Evaluation of NU-WRF Rainfall Forecasts for IFloodS

DI WU
Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, and Science Systems and Applications, Inc., Lanham, Maryland

CHRISTA PETERS-LIDARD
Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland

WEI-KUO TAO
Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland

WALTER PETERSEN
Code 610W, NASA Goddard Space Flight Center Wallops Flight Facility, Wallops Island, Virginia

(Manuscript received 27 July 2015, in final form 16 February 2016)

ABSTRACT

The Iowa Flood Studies (IFloodS) campaign was conducted in eastern Iowa as a pre-GPM-launch campaign from 1 May to 15 June 2013. During the campaign period, real-time forecasts were conducted utilizing the NASA-Unified Weather Research and Forecasting (NU-WRF) Model to support the daily weather briefing. In this study, two sets of the NU-WRF rainfall forecasts are conducted with different soil initializations, one from the spatially interpolated North American Mesoscale Forecast System (NAM) and the other produced by the Land Information System (LIS) using daily analysis of bias-corrected stage IV data. Both forecasts are then compared with NAM, stage IV, and Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation (QPE) to understand the impact of land surface initialization on the predicted precipitation. In general, both NU-WRF runs are able to reproduce individual peaks of precipitation at the right time. NU-WRF is also able to replicate a better rainfall spatial distribution compared with NAM. Further sensitivity tests show that the high-resolution runs (1 and 3 km) are able to better capture the precipitation event compared to its coarser-resolution counterpart (9 km). Finally, the two sets of NU-WRF simulations produce very close rainfall characteristics in bias, spatial and temporal correlation scores, and probability density function. The land surface initialization does not show a significant impact on short-term rainfall forecast, which is largely because of high soil moisture during the field campaign period.

1. Introduction

The warm season convection forecasting (Done et al. 2004; Kain et al. 2008; Weisman et al. 2008) is one of the most challenging weather forecast problems (e.g., Fritsch and Carbone 2004; Jankov et al. 2005), which happens over the central United States from May through August, often under “weakly forced” conditions (e.g., Carbone et al. 2002). Weakly forced conditions often refer to certain upper-level instabilities but without the presence of a surface front (e.g., because of the presence of an upper-level trough and advection of potential vorticity). Numerous efforts have been undertaken to improve the warm season precipitation forecast (e.g., Gallus 1999; Gallus and Segal 2000; Carbone et al. 2002).

Studies have demonstrated that choosing the convective scheme strongly affects simulated precipitation (e.g., Wang and Seaman 1997; Gallus 1999). Convective schemes are designed with different assumptions (e.g., mass flux and moisture adjustment) and are only suited for certain applications (e.g., certain grid resolutions and study regions). There are generally two types of...
convective schemes, mass-flux type and moisture-adjustment type. Generally, those schemes that are parameterized with moist downdrafts are more desirable for predicting warm season precipitation than those without (Wang and Seaman 1997). The Grell–Dévényi (GD) scheme (Grell 1993; Grell and Dévényi 2002) is an ensemble mass-flux-type scheme with multiple closure assumptions and various parameters including updraft, downdraft, entrainment, detrainment, and precipitation efficiency. Many studies suggest that the GD scheme overestimates the lighter rain rate and underestimates the moderate category (e.g., Mukhopadhyay et al. 2010), since it averages results from all ensemble members. The Betts–Miller–Janjić (BMJ) scheme (Betts 1986; Betts and Miller 1986; Janjić 1994) is a lagged convective adjustment scheme. It adjusts the model’s thermal and moisture structures toward reference profiles that reflect the quasi-equilibrium state established by deep convection (Betts 1986). Deep convection is generally parameterized and not explicitly predicted when the horizontal grid spacing is greater than about 10 km (e.g., Molinari and Dudek 1986; Kalb 1987; Zhang et al. 1988; Gallus and Segal 2000; Bélair and Mailhot 2001; Liu et al. 2001; Roberts and Rutledge 2003), but convective parameterization (CP) is typically avoided at higher resolution. This avoidance is appropriate because the conceptual basis for CP becomes increasingly ambiguous as the grid spacing decreases further (e.g., Kain et al. 2008; Molinari and Dudek 1992; Arakawa 2004).

The increasing of model resolution has generally improved quantitative precipitation forecasting (Mesinger 1996). Weisman et al. (1997) concluded that 4-km grid spacing is sufficient to not require a CP scheme for deep convection and squall lines. A typical thunderstorm cell is on the order of 10 km in all three directions. Thus, it is commonly believed that 1-km grid spacing is sufficient to resolve the basic thunderstorm structure. Cloud resolving models (CRMs) with 1-km horizontal grid spacing are suited to study these systems because they can explicitly resolve the mesoscale and larger convective-scale dynamics. With recent advances in computing power, regional-scale numerical weather prediction starts to use 1–4-km high resolution as well (e.g., Weiss et al. 2008; Lean et al. 2008; Seity et al. 2011; Baldauf et al. 2011).

The positive feedback between soil moisture and precipitation is also well known. The wet soil surface facilitates heat and moisture transport into the planetary boundary layer (PBL), so the moistening of lower PBL is likely to produce localized convective precipitation. Gallus and Segal (2000) found that precipitation amounts respond to the change of soil moisture differently among runs with different convective parameterizations. Case et al. (2011) found that high-resolution land surface initialization produces some small improvements for the forecasted precipitation. Tao et al. (2011) have shown that land surface initial conditions only play a secondary role in simulated rainfall from a squall line over the southern Great Plains (SGP), either by varying initial soil moirstures or by applying a high-resolution land surface initialization. The impact of surface initialization on precipitation is also location dependent. General circulation model (GCM) results show that the land initialization for June–August (JJA) seasonal rainfall is only useful for certain limited areas, while the impact of land initialization on seasonal temperature is much more extensive (Koster et al. 2003).

The Iowa Flood Studies (IFloodS) campaign was conducted in eastern Iowa as a pre-GPM-launch campaign from 1 May to 15 June 2013, with the goal of examining how well GPM and other blended products could be used for flood forecasting. The areas of focus for the IFloodS campaign were the Cedar and Iowa Rivers basin, the Turkey River basin, and South Fork Iowa River (Fig. 1). The Iowa–Cedar Rivers basin is 12620 mi² with a population of about 1 million (Iowa–Cedar Watershed Interagency Coordination Team 2015). The Turkey River drains 1545 mi². According to the Iowa Flood Center, the annual maximum discharges often occur in March or April. The maximum is likely caused by snowmelt or heavy spring rainfall when soils are often near saturation. The peak of flood occurrences is in March, following a secondary peak in summer (Iowa Flood Center 2013). The flood forecast usually not only depends on the forecasted rainfall amount but also on the size and orientation of the precipitating systems relative to the location and shape of the basins involved. Identification for spatial displacement errors from the precipitating systems is important because it is the dominant source of quantitative precipitation forecast (QPF) error (Ebert and McBride 2000). Poor QPF skill hinders hydrologic applications, particularly streamflow forecasting operations (Cuo et al. 2011).

To support deployment of ground-based instrumentation, our team at Goddard Space Flight Center (GSFC) conducted real-time forecasting with the NASA-Unified Weather Research and Forecasting (NU-WRF) Model, which was delivered daily to support 0900 local time forecast briefings to the campaign personnel. This effort required not only dedicated computational resources, but also a robust modeling system that is capable of simulating severe convective episodes typical of eastern Iowa during the active spring period. In this work, we provide a comprehensive evaluation of the precipitation forecasts from real-time NU-WRF forecasts. By doing so, we would like to understand 1) if there is added value in high-resolution
NU-WRF simulations using North American Mesoscale Forecast System (NAM) forcing and 2) whether there is a positive impact on precipitation forecasts with high-resolution surface initialization. We first describe the experimental design, including the modeling system, configuration, and evaluation datasets. Next, we present an evaluation of the precipitation forecasts based on an archive for the entire experimental period relative to ground data in addition to an operational forecast model. Finally, we discuss the implications of this work for future forecasting applications.

2. Experiment design

The NU-WRF (http://nuwrf.gsfc.nasa.gov) modeling system has been developed at GSFC as an observation-driven integrated modeling system that represents aerosol, cloud, precipitation, and land processes at satellite-resolved scales (Peters-Lidard et al. 2015). NU-WRF is a superset of the National Center for Atmospheric Research (NCAR) Advanced Research version of WRF (ARW) dynamical core model, achieved by fully integrating the GSFC Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2015), the WRF Chemistry (WRF-Chem)-enabled version of the Goddard Chemistry Aerosol Radiation and Transport (GOCART; Chin et al. 2000) model, the Goddard Satellite Data Simulation Unit (G-SDSU; Matsui et al. 2009), and custom boundary/initial condition preprocessors. Several NASA physical packages (microphysics and radiation) have also been implemented into NU-WRF. These physical processes include CRM-based microphysics (Tao et al. 2003; Lang et al. 2007, 2011, 2014) and radiation (Chou and Suarez 1999) schemes. All the above features are combined into a single software release, with source code available by agreement with NASA GSFC.

In this study, NU-WRF, version 3.4.1 (based on NCAR ARW, version 3.4.1), is employed to conduct high-resolution simulations. There are 60 vertical levels and three spatial domains with 9-, 3-, and 1-km grid spacing (Fig. 2), and time steps of 27, 9, and 3 s, respectively. The
GD cumulus parameterization scheme (Grell and Dévényi 2002) is adopted for the outer domain; no CP is used for two inner domains. The planetary boundary layer (PBL) parameterization employs the Mellor–Yamada–Janjić (Mellor and Yamada 1982) level-2 turbulence closure model through the full range of atmospheric turbulent regimes. The Goddard broadband two-stream (upward and downward fluxes) approach is used for the shortwave and longwave radiative flux calculations (Chou and Suarez 1999) and its explicit interactions with clouds (microphysics). In addition, the numerical simulations use the Goddard three-class ice (3ICE) scheme (Lang et al. 2011), which prognoses three types of ice hydrometeor species (i.e., cloud ice, snow, and graupel).

The LIS is a core component of NU-WRF. It is a flexible land surface modeling and data assimilation framework developed with the goal of integrating satellite- and ground-based observational data products and advanced land surface modeling techniques to produce optimal fields of land surface states and physics (Kumar et al. 2006; Peters-Lidard et al. 2007). The infrastructure can not only be directly coupled with the atmosphere, but it can also integrate high-resolution observations with the model forecasts to generate improved estimates of land surface conditions such as soil moisture, evaporation, snowpack, and runoff at 1-km and finer spatial resolutions and at 1-h and finer temporal resolutions.

During the IFloodS campaign period, two sets of 48-h NU-WRF forecasts were produced twice a day initialized at 0000 and 1200 UTC from 1 May to 15 June 2013. These forecasts require 7 h to produce with 2048 cores on the NASA Center for Climate Simulation (NCCS) supercomputer. One set of the NU-WRF uses NAM to provide initial and boundary conditions without LIS coupling (Table 1, denoted as WRF); the second set uses LIS to provide high-resolution surface initialization and online coupling with the atmospheric component of NU-WRF (Table 1, denoted as COUP).

The role of LIS in the COUP simulation is twofold: first, to provide physically consistent land surface initialization for NU-WRF and second, to interact with the surface layer and atmospheric components of NU-WRF.
and produce coupled water, energy, and momentum fluxes. To prepare the surface initial conditions, an offline LIS starts from 2 May 2008 to 1 May 2013 in cold start mode. For consistency, the Noah land surface model runs offline within LIS using the same domain configuration and soil and vegetation database as NU-WRF. The long spinup period is used for land surface states to achieve equilibrium for initialization (Cosgrove et al. 2003a; Rodell et al. 2005). The LIS offline spinup uses the Global Data Assimilation System (GDAS) and the North American Land Data Assimilation System, phase 2 (NLDAS-2), to provide forcing input.

Besides the 5-yr LIS offline run to kick off the first of the coupled simulations, there is also a short LIS offline restart run to provide initial land surface conditions for every forecast cycle. As shown in Fig. 3, the forecast starts every day at 0000 and 1200 UTC for 48-h integration. Each forecast consists of an LIS offline analysis and an online coupling between WRF and LIS. The offline analysis is forced by the previous day’s NU-WRF forecast and stage IV, and then forward integrates for 24 h to the current model initialization time. Stage IV data are used to provide hourly precipitation forcing for the LIS analysis, while NU-WRF output from the previous day supplements the atmospheric forcing. It should be noted that the stage IV rainfall product is only used to provide the forcing for the land surface model (LSM) during the analysis cycle and is not assimilated into the atmospheric component of the simulation. The soil initialization for the control simulations comes from spatially interpolated soil moisture and soil temperature from NAM. For both NU-WRF simulations, NAM provides the atmospheric part of the initial and boundary conditions.

It is worth noting that different LSM versions have been used in WRF and COUP simulations (Table 1). The WRF uses Noah LSM, version 3.4.1, while COUP uses a slightly older version (3.2.1; Ek et al. 2003) of the Noah LSM, which was the most recently implemented version of Noah in LIS at the time of the campaign. The changes from 3.2 to 3.4 focused on land ice and sea ice (NCAR 2012) and did not result in significant differences in soil moisture, runoff, or land surface fluxes based on other offline analyses (not shown). These changes only apply to offline simulations but not to a coupled WRF run, so the differences between different Noah LSM versions should not affect results in this study.

### 3. Data and methodology

Two observational datasets are employed for model evaluation. The stage IV (Lin and Mitchell 2005) quantitative precipitation estimation (QPE) is an analysis product that integrates real-time gauge and WSR-88D data and benefits from the National Weather Service River Forecast Centers’ manual quality-control (QC) step. Stage IV data are available at 4-km grid spacing with 1-h frequency. It is a widely used rainfall product for both hydrological and meteorological communities because of its national coverage, high spatial and temporal resolutions, and overall low biases (e.g., Tang et al. 2014; Seo et al. 2013). Its good performance in mean-square error (MSE) and total bias results from the effectiveness of bias correction and the manual QC procedures (Cunha et al. 2015).

Multi-Radar Multi-Sensor (MRMS) QPE integrates radar QPE, gauge QPE, local gauge bias-corrected radar QPE, and gauge orographic precipitation climatology QPE. MRMS QPE has a 2-min time interval in each 0.01° × 0.01° grid box. One improvement to the previous QPE product (Q2) is that MRMS uses the most advanced dual-polarimetric (DP) radar technologies to eliminate nonmeteorological echoes, and it also provides a more accurate spatial distribution of precipitation (Zhang et al. 2016). Despite the advantages of DP QPE in certain aspects, it does not necessarily provide an overall superior QPE than single-polarimetric (SP) QPE, such as stage IV. According to Cunha et al. (2015), DP QPE shows a higher MSE than stage IV estimates. However, stage IV also shows a decreased correlation with rain gauges with increasing rainfall threshold (greater than 5 mm h\(^{-1}\)) than DP estimates. In this study, we choose MRMS as a QPE reference for most statistical analysis.

### Table 1. Key model configurations for NU-WRF simulations. WRF refers to control run and COUP refers to coupled run in this study. Cumulus parameterization is only applied to outermost domain. Grid resolution refers to the innermost grid only.

<table>
<thead>
<tr>
<th>With LIS coupling</th>
<th>Forcing for surface initial condition</th>
<th>Nested</th>
<th>Cumulus parameterization for outer grid</th>
<th>Inner grid resolution (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRF</td>
<td>No</td>
<td>NAM</td>
<td>Yes</td>
<td>GD</td>
</tr>
<tr>
<td>COUP</td>
<td>Yes</td>
<td>Previous COUP forecast and stage VI</td>
<td>Yes</td>
<td>GD</td>
</tr>
<tr>
<td>WRF,9km</td>
<td>No</td>
<td>NAM</td>
<td>No</td>
<td>GD</td>
</tr>
<tr>
<td>WRF,3km</td>
<td>No</td>
<td>NAM</td>
<td>Yes</td>
<td>GD</td>
</tr>
<tr>
<td>WRF,BMJ</td>
<td>No</td>
<td>NAM</td>
<td>No</td>
<td>BMJ</td>
</tr>
</tbody>
</table>
Comparisons between stage IV and MRMS are also shown to give an uncertainty range between the observation datasets.

The model analysis is performed on the innermost grid (Fig. 2) of NU-WRF forecasts (40°–45°N, 97°–87°W), which output each hour with 1-km grid spacing. NAM forecasts are available at 3-hourly intervals with 12-km spatial resolution. Both 0000 and 1200 UTC forecasts are evaluated for a 48-h model integration period on each day for all modeling datasets from 1 May to 15 June 2015. All datasets are remapped onto the NAM grid and are intercompared at 3-hourly time intervals. Since the resolution of data makes a difference for intensity and spatial analysis, intercomparison on the same grid is required to ensure that all data are evaluated with the same methodology. However, smoothing data from high to low resolution will inevitably lose some finescale details. Spatial displacement is also a quite common error occurring in modeling datasets. A small shift in space will result in very low spatial correlation for the model. When the grid resolution becomes finer, these types of small shifts increase in frequency, which may cause noisier results. To avoid overpenalizing the spatial shift in model data, all data are averaged onto the NAM 12-km grid.

4. Evaluation of rainfall forecast

Figure 4 shows the accumulated precipitation from NU-WRF simulations and stage IV dataset from 1 May to 15 June 2013. NU-WRF shows a similar precipitation distribution as stage IV, with high accumulated precipitation over the region of Iowa and northern Illinois. However, NU-WRF has a higher peak value than stage IV over the region with high rainfall accumulation.

a. Rainfall time series

Figure 5 shows 3-hourly accumulated rainfall from two NU-WRF simulations (WRF and COUP), NAM, stage IV, and MRMS from 1 May to 15 June. Since there are several forecast cycles for each model, the value is calculated by averaging from all available forecasts at that time. The first 6-h simulations are considered the model spinup period and are thus removed from the analysis. The values in parentheses in Fig. 5 show the average precipitation over the whole period. Despite the overall overestimation comparing both stage
IV and MRMS, both NU-WRF runs capture the peak for most precipitation events. There are only slight differences between the two sets of NU-WRF runs; WRF produces slightly more rainfall than COUP. NAM is very close to the observations for average quantity (0.2 mm h\(^{-1}\)); however, it does not accurately replicate the individual events. From 16 May to 1 June, NAM tends to miss the peak and significantly underestimates the rainfall. However, NAM tends to overestimate the rest of the time. Overall, NAM has a better average rainfall over the whole period than NU-WRF; the NAM does not capture the timing and peak well for individual events.

Also shown in Fig. 5, the precipitation is brought to the area of interest in groups. There are seven wet periods that can be identified (Table 2), and two adjacent periods can be separated by at least one dry day (or with very light rain). Each precipitation period can be caused by a single convective/precipitating system or by a
succession of convective systems. Period I features a trough at 500 mb accompanied by a surface front, which is a typical characteristic for spring and cold season–type convections. Rainfall forecasts are generally accurate during period I, except that NU-WRF has overestimated rainfall compared with observations (Fig. 5). Strong upper-level support is also present for periods II, V, and VI, though no surface front or moisture boundaries are found during these periods. Rainfall forecasts are fairly accurate for these periods and successfully depict the timing of the observed peaks (Fig. 5). NU-WRF generally overestimated rainfall during these periods, and NAM has overestimated rainfall during periods V and VI. Three of these periods have either shortwave troughs (VII) or a combination of short waves and longwave troughs (III and IV) and no sign of surface front, which brings weak yet complicated forcing to the region. This is especially illustrated in period IV, where one short wave comes one after another, associated with a series of propagating systems at the surface. There are also some diurnal signals identified during period IV (Fig. 5). NAM struggles to produce an accurate precipitation forecast during period IV (Fig. 5). In contrast, NU-WRF is able to reproduce most rainfall peaks during period IV. Despite using NAM for the initial and boundary conditions, the finer-resolution NU-WRF simulations reflect improved features in reproducing these precipitation events. These series of weakly forced events demonstrate the characteristics of warm season MCSs (e.g., Carbone et al. 2002).

b. Rainfall statistics

The 3-hourly accumulated rainfall is averaged on the innermost domain and the statistical scores are calculated from 1 May to 15 June. Figure 6 shows rainfall statistics scores that are computed by averaging forecasts from 6 to 48 h of each forecast cycle; thus, each value point represents the score of one forecast cycle. Both bias and RMSE are calculated from 3-hourly accumulated precipitation. Both NU-WRF and NAM underestimate rainfall amount during periods III, IV, and VII, while the models overestimate the rest of the periods. NAM shows large negative bias during period IV and positive bias during other periods, so the overall low bias of NAM is balanced out from positive and negative biases. For the RMSE, all models show a similar score \([-0.55 \text{ mm (3 h)}^{-1}\) compared with MRMS, which means the models capture the domain-averaged precipitation evolution fairly well. The spread of the RMSEs among models \([0.02 \text{ mm (3 h)}^{-1}\) are even smaller than that between stage IV and MRMS \([0.14 \text{ mm (3 h)}^{-1}\). Models produce heavy precipitation events during periods III and IV, when high RMSEs are also presented. Figure 6c shows the correlations between different datasets and MRMS. NAM performed poorly during period IV, which is also reflected in Fig. 5, as NAM does not capture the timing and the magnitude of precipitation peak. Another period with low correlation is from 11 to 13 May, when little rainfall is brought to the area (Fig. 5). Low correlation is also observed between stage IV and MRMS during this dry period, which is caused by disagreement on the occurrence of very light rainfall between the two datasets. The models also have poor correlations with MRMS during this period. The differences between correlation scores of the two NU-WRF simulations are small enough (0.01) that it is within the uncertainties between observations (0.05).

Figure 7 shows domain-averaged 3-hourly accumulated rainfall with respect to the forecast lead time. Only 0000 UTC forecast cycles are used in Fig. 7; 1200 UTC forecasts reflect similar characteristics to the 0000 UTC forecasts (not shown). The statistical scores are computed from domain-averaged rainfall from all forecast cycles with respect to each forecast hour. The bias scores (Fig. 7a) are negative for WRF, COUP, and NAM in the first 6 h, which is due to the cold start mode, where all the precipitation values are initialized from zero. The NAM simulation has lower bias than the two NU-WRF simulations because of the cancellation of its underestimates (negative bias) during periods III and IV and overestimates (positive bias) during the other periods (Fig. 6). Despite the lower bias, NAM has higher RMSE \([0.73 \text{ mm (3 h)}^{-1}\) than the two NU-WRF
simulations [0.71 mm (3 h)^{-1}]. NAM also has a much lower correlation (0.64) compared with the two NU-WRF simulations (0.75 and 0.74). Both NU-WRF simulations have relatively high correlations with MRMS, which echoes features in Fig. 5 that NU-WRF has captured domain-averaged precipitation amount for individual events. The correlation trend between models and MRMS decreases slightly with each increase of forecast lead time and is expected as forecast skill decreases with time. The margin of the correlation between stage IV and MRMS (0.02) shows the uncertainty between the two datasets. The differences between the two NU-WRF simulations are small enough (0.01) compared to the margin between stage IV and MRMS, and WRF has slightly better correlation than COUP. It is also noted that the differences between the two NU-WRF correlations increase with forecast lead time, which reflects the initial small differences between the two models growing with time.

When considering spatial variability between different datasets and MRMS (Fig. 8), the spatial correlation scores are much lower than the area-averaged quantities (Fig. 7c). The spatial correlation is computed from a spatial map of precipitation every 3 h and averaged from all forecast cycles with the same lead time. The correlation between stage IV and MRMS decreases from 0.98 for area-averaged rainfall to 0.82 for spatial correlations. The score drops significantly for correlations between models and MRMS. For example, the correlation changes from 0.73 previously (Fig. 7c) to 0.16 (Fig. 8a) for COUP. Figure 8 also demonstrates that the forecast skills decrease with an increase of forecast hours. Despite the low spatial correlation, NU-WRF produces consistently higher correlation than the NAM forecasts, which demonstrates that NU-WRF produces a better rainfall spatial pattern than NAM.

Table 3 shows the spatial correlation scores for all seven different forecast periods. NAM shows consistently
lower scores than both NU-WRF forecasts, which echoes results from Fig. 8 that reflect that NAM has an overall lower correlation than NU-WRF for all the forecasts. The lower spatial score of NAM indicates a poor match of point-to-point precipitation with the observations. This particular spatial correlation identifies two error types associated with quantity and location. One caveat of this analysis is that the displacement error could be double penalized in some cases. Considering the situation where the simulated precipitation area has perfect intensity, shape, and orientation, but the location is shifted a few grid points off from the observed one, the spatial correlation may not reflect the excellent forecast; rather, it will give a low score because of the spatial displacement. Similar situations could be found in the case of spurious precipitation regions. There is some benefit to averaging the rainfall to a coarser grid so that the finescale spatial displacement error may be filtered out, such as what is presented in this study. By evaluating the datasets on a common 12-km grid, we are focusing more on the location and intensity of bigger systems, rather than the location of small thunderstorms. Still, the correlation may not necessarily reflect all aspects of the rainfall forecast for individual time slice. As shown in Fig. 9, NAM does not capture the heavy precipitation over eastern Iowa, but it has a higher correlation (0.3) with MRMS than WRF (0.25). With a relatively big sample size, the mean correlation is capable of describing the general characteristics of the datasets.

In Table 3, the mean spatial correlation scores also vary with different periods for each forecast. NAM has the lowest correlation during period IV (0.07), while NU-WRF did not show a significant reduction in correlation during this period (0.18). As mentioned in the

![Fig. 7. Time series of (a) bias [3-hourly rainfall; mm (3 h)^-1], (b) RMSE [3-hourly rainfall; mm (3 h)^-1], and (c) correlation for domain-averaged rainfall from WRF, COUP, NAM, and stage IV compared against MRMS with respect to forecast hours. All modeling datasets are from 0000 UTC forecasts.](image-url)
previous section, the synoptic setup in period IV features a session of shortwave troughs in the mid- to upper-level troposphere, accompanied by complex forcing at the surface, which is a strong characteristic in the warm season precipitation system. NAM has produced very low spatial correlation (0.07) with the observations, suggesting very poor forecast skills in period IV. Both NU-WRF forecasts produce higher spatial correlation (0.18) than NAM during period IV. The distinctive performances between NU-WRF and NAM during period IV are also prominent in Fig. 8c. NAM displays much lower correlation than NU-WRF throughout the 48-h forecast period during period IV (Fig. 8c) than in period III, when a relatively strong synoptic forcing prevails. The NU-WRF score at period IV did not drop significantly for either NU-WRF forecasts, suggesting the forecast skill is not as sensitive as NAM to the change of weather regimes.

The two NU-WRF runs have very similar correlations compared with each other, but their differences grow with time. As shown in Fig. 7c, there is little obvious difference between the two NU-WRF runs until 24 h into the forecast. It is a question of whether the spread between the two runs is caused by a systematic error that essentially resulted from model physics or by a random error that was caused by a small initial perturbation accumulated with time. In Figs. 6–8, the differences between the two NU-WRF runs fluctuate with time. In the end, the forecasts do not drift from each other, with very similar mean

<table>
<thead>
<tr>
<th>Period</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRF</td>
<td>0.22</td>
<td>0.20</td>
<td>0.15</td>
<td>0.18</td>
<td>0.11</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>COUP</td>
<td>0.22</td>
<td>0.21</td>
<td>0.15</td>
<td>0.18</td>
<td>0.11</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>NAM</td>
<td>0.20</td>
<td>0.17</td>
<td>0.10</td>
<td>0.07</td>
<td>0.09</td>
<td>0.13</td>
<td>0.07</td>
</tr>
</tbody>
</table>
values. There is no simple relationship between the rainfall statistics from the two NU-WRF runs, indicating no superiority of either model configuration. A systematic evaluation of this error behavior should be conducted in the future, but is beyond the scope of this study.

c. Rainfall time series PDF

The precipitation probability density function (PDF) is also evaluated for model and observational datasets. Since the resolution of the dataset affects the PDF of precipitation, all datasets are averaged onto the NAM grid for comparison. As shown in Fig. 10, NU-WRF compares very well with the observations for 3-hourly accumulated rainfall, but it slightly overestimates the frequencies for medium to heavy rainfall (2–16 mm). Considering the noticeable differences between stage IV and MRMS, the differences between NU-WRF and the observations are generally within the range of observational uncertainty. In contrast, NAM has produced outstandingly high frequencies for very light precipitation (0.25–0.5 mm), while significantly underestimating the frequencies for heavy rainfall (>4 mm). This feature is also evident from rainfall spatial distribution (Fig. 9). NAM produces a large light rain area, but fails to produce any heavy rainfall. It is an expected outcome, since the 12-km grid spacing in NAM limits the model’s capability to resolve cloud-scale convolutions. In this case, the coarse-resolution model relies on convective parameterizations to produce rainfall based on assumptions and empirical tuning for different regions and resolutions. Wang and Seaman (1997) suggested that convective precipitation produces more than 40% of the total rainfall for the 12-km model forecast in 24-h integration time, which leaves a large uncertainty in predicted rainfall that comes from convective parameterization. The small differences between the two NU-WRF runs indicate that different land surface initialization and differences within the LSM do not have a big impact on precipitation intensity. In contrast, the rainfall PDF is shown to be more sensitive to

FIG. 9. The 3-hourly rainfall accumulation (mm) for (a) MRMS, (b) NAM, (c) COUP, and (d) WRF at 0300 UTC 20 May 2013, which is 15 h since model initialization (at 1200 UTC 19 May 2013). The correlation between model and MRMS is shown in the lower-left corner of (b)–(d).
different cloud microphysics schemes (e.g., Tao et al. 2016; Wu et al. 2013).

Overall, there are noticeable differences between two NU-WRF simulations and NAM forecasts. Considering the WRF (i.e., the uncoupled run) and NAM, since the evaluation is performed on the same 12-km grid, the differences in forecasts are not due to data resolution (e.g., 12 or 1 km). WRF (the uncoupled run) uses NAM to provide initial and boundary conditions, which further excludes the possibility of different initial and boundary conditions to cause various forecast results. Therefore, the different rainfall forecast results between NAM and WRF are due to intrinsic differences between the two models, such as different physics parameterizations and model resolution.

5. Discussion

a. Sensitivity tests

NU-WRF and NAM predict very different rainfall characteristics (shown in section 4), especially during period IV. As mentioned previously, grid resolution and convective parameterization play a big role in warm season rainfall forecast. It is a question as to what extent model grid resolution and convective parameterization can affect simulated rainfall from NAM and NU-WRF. There are many other differences between the two models, such as different dynamical cores (i.e., ARW and NMM) and physics parameterizations that can lead to different results. Because of this complexity, it is important that the tests should be performed on the same modeling framework with identical model configuration and physical parameterizations, except for only varying grid resolution or convective parameterizations. The simulation period is selected during period IV, from 0000 UTC 19 May to 0000 UTC 31 May, when both NU-WRF and NAM forecasts are relatively poorer than other times (Fig. 5). The domain of interest is under weak forces, and warm season rainfall characteristics are evident during this time (Table 2).

Three sensitivity tests (WRF_9km, WRF_3km, and WRF) are performed to investigate the impact of grid setup and resolution on simulated precipitation (Table 1). As shown in Table 1, all three simulations with varying resolutions do not have LIS coupling and have the same physics parameterization. WRF represents the original NU-WRF forecast with 9-, 3-, and 1-km grid configuration (Fig. 2), while WRF_3km employs the two outer domains and WRF_9km uses only the outermost domain. In other words, the WRF_9km simulation tests how well the model performs without the two inner domains, and the WRF_3km adds the second domain on the basis of WRF_9km and with complexity in between WRF and WRF_9km. All three simulations use the GD ensemble cumulus scheme for the outermost grid (with 9-km resolution). From Fig. 11, WRF_3km and WRF produce similar amounts of precipitation and the same timing for peak precipitation, while WRF_9km is completely out of phase with the two finer-resolution runs. However, the timing of the peak in WRF_9km closely resembles NAM, both of which occur at 2100 UTC 29 May. Unlike NAM, WRF_9km is able to develop a more prominent peak, which is more comparable with the observations. It is widely recognized that many types of convections can be explicitly resolved with fine grid resolution (less than 4 km) with the absence of convective parameterization (Weisman et al. 2007).
However, the 9-km simulation relies on convective parameterization due to insufficient resolution to resolve subgrid-scale turbulence; thus, significant changes have been seen in the rainfall forecast (Fig. 11). Moving grid resolutions from 3 to 9 km generates more variabilities in rainfall field than surface initialization (Fig. 11, WRF and COUP).

As mentioned above, the predicted precipitation is very sensitive to different convective schemes applied in the forecast (e.g., Wang and Seaman 1997; Gallus 1999). To test the sensitivity of different convective schemes, the 9-km run with BMJ cumulus scheme (WRF_BMJ) is used to compare with the 9-km run with the GD cumulus scheme (WRF_9km; Table 1). The BMJ scheme has been adopted for operational application at the National Centers for Environmental Prediction (NCEP) and is used in the NAM forecast, while GD is utilized in NU-WRF simulations in this study. As shown in Fig. 11, WRF_BMJ has a smaller peak magnitude compared with WRF_9km. At the same time, WRF_BMJ produces rainfall time series that are the most comparable to NAM in all the simulations. GD and BMJ have very different assumptions. GD has 144 members that are an ensemble of mass-flux schemes, consisting of many different triggers (moisture convergence, CAPE removal, upward vertical motion, etc.) and parameters (entrainment rates, precipitation efficiency, etc.). For a given grid point, all members are averaged and output is sent back to the model. BMJ is an adjustment scheme that seeks to relax unstable soundings toward predetermined reference profiles. Regional model results indicate that the trigger function in the BMJ scheme may prevent and reduce convective activity over the SGP region, which results in reduced precipitation amount (Gochis et al. 2002). However, the performance of convective schemes is also quite sensitive to different weather regimes, grid resolutions, and combinations with other parameterization schemes, as suggested in many previous studies (e.g., Wang and Seaman 1997; Gallus 1999).

Both grid resolution and the choice of convective schemes have a strong influence on simulated rainfall amount. The forecasted rainfall amount does not show significant sensitivity to surface initialization. By reducing the model resolution and changing the convective scheme, NU-WRF simulation shows a similar precipitation peak and magnitude compared with the NAM forecast, which explains the main sources for different precipitation forecasts in the two models.

b. Soil moisture and rainfall

Of particular interest in this study is whether there are improvements in rainfall forecast by applying high-resolution and more accurate land surface initial conditions rather than interpolated fields from regional model forecast. As shown in previous sections, the differences in the rainfall forecasts between the two NU-WRF runs are rather small. One possible reason for such small differences is that the region of interest is under the influence of many heavy precipitation events during the campaign period; thus, with high soil moisture, the moisture transport from the surface is similarly high in both models. One indication for the above statement is the high evaporative fraction (EF), which is the ratio of latent heat to available energy at the land
surface. EF is a diagnostic parameter for the surface energy balance (energy-limited state or a moisture-limited state) that isolates soil moisture and vegetation from radiation and turbulent factors. Despite the strong diurnal periodicity for latent heat and sensible heat in the surface energy balance, EF is generally considered to be a constant during daytime hours (Nichols and Cuenca 1993; Crago 1996; Crago and Brutsaert 1996). Figure 12a shows the daily EF from the NU-WRF run, which is averaged from 0700 to 1800 local time. EF stays over 0.6 and gradually increases to 0.8 at the end of the campaign, which means that the energy fluxes to the surface energy budget are mainly contributed by latent heating. With high EF, the impact for precipitation processes with different soil initialization is minimized. WRF shows slightly higher EF (Fig. 12a) than COUP corresponding to higher top-layer soil moisture (Fig. 12b), which is also a persistent feature throughout the campaign period. Overall, both EF and the top-layer soil moisture are almost on top of each other for two NU-WRF runs.

It is well known that soil moisture has a relatively greater impact on precipitation under drier conditions. To illustrate the detailed soil moisture and EF structure, we chose one of the dry days as an example. The date 19 May is one of the drier days with less top-layer soil moisture; higher spatial variability is thus expected in both forecasts. Figures 13a–c show the evaporative fraction at local noon from an NU-WRF forecast initialized at 1200 UTC 19 May 2013. Figures 13d and 13e show the 0–10-cm-layer soil moisture at model initialization for the same forecast. Both models show very similar EF spatial distribution. The WRF soil moisture map shows lower spatial variability than COUP at model initialization. As previously mentioned, the smaller spatial heterogeneity for WRF comes from the interpolated soil moisture field from NAM that has a 12-km resolution, while COUP uses LIS analysis with a resolution that is consistent with 1-km NU-WRF grid. In addition, COUP uses stage IV–observed rainfall in LIS analysis that should result in a more accurate soil moisture map than WRF. The slightly drier top-layer soil moisture from COUP is also observed in Georgia and South Carolina during the summer season, which is actually closer to the observed soil moisture from the U.S. Department of Agriculture’s Soil Climate Analysis Network (SCAN; Case et al. 2011). The small EF differences reveal a similar partition in surface energy budget for both WRF and COUP. Figures 9c and 9d show the rainfall maps at 0300 UTC 20 May, which is 15 h since model initialization. Only small differences in precipitation forecasts are observed between the two NU-WRF runs for the forecast cycle at 1200 UTC 19 May, which is one of the relatively drier days during the field campaign. Despite the fine resolution and more accurate surface initialization in COUP, the high EF indicates similar surface moisture and energy transport to the atmospheric boundary layer in both
WRF and COUP, which minimizes the impact of surface initialization on precipitation processes.

6. Conclusions

Two sets of NU-WRF are used to provide 48-h real-time forecasts twice a day during the IFloodS field campaign from 1 May to 15 June 2013. One of the NU-WRF forecasts uses NAM interpolated land surface field as LSM forcing, while the other uses LIS spinup to provide land surface conditions, which assimilates the latest stage IV observed precipitation. These two sets of model datasets are compared with a low-resolution forcing dataset (NAM) and are evaluated with stage IV and MRMS.

Both NU-WRF simulations are able to reproduce the individual precipitation event during the field campaign period, while NAM is out of sync with observations for heavy precipitation events during period IV. In addition, NAM tends to overestimate the rainfall amount for light precipitation events (less than 1 mm h\(^{-1}\)) but underestimate rainfall for heavy precipitation events (e.g., 20 May, 25–30 May, and 13 June). Despite NAM forecasts’ better average rainfall over the 6-week period compared with NU-WRF, the NAM forecast skill is not necessarily better for individual events. NU-WRF is also able to produce a better rainfall PDF than NAM. NAM significantly underestimates the frequencies for heavy rainfall and largely overestimates frequencies for very light rainfall. While NU-WRF is able to produce PDF that is very close to the observed distribution from stage IV, the differences between NU-WRF and stage IV are within the differences between stage IV and MRMS. Overall, NU-WRF outperforms NAM in both time and spatial correlations with MRMS and PDF shape for rainfall forecast.

A series of NU-WRF sensitivity tests are conducted to further investigate the differences between NU-WRF and NAM results. The study shows that by reducing NU-WRF resolution to a coarser grid, the rainfall forecast skill is reasonably reduced, especially when the grid changes from 3 to 9 km. No significant differences are found in 1- and 3-km rainfall forecasts, which echoes previous studies that 1–4 km is generally considered to be the cloud-resolving scale, and no convective scheme is needed for such a scale. Additional tests were performed to demonstrate the impact of the choice of convective parameterization on precipitation forecast. By switching convective schemes from GD to BMJ, which is also the scheme adopted by NAM operational
forecasts, the NU-WRF forecasted rainfall amount is comparable to NAM forecast.

High-resolution surface initialization has the advantage of producing greater spatial variability and more accurate surface properties than forecasts without LIS analysis. However, the benefit of such surface initialization for the precipitation forecast is marginal. Two sets of NU-WRF simulations do not yield significant differences during the IFloods field campaign period. Instead, the study shows both NU-WRF runs predicted rainfall with similar bias, RMSE, correlation, and PDF. The differences between the two NU-WRFs are much smaller than the differences between NAM and NU-WRF or between the two observational datasets (stage IV and MRMS). Evaporative fraction indicates the relatively similar land surface moisture transport between the two NU-WRF simulations, which inhibit the degree of improvement in precipitation forecast for high-resolution land surface initialization.

Acknowledgments. This research was supported by the NASA Precipitation Measurement Missions (PMM) and Modeling and Analysis Program (MAP) solicitations through awards to PIs Peters-Lidard, Tao, and Petersen. This support is gratefully acknowledged. We also acknowledge the excellent computational and storage support provided by NASA’s Center for Climate Simulation (NCCS). Finally, the authors are grateful for the constructive comments from three anonymous reviewers.

REFERENCES


