Reproducing the September 2013 Record-Breaking Rainfall over the Colorado Front Range with High-Resolution WRF Forecasts

CRAIG S. SCHWARTZ
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 22 November 2013, in final form 13 January 2014)

ABSTRACT

Four convection-permitting Weather Research and Forecasting Model (WRF) forecasts were produced in an attempt to replicate the record-breaking rainfall across the Colorado Front Range between 1200 UTC 11 September and 1200 UTC 13 September 2013. A nested WRF domain with 4- and 1-km horizontal grid spacings was employed, and sensitivity to initial conditions (ICs) and microphysics (MP) parameterizations was examined. Rainfall forecasts were compared to gridded observations produced by National Weather Service River Forecast Centers and gauge measurements from the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS). All 1-km forecasts produced 48-h rainfall exceeding 250 mm over portions of the Colorado Front Range and were more consistent with observations than the 4-km forecasts. While localized sensitivities to both ICs and MP were noted, systematic differences were not attributable to the varied ICs or MP schemes. At times, the 1-km forecasts produced precipitation structures similar to those observed, but none of the forecasts successfully captured the observed mesoscale evolution of the entire rainfall event. Nonetheless, as all 1-km forecasts produced torrential rainfall over the Colorado Front Range, these forecasts could have been useful guidance for this event.

1. Introduction

Between 9 and 16 September 2013, unprecedented rainfall across portions of the Colorado Front Range produced widespread flooding. While flooding extended from the Wyoming border southward to Colorado Springs, the heaviest rain fell near Boulder, Colorado, where the official event total precipitation exceeded 430 mm (data available online at http://www.ncdc.noaa.gov). At least eight flood-related fatalities were reported in Colorado and economic losses due to flooding may exceed $2 billion (EQECAT 2014).

Most of the flood-producing rainfall occurred during the 48-h period between 1200 UTC 11 September and 1200 UTC 13 September. During this period, meteorological conditions were conducive for heavy rain, with low-level upslope winds; a diffluent, subtropical southwesterly flow aloft; and record-setting precipitable water for the month of September (http://www.crh.noaa.gov/unr/?n=pw). Given these ingredients, forecasters predicted heavy precipitation over the Front Range, but the observed rainfall far exceeded forecasters’ expectations. This underprediction may have been related to reliance on operational numerical weather prediction (NWP) model guidance with parameterized convection, including the Global Forecast System (GFS), North American Mesoscale Forecast System (NAM), and Rapid Refresh (RR) models, which also severely underpredicted the storm-total precipitation (discussed in section 3).

However, numerous studies have demonstrated that high-resolution, convection-allowing NWP models produce better forecasts of heavy precipitation than coarse-resolution models with parameterized convection (e.g., Colle and Mass 2000; Done et al. 2004; Colle et al. 2005; Kain et al. 2006; Lean et al. 2008; Roberts and Lean 2008; Weisman et al. 2008; Schwartz et al. 2009; Clark et al. 2010). Thus, this paper examines the utility of convection-permitting Weather Research and Forecasting Model (WRF) forecasts as guidance for this event by assessing whether the extreme rainfall between 1200 UTC 11 September and 1200 UTC 13 September over the Colorado Front Range could be successfully...
predicted. Note that this work focuses solely on certain NWP modeling aspects of the flood, whereas studying this case from an observational perspective and using an ingredients-based approach would be quite informative.

The next section describes the model configurations. Results are presented in section 3 before concluding in section 4.

2. WRF configurations

Weather forecasts were produced by version 3.3.1 of the nonhydrostatic Advanced Research WRF (Skamarock et al. 2008) model over a nested computational domain (Figs. 1a,b). The horizontal grid spacing was 4 km (400 × 400 grid points) in the outer domain and 1 km (801 × 801 grid boxes) in the inner nest. Both domains had 40 vertical levels and a 50-hPa top. The time step was 16 s in the 4-km domain and 4 s in the 1-km nest. Lateral boundary conditions (LBCs) for the 4-km domain were provided by GFS forecasts every 3 h. The 4-km domain provided LBCs for the 1-km nest, but additional interaction between the domains was not permitted to prevent the 1-km forecast from impacting the 4-km forecast.
Four 48-h WRF forecasts were configured with varying microphysics (MP) parameterizations and initial conditions (ICs), as convection-allowing models have demonstrated sensitivity to both of these aspects (e.g., Weisman et al. 2008; Schwartz et al. 2010; Bryan and Morrison 2012; Adams-Selin et al. 2013; Schwartz and Liu 2014). Two forecasts were initialized by interpolating the 0.5° × 0.5° 1200 UTC 11 September 2013 GFS analysis onto the nested computational domain, while another two were initialized by interpolating the ~12-km 1200 UTC 11 September 2013 NAM analysis onto the domain. Within these GFS- and NAM-initialized forecast sets, one experiment used the Morrison MP parameterization (Morrison et al. 2009) and the other the Thompson MP scheme (Thompson et al. 2008). The NAM-initialized forecasts also used GFS LBCs so that differences between forecasts were attributable solely to MP and ICs.

All other model settings and the following physical parameterizations were constant across the forecasts: the Rapid Radiative Transfer Model for Global Climate Models (RRTMG; Mlawer et al. 1997; Iacono et al. 2008) longwave and shortwave radiation schemes with aerosol (Tegen et al. 1997) and ozone climatologies, the Mellor–Yamada–Janjić (MYJ; Mellor and Yamada 1982; Janjić 1994, 2002) planetary boundary layer scheme, and the Noah land surface model (Chen and Dudhia 2001). No cumulus parameterization was used and positive-definite moisture advection (Skamarock and Weisman 2009) was employed.

3. Results

Precipitation forecasts were compared to gauge measurements from the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS; Cifelli et al. 2005; http://www.cocorahs.org) and analyses produced from National Weather Service (NWS) River Forecast Centers (RFCs; http://www.srh.noaa.gov/ridge2/RFC_Precip), CoCoRaHS observations are reported as 24-h precipitation accumulations while RFC observations are produced hourly on a ~4-km grid. These hourly RFC grids are similar to stage IV (ST4) analyses produced at the National Centers for Environmental Prediction (NCEP; Lin and Mitchell 2005) but potentially contain updates (J. Paul, NWS/ABRFC, 2013, personal communication), and they were more accurate for this case than hourly ST4 products. Comparison of the RFC and CoCoRaHS 48-h accumulated rainfall ending 1200 UTC 13 September reveals good correspondence (Fig. 1c).

Figure 2 shows forecasts of 48-h accumulated precipitation ending 1200 UTC 13 September from the GFS (interpolated onto a 0.5° × 0.5° grid), NAM (native ~12-km grid), RR (native ~13-km grid), and WRFs, as well as from the National Oceanic and Atmospheric Administration’s (NOAA) experimental High-Resolution RR (HRRR) model, which is a 3-km convection-permitting version of the Advanced Research WRF.2 These forecasts are overlaid with CoCoRaHS observations of 48-h rainfall exceeding 25 mm. Around Boulder and Estes Park, the operational and HRRR models produced precipitation <100 mm (Figs. 2a–d). Near Boulder, the NAM forecast was poorest. While the 4-km WRF forecasts (Figs. 2e–h) generated more precipitation than the operational and HRRR models near Boulder and Estes Park, the observed rainfall totals in these areas were not reproduced. There was limited 4-km sensitivity to MP, with the largest difference between the two GFS-initialized WRF configurations west and west-northwest of Boulder, where the Thompson MP scheme seldom produced rainfall amounts exceeding 150 mm while the Morrison MP scheme generated rainfall greater than 200 mm (Figs. 2e,f). Both NAM-initialized 4-km WRF forecasts (Figs. 2g,h) produced 48-h rainfall totals comparable to the 4-km GFS-initialized forecast with Thompson MP.

However, the 1-km forecasts generated substantially heavier 48-h rainfall than the 4-km forecasts (Figs. 2i–l). All 1-km forecasts yielded rainfall maxima exceeding 250 mm across parts of the Front Range that were closer to CoCoRaHS observations than the 4-km rainfall totals. The NAM-initialized forecast with Thompson MP (Fig. 2k) produced noticeably less rainfall around and between Boulder and Estes Park and the GFS (NAM)-initialized forecast with Thompson (Morrison) MP overpredicted in some areas (Figs. 2i,j). However, systematic behaviors attributable to MP were not detected. For example, considering the GFS-initialized forecasts, when Morrison MP was used, maximum 48-h rainfall amounts were reduced around Boulder and Estes Park compared to when Thompson MP was employed (Figs. 2i,j). Conversely, within the NAM-initialized forecasts

---

1 The HRRR model is expected to become fully operational at NCEP in 2014 (http://ruc.noaa.gov/hrrr).

2 The RR and HRRR models are initialized hourly and produce 18- and 15-h forecasts, respectively. To produce 48-h rainfall totals for the RR, the accumulated precipitation from forecasts initialized at 1200 UTC 11 September, 0600 UTC 12 September, and 0000 UTC 13 September were summed. The HRRR 48-h precipitation was produced by summing precipitation forecasts initialized at 1200 UTC 11 September, 0300 UTC 12 September, 1800 UTC 12 September, and 0900 UTC 13 September. The RR and HRRR forecasts initialized throughout the 48-h period varied substantially and sometimes produced heavier total rainfall than shown in Fig. 2. Thus, caution is urged when comparing the RR and HRRR 48-h rainfall accumulations with the WRF, NAM, and GFS forecasts presented herein.
Morrison MP produced substantially heavier precipitation northwest of Boulder than did Thompson MP. Some sensitivity to ICs was also noticed, as both GFS-initialized 1-km forecasts predicted heavy rainfall near Colorado Springs in agreement with observations that the NAM-initialized forecasts did not produce. No experiment predicted the localized heavy precipitation east of Denver.

48-hr accumulated precipitation (mm) between 1200 UTC 11 and 1200 UTC 13 September

(Figs. 2k,l), Morrison MP produced substantially heavier precipitation northwest of Boulder than did Thompson MP. Some sensitivity to ICs was also noticed, as both GFS-initialized 1-km forecasts predicted heavy rainfall near Colorado Springs in agreement with observations that the NAM-initialized forecasts did not produce. No experiment predicted the localized heavy precipitation east of Denver.
The operational, HRRR, and WRF 48-h accumulated precipitation forecasts were bilinearly interpolated onto the RFC grid and compared to the corresponding RFC observations between 39° and 41°N and 104° and 106°W, the region of heaviest rainfall. The fractions skill score (FSS; Roberts and Lean 2008) and multiplicative bias were calculated for various events defined as precipitation exceedance of certain thresholds $q$. Bias
is the number of forecast events divided by the number of observed events while the FSS employs a neighborhood approach (e.g., Ebert 2009) that relaxes the requirement that forecast and observed events match at the grid scale for a forecast to be considered perfect. Circular neighborhoods with radii of ~8, 20, and 40 km were used for the FSS. An FSS of 1 is perfect, while FSS = 0 indicates no skill.

The HRRR, operational, and 4-km WRFs underpredicted rainfall at all thresholds (Fig. 3a), although the 4-km underprediction was less severe for $q$ between ~75 and 150 mm (48 h)$^{-1}$. Conversely, the 1-km forecasts had biases near 1 for $q \leq 100$ mm (48 h)$^{-1}$ but the GFS- and NAM-initialized forecasts with Thompson and Morrison MP, respectively, had biases $\geq 1.8$ for $q \geq 200$ mm (48 h)$^{-1}$ (although care should be taken when interpreting the high biases at these thresholds due to small sample sizes). For all $q$, the HRRR model had higher FSSs than the operational models (Figs. 3b–d). However, for $q \geq 100$ mm (48 h)$^{-1}$ and all neighborhood sizes, all 4-km WRF forecasts had higher FSSs than the operational and HRRR models and FSSs were highest for the 1-km forecasts. The GFS-initialized 1-km forecast with Morrison MP consistently had the highest FSSs for $q \geq 200$ mm (48 h)$^{-1}$ and the NAM-initialized forecast with Thompson MP usually had the lowest 1-km FSSs. These statistics corroborate subjective impressions that the 1-km forecasts were better than the 4-km forecasts, which in turn, improved upon the operational and HRRR guidance.

To assess the temporal accuracy of the rainfall evolution, the total accumulated precipitation was computed each hour for the 1-km forecasts and RFC observations between 39° and 41°N and 104° and 106°W and normalized by the appropriate number of grid boxes over this area (Fig. 4). Observations reveal two distinct rainfall intensity peaks around 0700 UTC 12 September and 0300 UTC 13 September. Both 1-km NAM-initialized forecasts captured this bimodal pattern, although the maxima were earlier than those observed. The GFS-initialized 1-km forecasts depicted the timing of the first rainfall peak comparably to the NAM-initialized forecasts, but the second maximum was forecast differently depending on the MP scheme. The GFS-initialized WRF forecast with Thompson MP produced three distinct maxima after 1800 UTC 12 September, while the GFS-initialized forecast with Morrison MP better captured the second peak. It appears that the 1-km NAM-initialized forecast with Thompson MP produced lesser amounts of total rainfall than the other 1-km forecasts due to its comparatively lighter precipitation between ~0900 and 2000 UTC 12 September.

Forecast precipitation structures from the 1-km Thompson MP configurations during and surrounding the two rainfall maxima were also examined. At 0300 UTC 12 September, a band of rainfall stretched from Boulder eastward (Fig. 5a) in association with a low-level vorticity maximum near Denver (not shown). While both 1-km forecasts produced rainfall around Boulder at this time (Figs. 5e,i), the east–west-banded mesoscale feature was not replicated. However, between 0600 and 1200 UTC 12 September, the 1-km precipitation patterns more closely resembled the observations (Figs. 5b–d, f–h, and j–l), although there were many disparities regarding the details. Evidence of a low-level vorticity maximum near Denver was apparent in both simulations as spiraling rainfall motion was noted between 0600 and 1200 UTC 12 September. During the second observed rainfall maximum, rainfall rates increased from 2100 UTC 12 to 0300 UTC 13 September near Boulder before decreasing by 0600 UTC 13 September (Figs. 6a–d). However, during this period, the precipitation intensity around Boulder from the GFS-initialized WRF forecast with Thompson MP was erratic (Figs. 6e–h). Conversely, while rainfall rates from
the NAM-initialized forecast with Thompson MP increased near Boulder from 2100 UTC 12 September to 0000 UTC 13 September (Figs. 6i,j), rainfall intensity did not taper appreciably after 0300 UTC 13 September (Figs. 6k,l). It is interesting that this second period of heavy rainfall featured more “banded” and fewer cellular forecast structures than during the first rainfall maximum. This finding suggests greater atmospheric stability during
the second period, and the convective available potential
energy averaged between 39° and 41°N and 104° and
106°W was indeed lower during the second intense rain-
fall period (not shown). However, more exploration is
needed to understand the physical processes producing
these varied finescale structures.

4. Discussion and summary

Both the 1200 UTC 11 September GFS and NAM
analyses provided ICs for two 48-h WRF forecasts with
nested 4- and 1-km domains. Within the GFS- and
NAM-initialized sets, one forecast used the Thompson
MP scheme and the other the Morrison MP parameterization. The forecasts were evaluated with a focus on rainfall prediction over the Colorado Front Range between 1200 UTC 11 September and 1200 UTC 13 September, the period of heaviest precipitation during the record-breaking rainfall event.

All 1-km forecasts objectively produced more accurate 48-h rainfall accumulations than the 4-km forecasts. At times, the 1-km simulations reproduced some of the observed precipitation structures, but the forecasts were unsuccessful at capturing the mesoscale precipitation evolution throughout the 48-h period. While the NAM-initialized forecasts more accurately predicted the two observed rainfall intensity peaks than the GFS-initialized forecasts, broader systematic differences attributable to varying ICs or MP were not observed. As differences between the 1-km 48-h precipitation forecasts usually occurred at mesogamma scales, where noise may contribute substantially to errors (e.g., Lorenz 1969), it is difficult to attribute these localized differences to meaningful forcings. Clearly, differences between the four 1-km forecasts were minor compared to those between the 1- and 4-km forecasts.

These results corroborate previous studies that suggested rainfall forecasts may benefit from grid spacing finer than 4 km in topographically diverse areas (e.g., Colle and Mass 2000; Colle et al. 2005; Garvert et al. 2005; Lean et al. 2008; Roberts and Lean 2008; Smith et al. 2014). However, there are situations when 1- or 2-km forecasts do not produce more rainfall or better forecasts than 4-km simulations (e.g., Grubišić et al. 2005; Kain et al. 2008; Schwartz et al. 2009; Bryan and Morrison 2012; Johnson et al. 2013), particularly over flatter terrain, which indicates the heavier 1-km rainfall forecasts in this case were meaningful. As this study is limited by its small sample size, caution should be exerted when attempting to generalize the findings herein to other cases.

Testing of additional MP schemes and non-MP physical parameterizations is warranted for this event. Additionally, regional data assimilation techniques may improve the ICs and yield forecasts with a more consistent temporal and spatial representation of the observed precipitation structures. Despite some shortcomings of the 1-km simulations, this study demonstrates that 1-km WRF forecasts can produce precipitation totals that reasonably agree with the record-setting observations between 11 and 13 September over areas of the Colorado Front Range. While these 1200 UTC initialized forecasts would only have been available to forecasters a short time before the heaviest precipitation developed, an auxiliary WRF forecast employing Thompson MP but initialized from the 0600 UTC 11 September GFS analysis produced 48-h accumulated rainfall ending 1200 UTC 13 September similar to the corresponding 1200 UTC initialized forecast, although the heaviest precipitation was shifted slightly north (not shown). Nonetheless, this finding indicates at least some “run to run” consistency and suggests the signal for heavy rainfall was not restricted to forecasts initialized at 1200 UTC 11 September. It is encouraging that this simple initialization procedure of interpolating operational analyses onto a high-resolution WRF grid yielded successful forecasts, and the four 1-km forecasts initialized at 1200 UTC 11 September, viewed either individually or collectively as a “poor man’s ensemble,” would have been valuable forecast guidance and may have increased forecaster confidence of an extreme precipitation event.

Acknowledgments. The WRF simulations were performed on the Yellowstone supercomputer (ark:/85065/d7wd3xhc), which is maintained by the Computational and Information Systems Laboratory (CISL) of NCAR. Thanks to Morris Weisman, Wei Wang, and Glen Romine (NCAR/MMM) for helpful discussions and review. Matthew Bunkers (NWS) and several anonymous reviewers provided constructive comments.

REFERENCES


