boundary conditions for deterministic convection-permitting forecasts, also using WRF, over the eastern two-thirds of the conterminous United States (CONUS). In this study the authors evaluate the mesoscale assimilation system and the convection-permitting forecasts, at 15- and 3-km grid spacing, respectively. Experiments employing different physics options within the continuously cycled ensemble data assimilation system are shown to lead to differences in the mean mesoscale analysis characteristics. Convection-permitting forecasts with a fixed model configuration are initialized from these physics-varied analyses, as well as control runs from 0.5° Global Forecast System (GFS) analysis. Systematic bias in the analysis background influences the analysis fit to observations, and when this analysis initializes convection-permitting forecasts, the forecast skill is degraded as bias in the analysis background increases. Moreover, differences in mean error characteristics associated with each physical parameterization suite lead to unique errors of spatial, temporal, and intensity aspects of convection-permitting rainfall forecasts. Observation bias by platform type is also shown to impact the analysis quality.

1. Introduction

Forecasting squall lines, supercell thunderstorms, and other forms of organized moist convection is a problem at the forefront of numerical weather prediction (NWP; Fritsch and Carbone 2004). At the National Center for Atmospheric Research (NCAR), experimental forecasts with the Weather Research and Forecasting Model (WRF; Skamarock et al. 2008) have been performed every spring since 2003, both in support of field experiments [the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX; Davis et al. 2004); the Second Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX2; Wurman et al. 2012) and as part of the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT; e.g., Kain et al. 2006, 2008; Clark et al. 2012)]. A key finding from NCAR’s experimental forecasts is that initial conditions (ICs) often have a larger influence on short-term (0–36 h) convective-forecast skill than does the model configuration (Weisman et al. 2008). Because of this result, which points to the potential gains to be made by improving ICs for convective-scale forecasts, our 2011 experimental forecasts began from ICs produced by continuously cycling data assimilation with WRF and an ensemble Kalman filter. In this paper, we document the performance of the 2011 forecasts, with an emphasis on precipitation, and explore the factors that influence that performance and the quality of the analyses. We find that skill in these convective-scale forecasts is limited in significant part by systematic bias in the initial conditions, which in turn, in our continuously cycled system, arises from bias in WRF. We further show that reducing the model bias in the cycled analysis system improves ICs by examining forecast skill of convective precipitation.

Several efforts in convection-permitting NWP (CP; horizontal grid spacing of 4 km or less), both operational and experimental, have commenced in the last decade. This is motivated in part by evidence that resolution
sufficient to allow explicit representation of convection improves simulation of both the convection’s diurnal characteristics (Done et al. 2004; Liu and Moncrieff 2007) and its mode and structure (Clark et al. 2007; Kain et al. 2008; Schwartz et al. 2009; Sobash et al. 2011). The fine grid resolution and accompanying computational demand require limited-area forecast models (LAMs) and thereby lead to associated challenges for initial and boundary conditions.

Global analysis systems are based on continuous cycling data assimilation (CCDA), where skillful background forecasts lead to improvements in analyses, which in turn lead to further improvement in the background forecasts. LAM systems often employ alternative approaches, such as simply interpolating an external, lower-resolution analysis or forecast to the high-resolution grid, or using such an external analysis as the background field for high-resolution data assimilation at each assimilation time, or partial cycling, in which the data-assimilation system again starts from an external analysis but then cycles for a limited period (often 6 or 12 h). As examples, Kong et al. (2009) employ an external analysis as the background field for a three-dimensional variational assimilation of radar observations, while partial cycling is used operationally in NOAA’s Rapid Refresh system (S. Benjamin 2012, personal communication), NOAA’s North American Mesoscale Model (NAM) system (Rogers et al. 2009), and at Environment Canada (Fillion et al. 2010). The use of partial cycling in LAM systems was driven by improved performance statistics with this approach, despite the expectation that CCDA should perform better. The reasons for this difference between expectations and outcomes are not yet understood.

Ensemble Kalman filter (EnKF) data assimilation systems have already been demonstrated in a variety of real-time applications (e.g., Torn and Hakim 2008; Torn 2010; Ancell et al. 2011; Zhang et al. 2011b). For our application here to the convective-forecast problem, we employ the WRF and the EnKF as implemented in the Data Assimilation Research Testbed (DART; Anderson et al. 2009), a system that we will hereafter term WRF-DART. The analysis system was cycled continuously in real time over a 47-day period in spring 2011 and CP forecasts were initialized once daily from a single member analysis state. This long period of continuous cycling revealed significant biases in the analyses and the subsequent CP forecasts had limited skill, being noticeably inferior to forecasts initialized from the Global Forecast System (GFS) analysis, except at ranges less than 6 h. Similarly configured retrospective WRF-DART systems, but with varying physics, were also examined to assess sensitivities of the analysis bias to the chosen physics, and the dependence of the CP forecasts (with fixed physical parameterizations) on the analysis bias. While several studies have examined physics sensitivity in CP forecasts (e.g., Clark et al. 2008; Weisman et al. 2008; Gebhardt et al. 2011), so far only Torn and Davis (2012) have reported specifically on physics sensitivity in CCDA. Secondary objectives of this paper are therefore to illustrate how moderate bias in the LAM can degrade a CCDA system and to demonstrate that, despite this, CCDA offers a path to identify these model deficiencies, improve the model and thereby improve assimilation and forecast performance for LAMs.

The method of data assimilation is not the focus of this work, and is it likely to be crucial to the results presented here. Indeed, we expect that most reasonable assimilation systems, if continuously cycled, would exhibit a similar sensitivity to model bias for convective-scale forecasts. Our study leverages an EnKF assimilation scheme based on several properties that are appealing for convective-scale NWP in the longer term: (i) the EnKF provides an ensemble of initial conditions for ensemble forecasts, which are of particular interest at meso- and convective scales (Stensrud et al. 2009); (ii) it allows parameterizations that are uncertain, such as microphysics, to be easily changed or replaced, since no tangent-linear or adjoint model is required; (iii) it yields flow-dependent forecast covariances, such as between Doppler radial velocity and other components of the wind. Moreover, the EnKF has been shown to be effective at convective scales (Dowell et al. 2004; Aksoy et al. 2009) as well as at larger scales noted above.

The rest of the paper is organized as follows: in section 2, methodology for the assimilation system and forecasts is discussed including the forecast verification approach; section 3 covers the mean atmospheric state over the analysis period, analysis bias characteristics, and deterministic forecast character as relates to the parent analysis bias; a discussion of these results is provided in section 4 along with a summary and future directions in section 5.

2. Methodology

a. WRF description

The 50-member ensemble system is composed of WRF (version 3.2.1; Skamarock et al. 2008) short-duration forecasts over a limited-area domain at mesoscale (15-km horizontal) resolution covering the conterminous United States (CONUS), portions of the eastern Pacific, Canada, and the Gulf of Mexico, as shown in the largest domain in Fig. 1. For CP forecasts, a two-way nest (d02, 3-km horizontal resolution) is added by downscaling a selected ensemble member analysis state, while GFS forecasts
provide boundary conditions on the outer domain. The CP forecasts discussed herein were initialized from a member of the 0000 UTC ensemble analysis from 14 May–12 June 2011 and integrated 36 h.

Aspects of the default model physics configuration for the mesoscale analysis and CP forecasts are in Table 1. Choices were guided by experience in former spring experiment forecast success at NCAR (e.g., Weisman et al. 2008), as well as at NOAA’s Earth System Research Laboratory with the High-Resolution Rapid Refresh (HRRR) system (e.g., Alexander et al. 2010). The Kain–Fritsch (Kain 2004) cumulus parameterization used on the coarse domain was turned off for the high-resolution nest forecasts. Both domains used the Thompson microphysics scheme (Thompson et al. 2008) that includes prognostic snow, graupel, and cloud water mixing ratio, and both mixing ratio and number concentration for cloud ice and rain. Both domains also use the Mellor–Yamada–Janjic (MYJ) PBL scheme (Janjic 1994), the Noah land surface model (Ek et al. 2003), Rapid Radiative Transfer Model (RRTM) longwave (Mlawer et al. 1997), and Goddard shortwave scheme (Chou and Suarez 1994).

The Tiedtke cumulus scheme (Tiedtke 1989; Zhang et al. 2011a), released in WRF version 3.3, was ported to this WRF version and hereafter it is this modified WRF, version 3.2.1, using Tiedtke cumulus parameterization that is referred to as the “control” physics or TIE set, as shown in Table 1. Alternate physics sets of the analysis system are then the same as TIE except: (i) Kain–Fritsch for cumulus parameterization (KAF; Kain 2004); (ii) Yonsei University PBL scheme (YSU; Hong et al. 2006); and (iii) Morrison explicit microphysics (MOR; Morrison et al. 2009). Aside from the physics differences in the model used to advance the state between assimilation cycles, the four analysis system suites were otherwise configured identically and all used the same forecast model configuration as shown in Fig. 2. The impact of physics differences in the model advance on initial conditions for subsequent forecasts will be evaluated by quantifying the forecast skill from each analysis set.

b. DART ensemble data assimilation system

The DART ensemble data assimilation (EnDA) system for this study is the DART toolkit (Rev. 4869; available online at http://www.image.ucar.edu/DARES/DART/; Anderson and Collins 2007; Anderson et al. 2009) configured as an ensemble adjustment Kalman filter (Anderson 2001). Adaptive spatially and temporally varying inflation (Anderson 2009) is applied to the prior (background first guess) state to assist in
maintaining ensemble spread. Sampling error correction (Anderson 2012) is also applied to further improve spread and reduce influence from spurious correlations due to a limited ensemble size. Horizontal and vertical localization are used to overcome sampling errors using a Gaspari and Cohn (1999) weighting function with horizontal (vertical) weights diminishing to zero 1020 (13) km away from the observation location. As in Torn (2010), ensemble spread is further preserved where local observation density exceeds a count of 1600 within the localization ellipsoid by linearly reducing the localization length scale by the count overage ratio. The analysis is updated using DART from continuously cycled 6-h 50 member ensemble WRF forecasts. Additional DART configuration details are provided in Table 2.

c. Initial and lateral boundary conditions

The initial ensemble state is generated from the National Centers for Environmental Prediction (NCEP) GFS analysis valid at 0600 UTC 27 April 2011. The soil state is similarly drawn from the GFS analysis on this date and evolves freely thereafter for each ensemble member. Random perturbations were added to each member by sampling the NCEP background error covariance using WRFDA-3DVAR (three-dimensional variational data assimilation) (Barker et al. 2012) to produce an ensemble of perturbed ICs, as in Torn and Hakim (2008). The ensemble member lateral boundary condition (LBCs) perturbations are generated using the fixed covariance perturbation technique (Torn et al. 2006). A similar approach is used to generate LBC perturbations for each model advance between analysis times using a 6-h GFS forecast. The analysis and forecast period discussed herein is for the period 14 May–12 June 2011, so the initialization approach is expected to have little residual impact. The initial ensemble state used for the physics perturbation tests begins from the background KAF ensemble state at 0000 UTC 13 May 2011. As noted above, convection-permitting forecasts draw initial conditions from a single ensemble member state. The selected member was the closest to the mean ensemble state by normalized root-mean-square (RMS) difference. Unlike in the cycled analysis system, CP forecasts used unperturbed GFS forecast for LBCs.

The CP forecasts were also made daily using the same configuration of WRF (KAF) with initial conditions taken from GFS analysis at 0000 UTC between 14 May and 12 June 2011. These forecasts are used as a comparison set to CP forecasts launched from the DART

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cycled analysis domain</th>
<th>High-resolution forecast domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal grid</td>
<td>415 × 325, Δx = 15 km</td>
<td>1046 × 871, Δx = 3 km</td>
</tr>
<tr>
<td>Vertical grid</td>
<td>35 levels, ptop = 65 hPa</td>
<td>Same</td>
</tr>
<tr>
<td>Cumulus scheme</td>
<td>Tiedtke*</td>
<td>None</td>
</tr>
<tr>
<td>PBL scheme</td>
<td>MYJ</td>
<td>Same</td>
</tr>
<tr>
<td>Explicit microphysics</td>
<td>Thompson</td>
<td>Same</td>
</tr>
<tr>
<td>Radiation (longwave)</td>
<td>RRTM</td>
<td>Same</td>
</tr>
<tr>
<td>Radiation (shortwave)</td>
<td>Goddard</td>
<td>Same</td>
</tr>
<tr>
<td>Land surface scheme</td>
<td>Noah</td>
<td>Same</td>
</tr>
</tbody>
</table>

* For convection-permitting forecasts, the outer domain uses the Kain–Fritch scheme.

### Table 2. WRF options.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter type</td>
<td>EAKF</td>
</tr>
<tr>
<td>Adaptive inflation</td>
<td>1.0, 0.6 (initial mean, spread)</td>
</tr>
<tr>
<td>Adaptive localization threshold</td>
<td>1600</td>
</tr>
<tr>
<td>Localization type</td>
<td>Gaspari–Cohn</td>
</tr>
<tr>
<td>Horizontal localization half-width</td>
<td>510 km</td>
</tr>
<tr>
<td>Vertical localization half-width</td>
<td>6.5 km</td>
</tr>
<tr>
<td>Outlier threshold</td>
<td>3.0</td>
</tr>
<tr>
<td>Ensemble members</td>
<td>50</td>
</tr>
<tr>
<td>Sampling error correction</td>
<td>True</td>
</tr>
</tbody>
</table>

**Figure 2.** Schematic representation of the continuously cycled WRF-DART analysis configurations and inputs to the forecast model as described in section 2a.
ensemble analysis system ICs. Note that the NCEP Global Data Assimilation System (GDAS) does not assimilate surface observations and provides a global analysis at 0.5° horizontal resolution, or about 3 times coarser than the WRF-DART analysis. GDAS assimilates additional observations, such as satellite radiances, that are not assimilated in this study.

d. Observation sources and processing

The real-time system relied on observation processing and quality control (QC) from several sources. Observations were obtained from Global Systems Division (GSD) Meteorological Assimilation Data Ingest System (MADIS) for mandatory level rawinsondes \([u, v, T, T_d, \text{altimeter (Alt)}]\), standard aviation routine weather report (METAR) \((u, v, T, \text{Alt})\), as well as Aircraft Meteorological Data Relay (AMDar) reports \((u, v, T)\). MADIS processing and automated QC, (as described online at http://madis.noaa.gov/madis_qc.html) includes spatial and temporal consistency checks. Atmospheric motion vectors (AMVs; \(u, v\); Velden et al. 2005) were processed by and obtained from the Cooperative Institute for Satellite Studies Space Science and Engineering Center. Global positioning system (GPS) radio occultation observations (Kursinski et al. 1997) were processed by and obtained from Constellation Observing System for Meteorology Ionosphere and Climate (COSMIC).

A complete list of observation types and assumed observation errors appears in Table 3. Typical spatial distributions of assimilated observations (Fig. 3) were inhomogeneous by platform type, with AMVs over oceans, radiosondes evenly distributed over land, randomly distributed GPS profiles, AMDAR reports spreading from major airport hubs, and higher concentrations of surface observations in the eastern two-thirds CONUS. For this study dewpoint is used for moisture observations with observation error assignment following Lin and Hubbard (2004). This approach leads to larger dewpoint observation error assignment as the relative humidity decreases.

Further observation processing included (i) AMDAR report density was reduced by averaging observations over boxes of dimension 30 km in the horizontal and 25 hPa in the vertical, following Torn (2010); (ii) AMVs were also averaged spatially but over 60 km in the horizontal and were excluded over land; (iii) surface observations were excluded when the model terrain and station height differed by more than 300 m; (iv) radio occultation profiles were thinned to 15 levels; (v) observations within three grid lengths of the lateral boundaries were excluded; and (vi) observation errors for observations within five grid points distance of lateral boundaries were inflated to minimize analysis increments adjacent to the boundary edges, enhancing system stability.

Additional QC in the real-time system was based on the magnitude of the difference between the observation and prior ensemble mean. Observations were rejected when the squared difference exceeded 3 times the sum of the prior ensemble variance and observation error variance.

e. Precipitation verification method

The CP forecasts were verified with a focus on total precipitation. High-resolution NWP models are more prone to point displacement errors than coarser-resolution models. Thus, when point-by-point verification metrics are used to evaluate forecast quality, coarser-resolution models sometimes score better than higher-resolution configurations even if the higher-resolution forecasts subjectively produce more realistic features and provide greater forecast value (e.g., Mass et al. 2002; Done et al. 2004; Weisman et al. 2008). To reconcile this apparent disparity between realism, value, and quality of high-resolution forecasts, several objective “neighborhood approaches” (e.g., Ebert 2008; Casati et al. 2008; Gilleland et al. 2009) have been developed that relax the requirement that observed and forecast events match at the grid scale for forecasts to be considered statistically ideal. It is hoped that neighborhood approaches can

<table>
<thead>
<tr>
<th>Platform</th>
<th>Variable</th>
<th>Observation error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiosonde</td>
<td>Temperature</td>
<td>NCEP statistics</td>
</tr>
<tr>
<td></td>
<td>E–W, N–S winds</td>
<td>NCEP statistics</td>
</tr>
<tr>
<td></td>
<td>Dewpoint</td>
<td>Lin and Hubbard</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2004)</td>
</tr>
<tr>
<td>AMDAR (30 km, 25 hPa)*</td>
<td>Surface alimeter</td>
<td>2 hPa</td>
</tr>
<tr>
<td>METAR</td>
<td>Temperature</td>
<td>NCEP statistics</td>
</tr>
<tr>
<td></td>
<td>E–W, N–S winds</td>
<td>NCEP statistics</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>2 K</td>
</tr>
<tr>
<td>Buoy and ship reports</td>
<td>Temperature</td>
<td>2 K</td>
</tr>
<tr>
<td></td>
<td>E–W, N–S winds</td>
<td>1.75 m s(^{-1})</td>
</tr>
<tr>
<td></td>
<td>Altimeter</td>
<td>1 hPa</td>
</tr>
<tr>
<td></td>
<td>Dewpoint</td>
<td>Lin and Hubbard</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2004)</td>
</tr>
<tr>
<td>AMVs (60 km, 25 hPa)*</td>
<td>E–W, N–S winds</td>
<td>50% NCEP statistics</td>
</tr>
<tr>
<td>GPS (thinned to 15 levels)</td>
<td>RO refractivity</td>
<td>Kuo et al. (2004)</td>
</tr>
</tbody>
</table>

* Superobs (horizontal, vertical).
statistically corroborate subjective impressions of high-resolution model forecast utility.

The fractions skill score (FSS; Roberts 2005; Roberts and Lean 2008) is one measure that employs a neighborhood approach. Computing the FSS requires that the observations and model output be on a common grid. Thus, WRF forecasts of 1-h accumulated precipitation were bilinearly interpolated onto the stage IV grid (horizontal grid spacing of ~4.7 km) for the verification region shown in Fig. 1. The first step of calculating the FSS is to select precipitation thresholds $q$ to define an event and convert both the observed and forecast precipitation fields into fractional grids quantifying the probability of precipitation $\geq q$, as in Theis et al. (2005). Using circular geometry (Schwartz et al. 2009), a radius of influence $r$ (e.g., $r = 50$ km) is specified to construct a neighborhood around each of the $N$ grid boxes within the verification domain. All grid points whose centers fall within the radius of influence of the $i$th grid point are included in the neighborhood. A fractional value at the $i$th point is found by dividing the number of grid boxes with accumulated precipitation $\geq q$ within the neighborhood by the total number of points in the neighborhood.

The FSS compares the observed $o_i$ and forecast $f_i$ fractions. First the “fractions Brier score” (FBS; Roberts 2005) is found:

$$FBS = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2.$$  (1)

The FBS is identical to the Brier score (Brier 1950), except $o_i$ is allowed to vary between 0 and 1. To produce a measure of skill, the FBS is compared to a reference forecast. Roberts (2005) and Roberts and Lean (2008) used the worst possible FBS (FBS$_{\text{worst}}$) as the reference, which occurs when there is no overlap of nonzero fractions and is expressed as

$$FBS_{\text{worst}} = \frac{1}{N} \sum_{i=1}^{N} (f_i^2 + o_i^2).$$  (2)

Finally, the FSS is calculated as

$$\text{FSS} = 1 - \frac{\text{FBS}}{\text{FBS}_{\text{worst}}}.$$  (3)

---

**Fig. 3.** Spatial distribution of assimilated observation platforms from a single 0000 UTC analysis cycle. Observation counts are for unique observations within each platform type.
The FSS varies from 0 to 1, with a perfect forecast achieving a score of 1. At large scales approaching the size of the model domain, more biased forecasts yield lower FSSs, but at smaller scales, it is possible that a high bias may increase the FSS (Mittermaier and Roberts 2010). Computing the FSS for various values of $r$ indicates the scales at which forecasts reach a desired level of skill. For results presented herein, accumulation interval was 1 h, with no applied bias correction.

3. Results

a. Mean state characteristics

Mean atmospheric conditions for the TIE cycling experiment over the entire period (0000 UTC 13 May through 0000 UTC 12 June 2011; see Fig. 4) featured a mean upper-level trough in the Pacific Northwest and weak ridging in the eastern CONUS leading to mean west-southwest flow aloft over the central plains and Midwest (Figs. 4a,b). A low-level ridge over the deep South favored southerly low-level flow and moisture transport from the Gulf of Mexico into the Midwest where a lower-troposphere frontal boundary resided on average (Figs. 4c,d). Upper-air disturbances passing through the mean pattern triggered numerous heavy rainfall events over a similar swath from the central plains through the Ohio Valley.

b. Analysis bias by physics suite

During the spring 2011 season the KAF configuration was run in real-time. High-resolution forecasts that used this analysis frequently displayed slower eastward propagation of weather systems than was either observed and or present in forecasts from other analysis systems. Convection was subjectively deemed as poorly forecast.
more often than in prior seasons (e.g., Weisman et al. 2008), owing to both these displacement errors and with the development of too little intense convection, particularly during the first 12 forecast hours. The latter was unexpected since the downscaled mesoscale analysis had initially nonzero values in hydrometer fields and accompanying mesoscale ascent providing a ‘‘moist’’ start. Discussions with colleagues and sensitivity testing eventually led to investigating model error characteristics by employing various physics suites in the WRF-DART analysis system, the results of which follow.

Vertical profiles of time averaged mean fits of the $6$-h forecast ensemble mean to all assimilated radiosonde (AMoDAR) observations for the period 13 May–12 Jun 2011 are shown in Figs. 5a–c (5d,e). Figure 5f is identical to Fig. 5a except the posterior (analysis) ensemble mean fit to radiosonde temperature observations is shown. For all physics suites, 6-h forecasts (priors) are both too warm and too humid with horizontal wind speeds that are on average too slow relative to radiosonde observations within most of the troposphere. Aside from synoptic variability, the temporal variability of model bias and RMS error in forecast and analysis compared to observations is quite small in the mid- to upper troposphere (not shown). Forecast temperatures in the lower troposphere are closest to observations (smallest bias) when the Tiedtke scheme is used (e.g., Fig. 5a). Differences in the trigger mechanisms for shallow convection in the cumulus parameterizations (Torn and Davis 2012) may account for the significantly larger positive temperature bias with KAF. Weaker coupling between the PBL and free atmosphere likely contributes to the smaller negative wind speed and moisture bias in the lower troposphere relative to the Tiedtke-based analyses. Spatially, the KAF warm bias maximum relative to radiosonde temperatures was most pronounced near the 700-hPa level from the northern plains and broadening in zonal extent toward the East Coast. MOR (Fig. 5a) features a unique warm bias in the upper troposphere. In a relative sense, the Morrison scheme spatial extent of cloud ice coverage was found to be about 3 times greater than Thompson (not shown). This difference is likely due to the Thompson scheme’s more aggressive conversion of cloud ice to snow (H. Morrison 2011, personal

Fig. 5. Vertical profiles of time averaged, ensemble mean, prior mean error with respect to (a) mandatory level radiosonde temperature, (b) mandatory level radiosonde dewpoint, (c) mandatory level radiosonde horizontal wind speed, (d) AMDAR temperature superobs, (e) AMDAR horizontal wind speed superobs for period 13 May–12 Jun 2011, and (f) as in (d), but for analysis mean error with respect to mandatory level radiosondes.
communication along with the large upper limit on cloud ice concentration in the Morrison scheme. The RRTM longwave radiation scheme is sensitive to the cloud ice layer, such that MOR has a relative warm bias within and immediately below the layer where cloud ice is more prevalent. YSU shows enhanced positive lower to midtroposphere moisture bias (Fig. 5b), along with the largest slow wind speed bias (Figs. 5c,e). Relative to MYJ, the YSU PBL scheme has a known tendency toward a warmer, drier PBL with a deeper mixed layer (e.g., Weisman et al. 2008; Trier et al. 2011). This may lead to stronger coupling of the PBL and free atmosphere promoting a larger upward transport of moisture (contributing to a higher midtropospheric moisture bias) and greater momentum drag (larger negative wind speed bias).

There are strong indications that the analysis bias relative to radiosondes is influenced not only by systematic error in the model but also by bias differences between observation platforms. Comparing mean temperature bias in the prior state relative to radiosonde (Fig. 5a) and AMDAR (Fig. 5d) observations shows considerable differences, especially in the upper troposphere. The background fit to AMDAR temperature observations suggest a cool model bias and indicates that AMDAR temperatures are on average 0.2–0.6 K warmer than radiosondes. This is broadly consistent with Ballish and Krishna Kumar (2008) who proposed bias correction by aircraft type ranging from 0.4 to −1.2 K, with a correction of −0.6 K suggested for the most common aircraft type at flight level (300–150 hPa). Despite our use of AMDAR “superobs,” assimilated AMDAR significantly outnumber radiosonde observations (e.g., Fig. 3) between 900 and 200 hPa, with both having similar observation error assignments. As a consequence, the analysis fits AMDAR temperatures much more closely than radiosonde temperatures, with greatest misfit relative to radiosondes at flight level (approximately 300–200 hPa). A further indication that the AMDAR temperatures are biased is that the mean fit to radiosonde temperatures improves slightly from the analysis (Fig. 5f) to the forecast at flight level. Despite the indicators in our study that AMDAR observation bias is likely a leading contributor to the temperature bias in the midtroposphere (e.g., Figs. 5a,d), Benjamin et al. (2010) identified AMDAR observations as having a large positive impact toward reducing forecast error in the Rapid Update Cycle (RUC) system. Within GSD AMDAR observation bias is tracked by specific aircraft tail numbers, generating rejection lists as needed (B. Moninger 2011, personal communication), but these suspect observations are still processed by WRF-DART system.

The wind speed bias also differs between radiosondes and AMDAR observations. The mean characteristic bias from the background state horizontal wind speed in the midtroposphere relative to radiosonde (AMDAR) observations is −0.6 m s$^{-1}$ (−0.4 m s$^{-1}$), as shown in Figs. 5c and 5e, respectively. Thus AMDAR wind speeds have a bias of roughly +0.2 m s$^{-1}$ relative to radiosonde observations. Similar to the temperature bias, the much greater number of wind observations from AMDAR than from radiosondes means that the wind speed bias in the analysis fit to AMDAR observations is near zero while a −0.2 m s$^{-1}$ remains in the analysis fit to radiosondes. Notably, the zonal wind component is the primary contributor to the slow wind speed bias and is spatially inhomogeneous as described later.

While assimilation of relatively dense surface observations can be key to forecasts of convective triggering (Fujita et al. 2007), the surface fields also exhibit substantial bias. Figure 6 shows time series of METAR station averaged RMS error, bias, and total spread (square root of the sum of observation error variance and ensemble variance) for both background and analysis (e.g., “sawtooth” plot) fits to assimilated surface observations for TIE. Other physics suites had similar performance, except YSU, which had characteristic shifts in the surface state toward a warmer and drier PBL while maintaining similar spatial patterns in misfit with the observations (not shown). RMS error and total spread for all surface observation types are fairly consistent over time and are generally of the same magnitude, with variations mostly attributed to synoptic-scale waves moving through the domain. Focusing on bias, assimilated altimeter observations (Fig. 6a) have a consistent negative bias in prior fit to observations (model surface pressure, i.e., total mass, is generally less than observed) with relatively large increments every cycle. Temperature mean error at 2 m (Fig. 6b) displays diurnal and synoptic-scale variability, though generally neutral bias overall. Bias in 2-m dewpoint (Fig. 6c) is nearly always positive indicating the model surface conditions are too moist with a significant diurnal cycle. For 10-m winds (Figs. 6d,e) temporal changes in the prior fit are dominated by synoptic variability within the domain, with a general tendency for the north–south (east–west) wind component in the model to have a stronger southerly (westerly) wind than observed near the surface.

Surface data assimilation with EnKF assimilation systems has been shown to improve thermodynamic character through the depth of the PBL, particularly during the afternoon when the PBL is deep (e.g., Hacker and Snyder 2005; Fujita et al. 2007). A spatial plot of mean observation-minus-forecast difference for
surface temperature, shown in Fig. 7b, closely matches the mean temperature analysis increment in Fig. 7a. This suggests that surface temperature observations dominate the temperature state increments at low levels. Consistent with prior work, the same spatial mean increment pattern in temperature was uniform within the PBL (e.g., Fig. 7e) while a different mean increment pattern is evident above the PBL (e.g., Fig. 7f). Surface moisture increments (Fig. 7c) primarily indicate positive moisture bias was prevalent across much of the CONUS east of the Rockies. Moisture analysis increments were also similar in spatial character through the depth of the PBL (not shown). Thus, mean thermodynamic analysis increments in the PBL were consistent with a persistent bias between the ensemble mean and surface observations. The smallest mean increments for thermodynamic state within the PBL, when averaged across the entire domain, were noted with YSU. While analysis increments from surface observations aided in reducing bias in the background within the PBL, these adjustments may be too small or too infrequent in our study to compensate for the pace of model state drift, at least for the physics combinations tested in this study.

CONUS scale circulation errors were also noted. At the surface, a positive (negative) pressure analysis increment (Fig. 7d) is noted where the low-level temperatures are too warm (cool) (Figs. 7a,e). A consistent anticyclonic mean wind increment (Fig. 6d, wind vectors) is centered over the positive pressure increment and beneath the mean upper-tropospheric trough pattern (Fig. 4a).
FIG. 7. Spatial distribution of time averaged mean analysis increments from TIE for (a) potential temperature and (c) water vapor mixing ratio with wind vector increments at lowest model level; (d) surface pressure and 10-m wind vector increments; potential temperature with wind vector increments at sigma levels (e) 0.921 35 and (f) 0.742 035 (ascending from 1 at the lowest model level); and (b) time averaged mean analysis bias ($O - F$) to METAR temperature observations for the period 13 May–12 Jun 2011.
c. Analysis system physics suite influence on convection-permitting forecasts

Recall all CP forecasts used an identical model configuration (KAF) with varying ICs supplied from each physics suite. The analyses were downscaled to 3-km horizontal resolution over domain 2 with disabled cumulus parameterization, with precipitation verification over a subsection of the inner domain (see stippled region in Fig. 1). Mean 36-h accumulated rainfall for CP forecasts initialized from WRF-DART (Figs. 8a–d) and GFS (Fig. 8e) analyses show similar large-scale accumulated rain patterns and compare favorably with NCEP stage-IV accumulated precipitation (Lin and Mitchell 2005; Fig. 8f). However, there is a general bias toward excess coverage at higher accumulation thresholds in the model forecasts than was observed, along with a wider latitudinal band of heavy precipitation. Further, the model forecasts predicted more rainfall across the northern plains than observed, yet too little across northern Kansas and parts of eastern Texas into the Ozarks. All forecasts agreed well with observations on limiting accumulated precipitation in the southern United States, while varying on the northern extent of this dry region.

Individual WRF-DART forecasts often developed severe convection proxies too far west relative to observations, and often also west of forecasts from GFS ICs. This contrasts with eastward forecast displacement errors noted by forecasts as described in Clark et al. (2012) during the 2010 Hazardous Weather Testbed. An example from a particular case is demonstrated by 24-h forecasts of simulated reflectivity from each analysis configuration and the corresponding radar composite (Fig. 9). Most of the forecasts place the intense convection over the central plains too far west and produce more precipitation across the northern plains than was observed. For KAF, the convection is farthest west in the central plains and less robust, yet this forecast, along with GFS, has a fair representation of the less strongly forced precipitation farther east in the upper Mississippi River valley. The forecasts from Tiedtke-based analyses (TIE, MOR, and YSU) produce greater coverage of convection with similar overall patterns. All show relatively poor forecasts for the weakly forced convection farther east. It is more difficult to assess the apparent
excess precipitation in the northern intermountain region due to greater uncertainty in observation quality.

While differences between forecasts varied, representative placement and timing differences for 12–36-h forecasts and observations are also shown in a Hovmöller diagram (as in Carbone et al. 2002) for KAF and GFS based ICs in Fig. 10. During this 10-day period, where overlain model and stage-IV precipitation swaths do not overlap such that model forecast precipitation is farther west and or later in time, along with stage-IV precipitation earlier and farther east, forecast precipitation systems are shown to be progressing from west to east slower than observed. While this pattern is particularly evident at times with KAF (e.g., Fig. 10a, 24–27 May), there are similarities in forecasts from GFS suggesting that the CP forecast model error also contributes to this signal. At other times, reversed patterns are evident suggesting general unreliability in the timing and location of precipitation event forecasts.

The areal coverage (Fig. 11), FSS (Fig. 12), and bias (Fig. 13) aggregated over all CP forecasts for accumulation thresholds of 0.2, 1.0, and 10 mm h\(^{-1}\) provide some insight into differences in forecast skill for each analysis. FSS is computed with a radius of influence (ROI) of 50 km, results were qualitatively similar with different ROI. First, based on areal coverage comparisons with stage IV, forecasts from all the WRF-DART analyses exhibit faster spinup than forecasts with GFS ICs (Fig. 11). Precipitation coverage reaches the observed coverage more rapidly for intense rainfall rates, especially with MOR. Over the first 12 h, with the exception of KAF analyses, all the WRF-DART analyses produce forecasts with superior FSS to those from GFS at the initial hour and some through 8–10 h (Fig. 12), while the GFS analyses clearly lead to more skillful forecasts during the next day, from local noon through the evening hours, particularly at higher rain rates (Fig. 12c; forecast hours 18–28). The forecasts from KAF analyses develop limited precipitation during the first 24 h.

Because of the relatively modest sample size (30) confidence intervals are large at individual forecast hours. To better highlight differences, FSS was aggregated over early (0–12 h) and late (18–36 h) forecast periods as a function of ROI (Fig. 14) and bootstrap confidence
intervals (CIs) were selectively added. The following procedure was applied to generate aggregate scores and bootstrap CIs based on hourly rainfall forecasts:

1) The FBS [Eq. (1)] and FBS\textsubscript{worst} [Eq. (2)] scores computed for hourly rainfall accumulations were summed over the hours of interest and all days.

2) FSS [Eq. (3)] was computed using the summed (i.e., aggregated) FBS and FBS\textsubscript{worst} quantities from step 1.

3) 10,000 resamples were drawn from a list of all FBS and FBS\textsubscript{worst} pairs that were elements of the sums described in step 1 to compute the bounds of the 90% confidence interval.

The most notable distinctions between GFS, KAF, and the forecasts from analyses that include the Tiedtke cumulus parameterization. An anonymous reviewer

![Time–longitude (Hovmöller) diagrams of meridionally averaged, hourly accumulated precipitation within the verification region for explicit forecasts from (a) KAF and (b) GFS analysis initial conditions for 12–36-h forecasts initialized from 20 to 30 May 2011. Color fill indicates where model forecast only exceeds 0.15 mm (blue), stage-IV precipitation analysis only exceeds 0.15 mm (red), and where both exceed 0.15 mm (gray).](image)

**FIG. 10.** Time–longitude (Hovmöller) diagrams of meridionally averaged, hourly accumulated precipitation within the verification region for explicit forecasts from (a) KAF and (b) GFS analysis initial conditions for 12–36-h forecasts initialized from 20 to 30 May 2011. Color fill indicates where model forecast only exceeds 0.15 mm (blue), stage-IV precipitation analysis only exceeds 0.15 mm (red), and where both exceed 0.15 mm (gray).

![Time mean areal coverage of precipitation at rain rates of (a) 0.2, (b) 1.0, and (c) 10 mm h\textsuperscript{-1} by forecast hour from NCEP stage IV (black), KAF (green), TIE (violet), YSU (red) and MOR (gold) physics suite, and GFS (blue) for initial conditions for convection-permitting forecasts over the verification domain.](image)

**FIG. 11.** Time mean areal coverage of precipitation at rain rates of (a) 0.2, (b) 1.0, and (c) 10 mm h\textsuperscript{-1} by forecast hour from NCEP stage IV (black), KAF (green), TIE (violet), YSU (red) and MOR (gold) physics suite, and GFS (blue) for initial conditions for convection-permitting forecasts over the verification domain.
suggested applying a bias correction. However, over this domain, zeroes dominated the precipitation fields and the physical thresholds corresponding to percentiles were very small. To avoid this problem, we calculated the percentiles by excluding zeroes and from a climatological perspective, as in Schwartz et al. (2010). The results did not meaningfully differ from those using absolute thresholds.

There is a notable dip in intense convection coverage for forecasts from TIE analyses during the local early morning hours (Fig. 13c; forecast hours 10–18), and to a lesser extent with MOR. Forecasts from YSU analyses have the largest positive bias (Fig. 13a), especially at higher rain rates, yet FSS is similar to other Tiedtke-based-analysis forecasts (Fig. 12). This is consistent with the analysis system based on YSU also having the largest lower- to midtropospheric moisture bias (Fig. 5b). Overall, MOR analyses led to the most skillful forecasts during the first few hours, despite having the largest upper-tropospheric temperature bias (e.g., Figs. 5a,d).

4. Discussion

a. Short-range forecast performance (0–12 h)

Forecast skill (Figs. 12 and 14) and spinup (Fig. 11) are uniformly better during the first 12 h of the forecasts with WRF-DART than GFS ICs. The higher horizontal resolution in WRF-DART analyses contributes some of the forecast advantage over GFS during the first few hours of the forecast. Skamarock (2004) noted 6–12-h time windows for mesoscale kinetic energy spectrum to
reach maturity when starting from a 40-km Eta Model analysis in a 4-km forecast. The assimilation of surface observations into the WRF-DART system also leads to an improved surface analysis over GFS ICs. Fujita et al. (2007) found surface data assimilation improved their ensemble forecasts out to 6 h. Finally, the downscaling approach used in our study to initialize the inner nest leads to a moist start for the initial state from parameterized precipitation fields and accompanying mesoscale ascent.

b. Longer-range forecast performance (12–36 h)

Beyond the first 12 h, the WRF-DART analysis advantage faded while GFS analysis based forecasts on average led to superior precipitation forecast skill with reduced positive bias. All CP forecasts show large positive precipitation biases during the second diurnal cycle (Fig. 13). For example, note in Fig. 9 all forecasts produce more significant coverage of precipitation in the northern high plains than was observed. Likely some of this areal coverage bias shared by all forecasts owes to model bias from the common high-resolution forecast model.

Previous studies of numerical simulations of warm season rainfall system propagation (e.g., Carbone et al. 2002) have found explicitly resolved convection can represent the main diurnal characteristics (e.g., Liu and Moncrieff 2007; Weisman et al. 2008; Clark et al. 2009). Unexpectedly, the Tiedtke-based analysis led to forecasts with too little intense rainfall during the local early morning hours (Fig. 13c). The background forecasts in all the analyses have positive moisture biases through the depth of the troposphere (Fig. 5b; GFS not shown). Since increased precipitation bias can lead to higher FSS at small scales, the relative merits of different WRF-DART analyses configurations is difficult to disentangle. The exception is the KAF analyses, which consistently underperformed during the first 24 h of the forecast.

Subjective evaluation of CP forecasts during forecast hours 18–30 h often revealed more realistic system
propagation characteristics and fewer poor convective forecasts with GFS initial state than WRF-DART analyses. Testing of physics modifications to the CP forecast model configuration, using the KAF analysis for ICs, included microphysics and PBL physics sensitivity. These tests showed limited forecast sensitivity in the placement of convective features, similar to results from prior studies (e.g., Weisman et al. 2008). Examining the mean GFS and TIE analysis states revealed significant large-scale temperature and wind differences, where the GFS analysis state was generally closer to radiosonde observations, particularly with regard to bias. Exceptions were with (i) fits to surface observations, with the GFS analysis further from observations, and (ii) fits to AMVs off the west coast of the Baja California Peninsula toward the southwest corner of the domain. The TIE analysis fit to radiosonde observations was noted to have higher day-to-day variability than GFS. Further reduction in analysis temperature bias may be possible by adjusting current observation processing approaches to effectively reduce (increase) the impact of AMDAR (radiosonde) observations. Reducing the amount of vertical localization of surface altimeter observations might also lead to a more significant correction of midtropospheric temperatures.

Improvements in longer-range predictability also appear possible from further reduction in large spatial scale model errors. For example, the relative warm bias noted in the Pacific Northwest from mean temperature analysis increments (Figs. 7c,f) is loosely collocated with the mean upper-level trough (Figs. 4a,b). By thermal wind balance, a reduction in the lower-troposphere baroclinicity from this warm bias within the mean trough would reduce upper-level flow. Consistent with this, mean analysis increments of wind in the mid- to upper troposphere correct a slow wind bias in the base and along the lee side of the mean upper trough (Fig. 15a). Despite these corrective increments, slow wind speed bias remains present in this area of the analysis state and elsewhere (Fig. 15b). Additional observations may be needed in the eastern Pacific region to address some of these inconsistencies. The reduced upper-level wind speeds would contribute to slower eastward progression of embedded upper-level disturbances that aid in initiating convection and driving convective evolution during forecasts. A westward bias was present at times in precipitation forecast swaths (e.g., Fig. 10) and in placement of intense convection (e.g., Fig. 9). Finally, model bias in the CP forecast model is as yet unexamined for potential interaction with the bias characteristics of the parent analysis system. Forecast errors associated with this bias adjustment process warrant future investigation.

Finally, several operational forecast centers have identified problems with CCDA with limited-area models. For example, Rogers et al. (2009) found the NAM synoptic-scale forecasts were more skillful from GFS analysis than from the NAM data assimilation system, prompting their use of a partial cycling strategy. Other forecast centers have adopted a similar partial cycling strategy (Environment Canada; Fillion et al. 2010; S. Benjamin, ESRL, 2012, personal communication) with
like justifications. Whether these errors are associated with model bias that develops over many cycles, thus similar to the errors shown here, has not been detailed in these prior studies.

c. Precipitation and moisture bias

High precipitation bias is a known problem in CP forecasts (e.g., Weisman et al. 2008; Schwartz et al. 2009, 2010) despite significant mitigation from numeric improvements (Skamarock and Weisman 2009). Aside from KAF the WRF-DART-based precipitation forecasts were higher in bias than GFS (Fig. 13). The precipitation bias in our results complements analysis moisture bias in both the PBL (Figs. 6c and 7c) and throughout the troposphere (Fig. 5b). Along these lines, YSU had the highest moisture bias in the lower troposphere (Fig. 5b) and the highest precipitation bias (Fig. 13). Model errors in the CP forecast model likely further enhance moisture bias.

Consistent with the overly moist forecasts, negative analysis increments in boundary layer water vapor (e.g., Fig. 7c) were found across much of the central plains and Midwest. Despite daily mean water vapor increments of up to 1 g kg\(^{-1}\) through the depth of the PBL, remaining or quickly redeveloping surface moisture bias likely contributes to the deeper tropospheric moisture bias (Fig. 5b). Further improvements in the moisture analysis might be possible by assimilating additional tropospheric moisture observations (Jones and Stensrud 2012).

Comparing mean rainfall patterns for explicit forecasts against NCEP stage-IV rainfall (Fig. 8), a swath of significant rainfall overlaps only the northern part of the region with surplus boundary layer water vapor (Fig. 7c). Examining the time series plots of ensemble mean fit to surface observations (Figs. 6b,c) reveals a strong diurnal cycle in temperature (moisture) bias, peaking daily at 1800 (0000) UTC. Similarly, partitioning mean increment contributions by analysis hour (0000, 0600, 1200, and 1800 UTC) confirms the largest temperature and moisture increments were at 1800 and 0000 UTC, respectively. This characteristic diurnal cycle in the fit to observations and analysis increments was noted for all physics suites.

The surface temperature and moisture bias characteristics were not particularly sensitive to the physics suites. Elevated moisture and temperature in the boundary layer may owe to excess latent and sensible heat flux. Time-averaged spatial coverage of cloud ice in MOR was found to be about 3 times greater than with Thompson microphysics. The greater cloud ice with MOR clearly impacted the longwave radiation scheme producing significant warming of upper-tropospheric temperatures (e.g., Figs. 5a,d) but no notable impact on surface fluxes.

An examination of how the Goddard shortwave radiation scheme interacted with the explicit microphysics schemes warranted further investigation. Moreover, excess downwelling shortwave energy reaching the surface may be due to any or all of (i) insufficient cloud cover, (ii) insufficient sensitivity to cloud cover by the shortwave radiation scheme, and (iii) lack of aerosol treatment, or perhaps other sources.

d. Representativeness of ensemble-derived initial states

Ideally, we would capitalize on the ensemble of analyses provided by the EnKF and make an ensemble forecast at CP resolution. Our use of a single, deterministic CP forecast is dictated solely by the computational cost of an ensemble of CP forecasts. Faced with similar constraints, other EnKF studies have also employed single deterministic forecasts when evaluating the influence of the EnKF on forecasts. Torn and Hakim (2008), Buehner et al. (2010), and Liu et al. (2012) evaluate deterministic forecasts from the ensemble-mean EnKF analysis. The ensemble-mean analysis is expected to have the minimum RMS error but also will spatially smooth small-scale features, such as tropical cyclones or convective cells, whose position is even slightly uncertain. Thus, Torn (2010) and Cavallo et al. (2013) considered tropical-cyclone forecasts initialized from the analysis of a single member selected at each analysis time. We have taken the same approach here, though our criterion for selecting the member differs from theirs.

To examine the relative impact of selecting a single ensemble member analysis for ICs for forecasts in our study, tests were done using the ensemble mean state for ICs. On average these forecasts were less skillful and displayed slower spinup for the first few hours, but thereafter were more skillful than forecasts from an individual member analysis. In fact, forecasts from the TIE ensemble mean analysis were more similar to GFS in forecast skill beyond 24 h, although they included a higher precipitation bias (not shown). The shift in systematic precipitation bias from using the ensemble mean was unexpected and warrants future investigation.

5. Summary

This study examined present capabilities and challenges for EnDA with real observations for application to future CP forecast systems. To understand sensitivity to WRF bias in a continuously cycled EnDA system, several model physics configurations were tested. A once-daily 15-km analysis from the 0000 UTC cycle from each model configuration suite provided initial and
LBCs for 3-km CP forecasts. Different physics suites for the analysis system were shown to generate unique model background error characteristics that were carried into the analysis. The forecast skill and bias characteristics for the CP forecasts that used the analysis for ICs were found to also have unique error characteristics and some similarities with background error characteristics of the parent analysis system. Both observations and model physical parameterizations were shown to be sources of bias. While the EnKF considered here functions well despite the biases in the background forecasts and observations, this study presents a clear need to mitigate bias errors to improve ICs and subsequent CP forecasts in continuously cycled data assimilation (CCDA) systems.

We considered the sensitivity of CP forecasts to analyses produced using four different physics suites in the assimilation system: TIE uses the configuration of WRF given in Table 1, KAF replaces the Tiedtke convective parameterization with Kain–Fritsch, YSU replaces the PBL scheme with the Yonsei University scheme, and MOR replaces the microphysics with the Morrison scheme (see Fig. 2). During the first 12 h, forecasts from MOR ICs (KAF) led to the most (least) skillful precipitation forecasts, while YSU (KAF) had the highest (lowest) precipitation bias. During the next full diurnal cycle in the forecast, ICs from YSU (KAF) led to the most (least) skillful forecasts, while YSU (KAF) still had the highest (lowest) bias. These results may not generalize to other domains, seasons, or model configurations. However, users of CCDA systems and regional climate modelers may benefit from using a similar approach to compare their model climatology with observations.

Forecasts from WRF-DART analysis ICs all provided more skillful CP forecasts than those from a GFS analysis during the first 12 h, particularly the first 5–6 h. The downscaled microphysics fields and mesoscale circulations, improved surface analysis, and finer spatial resolution may be individually or collectively beneficial. Beyond the first 12 h, forecasts from GFS analyses were generally more skillful. The improved forecasts from GFS beyond 18 h are consistent with findings at other forecast centers that have found poorer forecast performance from CCDA systems than a partial cycling approach.

Substantial bias remains in the background ensemble DA system, likely tied to model bias with an emphasis on physical parameterization characteristics, as well as observation bias characteristics by platform. Reduction of these bias sources in the cycled WRF-DART analysis system will likely lead to improved forecasts when the analysis product is used for forecast ICs. A significant warm temperature bias in the lower troposphere was noted within a mean upper-level trough in the Pacific Northwest. This temperature bias feature was consistent with background low surface pressure bias and cyclonic mean wind field bias patterns at the surface. Aloft, wind speeds were too slow through the base and lee side of the upper trough, consistent with weather disturbances and accompanying precipitation episodes passing through the mean pattern having slower forward motion than observed on average. Substantial positive moisture bias was noted both in the boundary layer and throughout the troposphere. Moreover, mean forecast model background errors were consistent with subsequent forecast error characteristics for convection-permitting forecasts that used the analyses for ICs. Observation bias differences were noted between the background fit to AMDAR and radiosonde observations. Further improvements in the spinup of convection during the first few hours of the forecast could likely be achieved by applying a cloud analysis to the IC state (e.g., Kain et al. 2010).

In the future we plan to explore CP ensemble forecasts, where the initial ensemble is drawn from a WRF-DART mesoscale ensemble system. While convection-permitting ensemble forecasts are becoming more common (e.g., Kong et al. 2007; Xue et al. 2008, 2009, 2010), there is much to be explored in probabilistic prediction of convection, in particular how best to design an ensemble (Clark et al. 2008). Constructing a CP ensemble with appropriate growth of ensemble spread to approximate forecast uncertainty is crucial, but based on results found in this study and others a multiphysics or multi-parameter approach might generate biased probabilities. Stochastic energy backscatter (Berner et al. 2011) might be a better approach toward enhancing reliability through meaningful growth of ensemble spread, without sacrificing sharpness. The forecast skill from ensemble mean ICs should be compared with mean forecasts from an ensemble to understand if the behavior found in this study is consistent. Efforts will also continue to identify and hopefully improve bias tendencies in WRF. Given that mean increments made significant bias corrections to the background, more frequent cycling may reduce bias in the background and/or further constrain model error growth. The potential benefit from more frequent cycling needs to be weighed against the potential negative influence from initial state imbalance and limited availability of thermodynamic observations above the PBL. Collectively, these efforts appear likely to produce more skillful forecasts with CCDA systems.

Acknowledgments. Funding for the lead author was primarily provided by the NCAR’s Short-Term Explicit
Prediction program. Helpful discussions with Ryan Torn, David Dowell, So-young Ha, Jimy Duhdia, Dave Gill, Kevin Manning, Peggy LeMone, and Fei Chen all aided in various aspects of this study. We are particularly grateful to Wei Wang, Kevin Manning and Ming Chen who configured and ran the real-time convection-permitting WRF model forecasts. The NCAR DART team assisted in technical aspects of the DART software, especially Nancy Collins and Tim Hoar. We appreciate contributions from two anonymous reviewers that led to significant improvements in the paper. Real-time observation data access was provided by MADIS, COSMIC, and SSEC. Ryan Torn provided software and scripts that aided in running the real-time WRF-DART system. We would like to acknowledge high-performance computing support provided by NCAR’s Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Radar composite data was obtained from the Iowa Environment Mesonet website.

REFERENCES


