Performance of Operational Model Precipitation Forecast Guidance During the 2013 Colorado Front-Range Floods

Thomas M. Hamill

NOAA Earth System Research Lab, Physical Sciences Division
Boulder, Colorado

Submitted to *Monthly Weather Review*

Revised 5 February 2014

Corresponding author address:

Dr. Thomas M. Hamill
NOAA Earth System Research Lab
Physical Sciences Division
R/PSD1
325 Broadway
Boulder, CO 80305-3328
tom.hamill@noaa.gov
phone: (303) 497-3060
fax: (303) 497-6449

ABSTRACT

During the period of 9 - 16 September 2013, more than 17 inches (~ 432 mm) of rainfall fell over parts of Boulder County, CO, with more than 8 inches (~203 mm) over a wide swath of Colorado's northern Front Range. This caused significant flash and river flooding, loss of life, and extensive property damage. The event set a record for daily rainfall (9.08 inches, or > 230 mm) in Boulder that was nearly double the previous daily rainfall record of 4.8 inches (122 mm) set on July 31, 1919.

The operational performance of precipitation forecast guidance from global ensemble prediction systems and the National Weather Service's global and regional forecast systems during this event is documented here, briefly in the article and more extensively in online appendices. While the precipitation forecast guidance uniformly depicted a much wetter than average period over northeastern Colorado, none of the global nor most of the regional modeling systems predicted precipitation amounts as heavy as analyzed. Notable exceptions to this were the Short-Range Ensemble Forecast (SREF) members that used the Weather Research and Forecast (WRF) Advanced Research WRF (ARW) dynamical core. These members consistently produced record rainfall in the Front Range. However, the SREF's record rainfall was also predicted to occur the day before the heaviest actual precipitation as well as the day of the heaviest precipitation.

1. **Introduction**.

During the period from 12 UTC (6 AM Mountain Standard Time) 9 September (Sep) 2013 to 12 UTC 16 Sep 2013, gauge measurements showed that more than 17 inches (~ 432 mm) of rainfall fell in several areas of the Front Range of Northern Colorado, with the precipitation maximum nearly directly over the city of Boulder. A large area of > 8 inches (~203 mm) accumulated precipitation extended across a wide swath of the Front Range. The peak precipitation periods were the evenings of 11 and 12 Sep, though heavy rainfall also occurred on 9 and 15 Sep. Figure 1 provides a map of the estimated analyzed precipitation over Colorado and New Mexico, with additional panels showing the day-by-day accumulations in the northern Front Range and Denver metropolitan area. There were several areas with very heavy precipitation, with especially heavy rainfall also occurring in Aurora, CO just east of Denver, another small area of very heavy precipitation southwest of Colorado Springs, and extensive heavy rainfall in central and southern New Mexico. The largest impacts, though, were in the northern Front Range and later along river basins to the east. According to the Federal Emergency Management Agency, in their preliminary disaster declaration (through 30 Nov 2013), 1,500 houses were destroyed and ~19,000 damaged. Four hundred eighty-five miles of roadway were damaged, including most roads into the mountains in the northern Front Range, making many homes impossible to reach, except on foot. Thirty state highway bridges were destroyed and 20 severely damaged. Twenty-seven state dams sustained damage; 150 miles of railroad track were damaged. Nine people died as a result of the storms and flooding.

This article will analyze the performance of operational precipitation forecasts over the northern Front Range, especially Boulder County, though the maps herein will allow the reader to examine the performance of the models over larger regions. This article does not present new research; the purpose is simply to document the performance of the operational guidance available to forecasters at the time.[1] Because the forecasts were, for the most part, unexceptional, it is likely that this event will become a focus of intense study in the months and years to come. This article was written to document the performance of the operational models, as these may become a useful baseline for future comparison. Previously, some characteristics of global ensemble-mean predictions for this storm were examined in Lavers and Villarini (2014), including the ability of global ensemble systems to predict the accumulated precipitation more faithfully over the state of Colorado than over a 0.5-degree box around Boulder.

2. **Precipitation analysis data and the forecast models.**

Both "Stage IV" and "AHPS" (Advanced Hydrologic Prediction System) precipitation analysis data were used in this study. Each provides data on ~ 4-km grids over the contiguous US. Stage IV data is available at hourly and six-hourly

[1] Three online appendices accompany this article. Appendix A [insert URL] provides information on the forecast models and a brief synoptic overview of observed/analyzed conditions. Appendix B [insert URL] provides information on precipitation forecasts from global medium-range ensembles. Appendix C [insert URL] provides information on precipitation forecasts from the shorter-range prediction systems.

intervals, though more quality control is applied to the six-hourly data. The AHPS precipitation analyses are provided only every 24 h and agglomerates the four six-hourly Stage IV analyses. Generally, the procedure here was to use the most compact and highest- quality data available from these three sources whenever possible. Consequently, for accumulated precipitation forecast plots that span > 24-h periods, AHPS data was used as much as possible, supplemented by Stage IV six-hourly data and hourly data only when necessary. One instance where the use of hourly data was necessary was in the creation of plots of forecast and analyzed time series of accumulated precipitation over multi-day periods. In such cases, hourly Stage IV data was used, but the accumulated precipitation amounts over the multi-day periods were scaled (generally upwards) to be consistent with the amounts from the more quality-controlled six-hourly Stage IV and 24-hourly AHPS data. A description of the AHPS precipitation analyses are provided at http://water.weather.gov/precip/about.php. Stage IV data is documented in Lin and Mitchell (2005) and at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/ .

The following forecast modeling systems were examined in this study: medium-range global ensembles from the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS), the European Centre for Medium-Range Weather Forecasts (ECMWF), the United Kingdom (UK) Met Office, and the Canadian Meteorological Center (CMC). Regional ensembles were also examined from the NCEP Short-Range Ensemble Forecast (SREF) system. Deterministic forecasts were examined from the NCEP Global Forecast

System (GFS), the NCEP regional North American Mesoscale (NAM) system, and the Rapid Refresh Model (RAP). A more extensive documentation of the model configurations is provided in online appendix A.

Native forecast model resolutions varied widely. However, for the global ensemble predictions, the forecast data for the CMC, ECMWF, and the NCEP GEFS systems beyond day +8 were not available on their native grid as obtained from the TIGGE (THORPEX Interactive Grand Global Ensemble; Bougeault et al. 2009) archive. For these global ensembles, raw data was obtained from TIGGE at the highest possible resolution and then interpolated to a 0.2-degree grid before display or analysis. Some of the global models used native grid spacings that were relatively coarse for such a local event; for example, the UK Met Office used a 0.83-degree grid with its ensemble prediction system. When interpreting the subsequent results, especially for comparisons over Boulder County, which will be a box of 0.5 x 0.6 degrees, the reader should be aware that the forecast models cannot be expected to provide this level of detail.

3. Results.

A very abbreviated set of the most pertinent results are presented here; a much more complete set of forecast results spanning a large range of initialization times are presented in online appendices B (for global ensemble forecasts) and C (for shorter-range and deterministic forecasts). Figure 2 shows time series of global ensemble forecasts of accumulated precipitation from the four global ensemble systems, in this plot for forecasts initialized 12 UTC Monday, 08

September 2013, ~60 h before the onset of the heaviest precipitation. The three panels show the precipitation guidance approximately over Boulder County, and then over progressively larger areas. These larger areas were included because theory and practice suggests that precipitation forecast skill should be larger over larger areas (Islam et al. 2001, Gallus 2002). Hence, we seek to determine whether precipitation forecast consistency with the analyzed data improves with increasing scale. Precipitation forecast accuracy was not evaluated objectively, for example with threat scores or ranked probability skill scores. Such statistics are commonly only significant when evaluated over many dozens of independent events.

Figure 2 shows that over Boulder County, with the exception of CMC, the ensemble systems for this initialization time were generally predicting total accumulations in excess of 50 mm over Boulder County. None produced accumulated precipitation anywhere near the analyzed amount, which was ~ 250 mm, though some of this can be attributed to the coarser model grid spacing. For similar forecasts at other lead times (online appendix B), there were occasionally one or two members with total accumulations up to 60% of observed. The ensemble guidance produced greater precipitation amounts over Boulder County as the event got closer, but then for the several lead times just prior to the onset of heaviest precipitation, the ensembles again forecasted somewhat lighter precipitation amounts. This happened with all four models. At the intermediate scale in Fig. 2b, the ensemble predictions still under-forecasted the rainfall accumulation, though the discrepancy between analyzed and forecast was lessened. Finally, Fig. 2c shows that the precipitation forecasts were even more consistent with the analyzed accumulation over the largest region, as suggested in the previous literature, including in Lavers and Villarini (2014) for this case.

Was the deficiency of precipitation noted in the forecast ensembles in Fig. 2a merely a consequence of the models' coarse grid spacing? This can be examined in part by examining the spatial patterns of accumulated precipitation. Figure 3 maps the analyzed precipitation and the four global systems' ensemble-mean forecasts. For ease of interpretation, Fig. 3d also shows a coarser ~1-degree smoothed precipitation analysis, more consistent with a resolution the forecast model can potentially predict. Figures 3b-c show that both the NCEP and ECMWF systems were forecasting a local maximum of precipitation near Boulder County, with the maximum in the NCEP system displaced slightly west of the analyzed position. The NCEP forecasts also under-forecasted the precipitation through much of New Mexico. ECMWF predicted the heavier precipitation along the Front Range, consistent with the analyzed pattern but missed the extension of heavy precipitation to the southeast of Boulder and somewhat in eastern New Mexico. Figures 3 e-f show that the UK Met Office forecast maximum in the northern Front Range was weaker and further east, and the CMC forecast maximum at this time were much weaker and slightly further east. Generally, across many initial times, ECMWF and NCEP's GEFS produced better pattern forecasts, though their amplitudes were consistently too low, even with respect to the 1-degree smoothed analyses in Fig. 3d. While this can be due in part to the "smearing" effect of ensemble averaging precipitation that occurs when members' maxima are in different locations, it is apparent that the overall ensemble-mean patterns of heavy precipitation were different than for the analyzed. The deficient precipitation noted in Fig. 2a is hence likely to be due in part to errors in the pattern of precipitation that was forecast, not just due to the coarse grid spacing.

We now turn our attention to shorter-range forecasts. Figure 4 shows plume diagrams of accumulated precipitation for the forecasts initialized around 00 UTC 11 Sep 2013, i.e., Tuesday evening, 24 h before the onset of heaviest precipitation in Boulder County (the SREF was actually initialized at 03 UTC 11 Sep 2013). Forecasts from the GEFS, SREF, and deterministic GFS and NAM were considered. The two deterministic forecast models show much lighter than analyzed accumulations, and the GEFS system also significantly under-forecasted the accumulated precipitation. In the SREF system, however, there were several members with accumulated precipitation that was remarkably consistent with the analyzed precipitation. At the intermediate and larger scales in Fig 4(b) and 4(c), there was greater consistency between forecast and analyzed precipitation amounts across the modeling systems.

Figure 5 shows stamp maps for the SREF system, indicating that it was the members that used the WRF/ARW model that produced the exceptionally high precipitation. These show that the SREF's ARW forecasts were rather consistently producing heavy precipitation along the northern Front Range and generally heavy precipitation in much of Colorado down through central New Mexico. SREF system WRF/ARW forecasts initialized several days prior to the event also produced heavy precipitation on Tuesday, a day before the heaviest precipitation, as shown in data presented in online appendix C. Hence, despite the superior forecasts of the SREF

WRF/ARW members for the northern Front Range, it is possible that because heavier precipitation forecasts from those earlier initializations did not occur, forecasters might have discounted somewhat the heavy precipitation in later guidance.

The reasons behind the superior forecasts for the SREF members that used the WRF/ARW are not yet understood. The SREF members used three models, two different control initial conditions, and different perturbations for each member. Further data, presented in the online appendix C, show the mean SREF initial conditions for 10-m and 700 hPa analyzed winds, CAPE (convectively available potential energy), and total precipitable water. This also shows the deviations from the mean of the initial analyses used for the WRF/ARW members, the WRF/NMMB, and WRF/NMM. There was no "smoking gun" signature in the local initial conditions that would lead one to conclude obviously that WRF/ARW members would produce much more heavy Front-Range precipitation as a result of their initial state. There was no dramatically enhanced upslope flow, nor especially higher CAPE, nor much greater precipitable water for the WRF/ARW initializations.

At very short lead times, forecasters may examine guidance from the WRF Rapid Refresh, i.e., the RAP. It has been shown (Benjamin et al. 2009) that the radar reflectivity assimilation in the RAP has improved short-range forecast guidance of precipitation and reduced spin-up problems relative to other NCEP forecast systems without the digital-filter initialization to radar data. Figure 6 shows plume diagrams for the RAP. Unfortunately, for this case the RAP guidance almost always dramatically under-estimated the rate of accumulation of precipitation over Boulder

County during the period of most intense rainfall. However, the RAP guidance was more consistent with the analyzed accumulation when considering the forecasts over larger regions. Still, the RAP guidance would not have alerted forecasters to the potential for heavy rainfall near Boulder.

Interestingly, the RAP system used the WRF/ARW model, as did the SREF system that produced members that forecast the precipitation in the northern Front Range better than other systems. The mere usage of WRF/ARW apparently was not the crucial key to the SREF's improved forecasts over the northern Front Range. The RAP's 13-km grid spacing was similar to the SREF's 16-km. Perhaps the choice of parameterization may have been the ultimate source of the differences.

4. **Conclusions**.

This article briefly described the performance of precipitation forecast guidance leading up to the flash and river floods in the Front Range and in eastern Colorado, 9-16 Sep 2013. The article considered both global ensemble predictions from the NCEP GEFS as well as the ECMWF, UK Met Office, and CMC ensemble systems. Shorter-range forecast guidance from the NCEP GEFS, GFS, NAM, SREF, and RAP forecasts were also examined. Extensive online appendices are provided that provides model configuration details and additional plots of the analyzed conditions and forecast guidance for many other initial times.

The global ensemble prediction systems indicated that an abnormally wet pattern was to be expected in northeastern Colorado during 9-16 Sep 2013. However, the extent of the actual wetness near Boulder was not captured by any of the global ensemble prediction systems. This result is consistent with Lavers and Villarini (2014). Shorter-range prediction systems also dramatically under-forecasted the precipitation amount. Some noteworthy exceptions were the members of the SREF system that used the WRF/ARW model. These members produced very heavy precipitation in northern Colorado at the time when it was observed. Earlier runs, however, produced forecasts of heavy precipitation prior to the actual heavy precipitation. Interestingly, forecasts from the RAP system, which has very similar initial conditions and which also uses WRF/ARW model, did not produce heavy precipitation.

The WRF/ARW simulations in the SREF do suggest that the heavy precipitation in the northern Front Range of Colorado was somewhat predictable. Other scientists (e.g., personal communication, R. Shumacher, 2013) have also generated higher-resolution WRF/ARW simulations that forecasted the storm better than most of the operational guidance. It may be that the WRF/ARW system was more predisposed to produce heavy precipitation when run with certain combinations of parameterizations. Further experimentation is suggested to understand what model aspects were particularly important to producing heavy precipitation over the northern Front Range. Ideally, it would be interesting to examine other high-impact cases such as the May 2010 Nashville floods (Moore et al. 2012) and determine if there are any general principles for model configurations to improve QPF.

NOAA, the National Oceanic and Atmospheric Administration, has recently emphasized research and development on other high-impact events such as hurricanes relative to quantitative precipitation forecasting. The largely unexceptional forecasts during this event remind us that improving precipitation forecast guidance is still an urgent necessity within NOAA. Plans have previously been formulated that still provide useful a useful roadmap for how NOAA can improve its warm-season quantitative precipitation forecasts (Fritsch and Carbone 2004). Perhaps this event will spur NOAA to dust off and vigorously pursue such plans.

Acknowledgments:

Goeff DiMego, Geoff Manikin, Jun Du, Yuejian Zhu, and Glenn White of NCEP/EMC are thanked for providing information on accessing model data. Gary Bates of ESRL/PSD is thanked for help with the data processing. Seth Gutman is thanked for help obtaining the GPS total precipitable water time series shown in online appendix A. This publication was partially supported by a NOAA Office of Weather and Air Quality (OWAQ) USWRP grant. This project also used data from ECMWF's TIGGE archive. TIGGE is supported by the World Meteorological Organization's THORPEX program. Russ Schumacher (Colorado State), Wallace Hogsett (NCEP/WPC) and Jeff Whitaker (ESRL/PSD) are thanked for their consultations.

References

Benjamin, S. G., and co-authors, 2009: Technical review of the rapid refresh / RUC project. NOAA/ESRL/GSD Internal Review. [Available online at http://ruc.noaa.gov/pdf/RR-RUC-TR_11_3_2009.pdf].

Bougeault, P., Z. Toth, many others, T. M. Hamill, and many others, 2009: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull Amer. Meteor. Soc.*, **91**, 1059-1072.

Fritsch, J. Michael, R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: a USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–965. doi: http://dx.doi.org/10.1175/BAMS-85-7-955

Gallus, W. A., Jr., 2002: Impact of verification grid box size on warm season QPF skill measures. *Wea. Forecasting*, **17**, 1296-1302.

Islam, S., R. L. Bras, K. A. Emanuel, 1993: Predictability of mesoscale rainfall in the tropics. *J. Appl. Meteor.*, **32**, 297–310.

Lavers, D. A., and G. Villarini, 2014: Were global numerical weather prediction systems capable of forecasting the extreme Colorado rainfall of 9-16 September 2013? *Geophys. Res. Letters*, to appear. doi: 10.1002/2013GL048282.

Lin, Y. and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. *Preprints, 19th Conf. on Hydrology, American Meteorological Society*, San Diego, CA, 9-13 January 2005, Paper 1.2.

Moore, B. J., P. J. Neiman, F. M. Ralph, and F. E. Barthold, 2012: Physical processes associated with heavy flooding rainfall in Nashville, Tennessee, and vicinity during 1–2 May 2010: The role of an atmospheric river and mesoscale convective systems. *Mon. Wea. Rev*., **140**, 358–378.

FIGURE CAPTIONS

Figure 1: Accumulated precipitation analyses. The large panel shows the accumulated precipitation, taken from AHPS analyses, for the period 12 UTC 9 Sep 2013 to 12 UTC 16 Sep 2013. The smaller panels show the 12 UTC – 12 UTC accumulated precipitation for individual days, focusing on the Boulder-Denver metro area and the northern Front Range. Boulder County is in the center of these smaller panels.

Figure 2: "Plume" diagrams of accumulated precipitation forecasts and the analysis for four global ensemble prediction systems, initialized 12 UTC 08 Sep 2013. The three panels provide the forecast and analyzed accumulated precipitation averaged over three increasingly large areas, denoted by the red box in each figure. Only the first 20 members of each ensemble prediction system are displayed.

Figure 3. (a) Analyzed precipitation for the period 12 UTC 09 Sep 2013 – 12 UTC 16 Sep 2013. Corresponding smoothed analyses (1-degree grid spacing) are shown in panel (d). Panels (b), (c), (e), and (f) present the ensemble-mean forecasts from the NCEP GEFS, the UK Met Office, ECMWF, and CMC, respectively.

Figure 4. As in Fig. 2, plume diagrams, but only for shorter-range deterministic and ensemble forecasts produced at NCEP, here initialized 00 UTC 11 Sep 2013.

Figure 5: Stamp maps of analyzed and accumulated precipitation forecasts from the NCEP SREF system, initialized at 03 UTC 11 Sep 2013. Individual panels

show the different member forecasts. The top row shows the member forecasts that used the WRF/ARW forecast model. The middle row shows member forecasts that used the WRF/NMMB forecast model. The bottom row shows member forecasts that used the WRF/NMM model. “ctl,” “n1”, and so on are the perturbation number.

Figure 6: Plume diagrams as in Fig. 2, but for RAP forecasts initialized every 3 hours, plotted over the period from 12 UTC 9 Sep 2013 to 12 UTC 16 Sep 2013. Each RAP forecast extends to +18 h lead time.

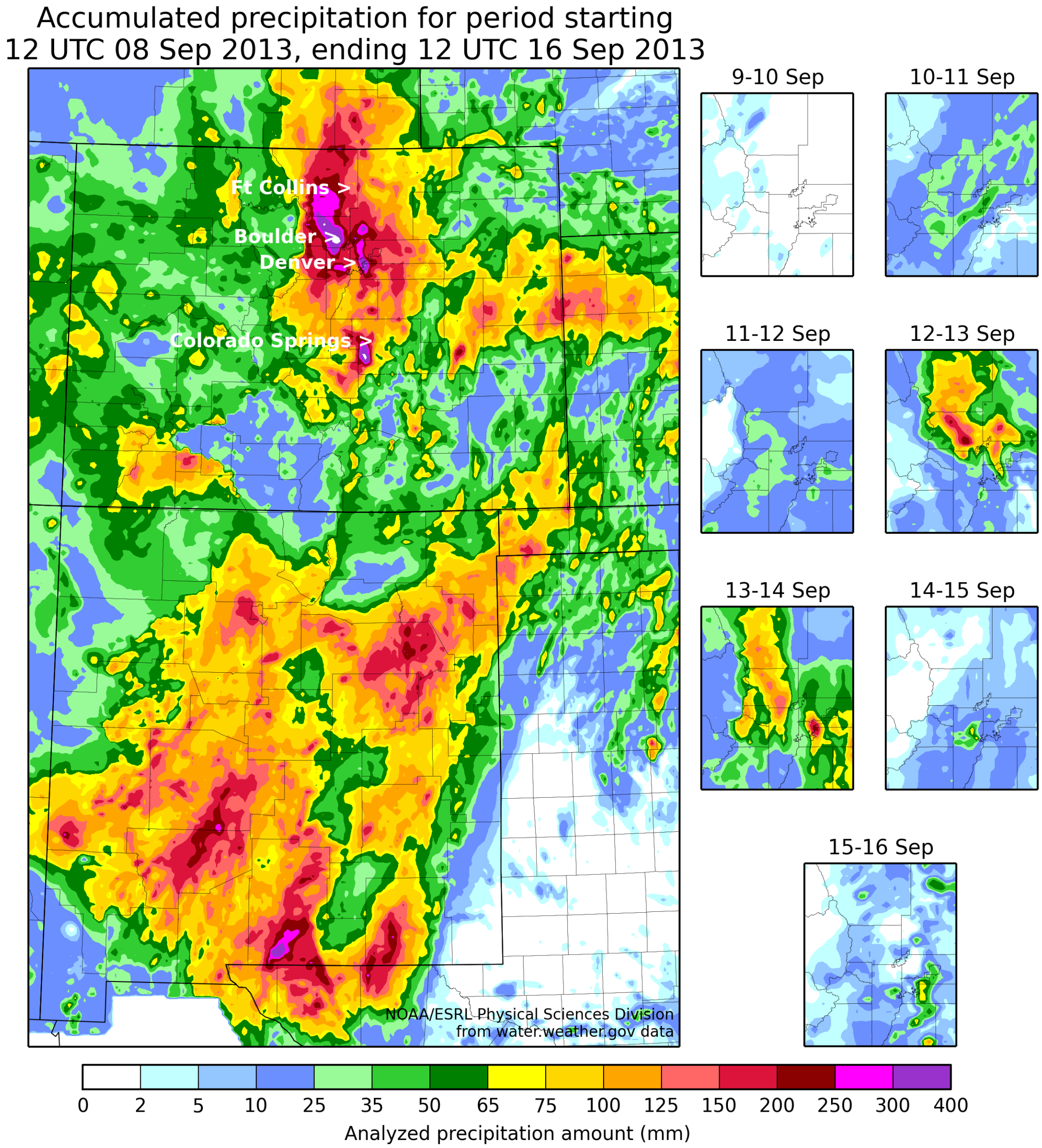

Figure 1: Accumulated precipitation analyses. The large panel shows the accumulated precipitation, taken from AHPS analyses, for the period 12 UTC 9 Sep 2013 to 12 UTC 16 Sep 2013. The smaller panels show the 12 UTC – 12 UTC accumulated precipitation for individual days, focusing on the Boulder-Denver metro area and the northern Front Range. Locations of major cities are roughly at the tip of each ">" symbol.

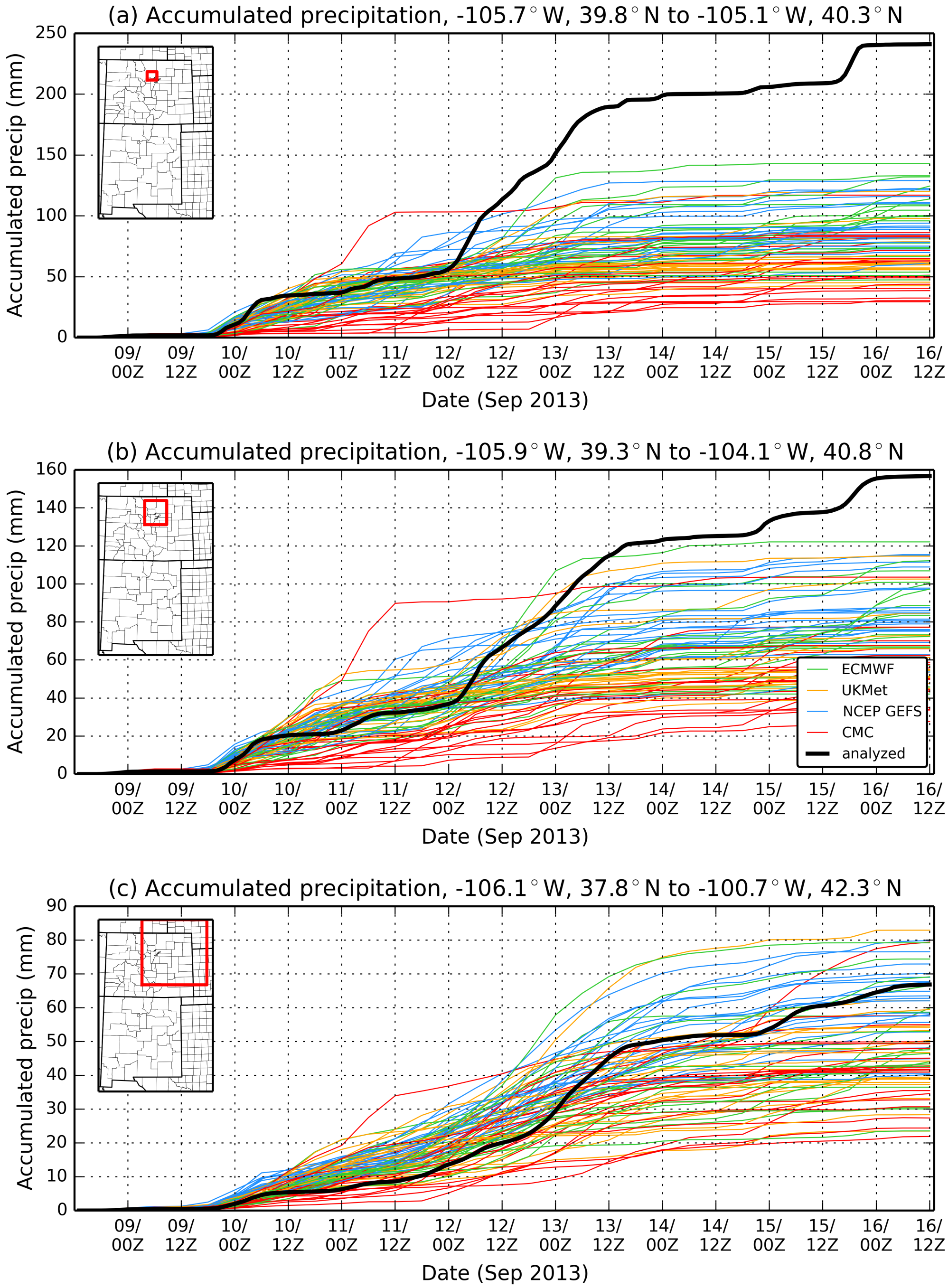

Figure 2: "Plume" diagrams of accumulated precipitation forecasts and the analysis for four global ensemble prediction systems, initialized 12 UTC 08 Sep 2013. The three panels provide the forecast and analyzed accumulated precipitation averaged over three increasingly large areas, denoted by the red box in each figure. Only the first 20 members of each ensemble prediction system are displayed.

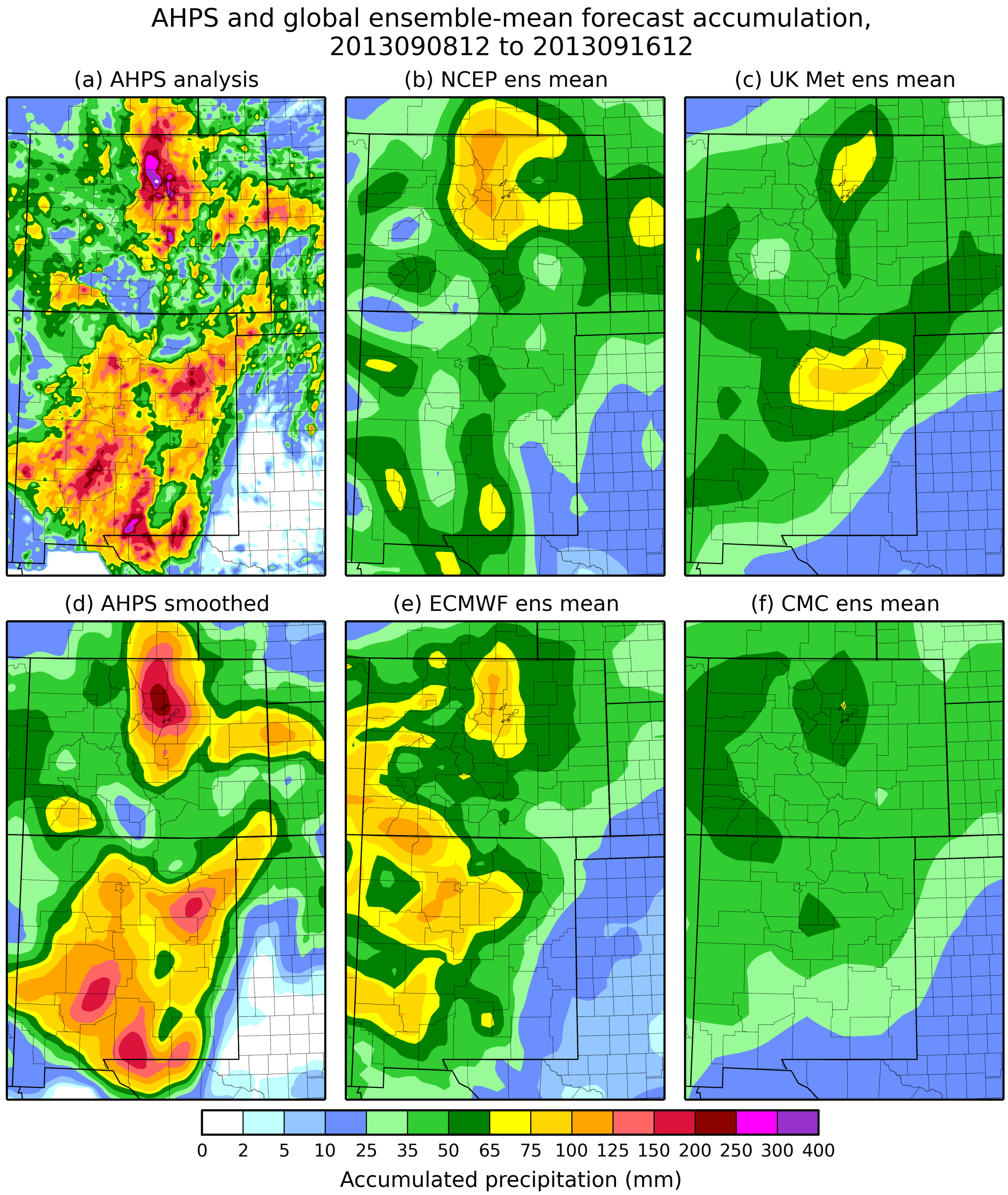

Figure 3. (a) Analyzed precipitation for the period 12 UTC 08 Sep 2013 – 12 UTC 16 Sep 2013. Corresponding smoothed analyses (1-degree grid spacing) are shown in panel (d). Panels (b), (c), (e), and (f) present the ensemble-mean forecasts from the NCEP GEFS, the UK Met Office, ECMWF, and CMC, respectively.

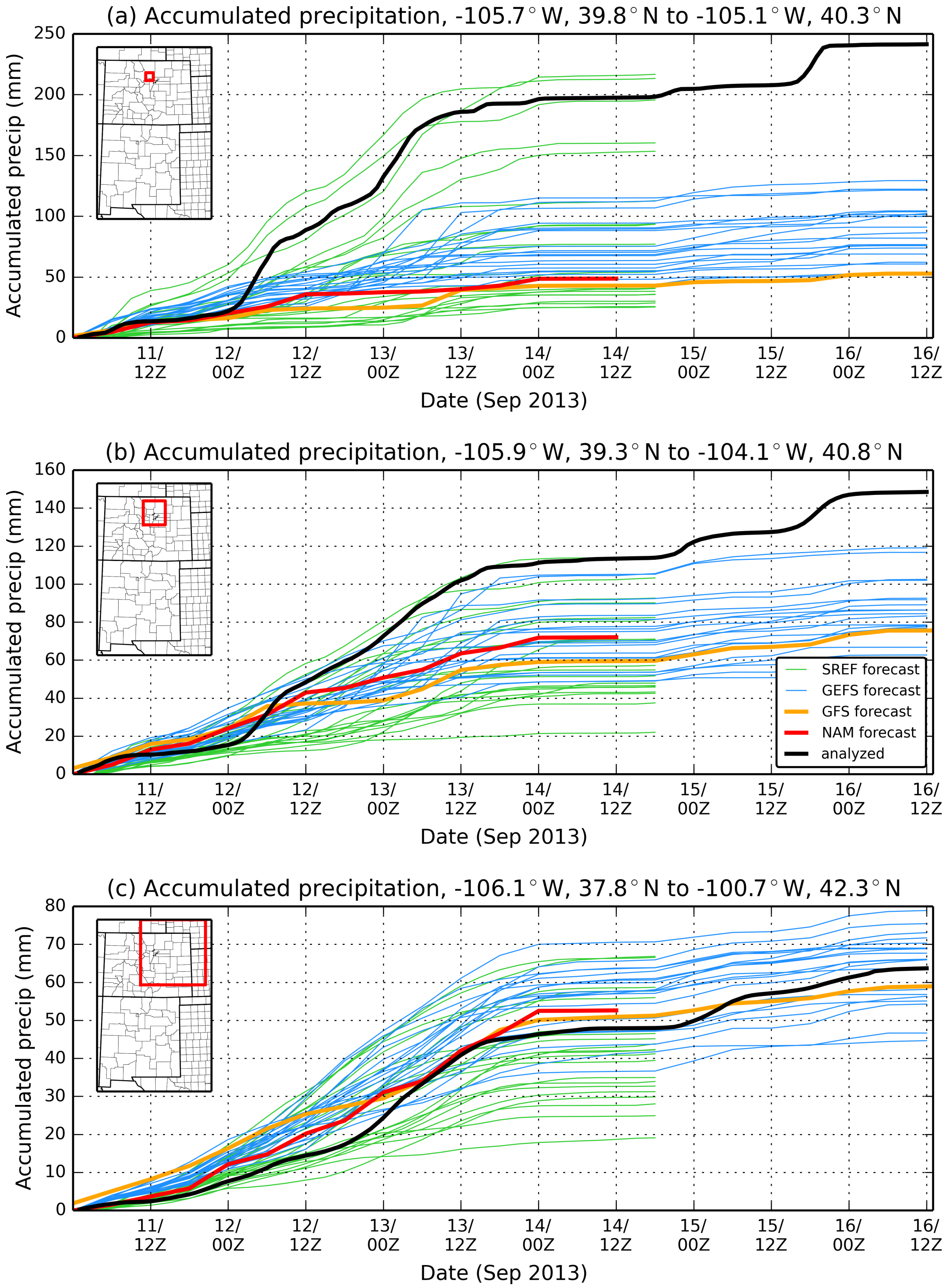

Figure 4. As in Fig. 2, plume diagrams, but only for shorter-range deterministic and ensemble forecasts produced at NCEP, here initialized 00 UTC 11 Sep 2013.

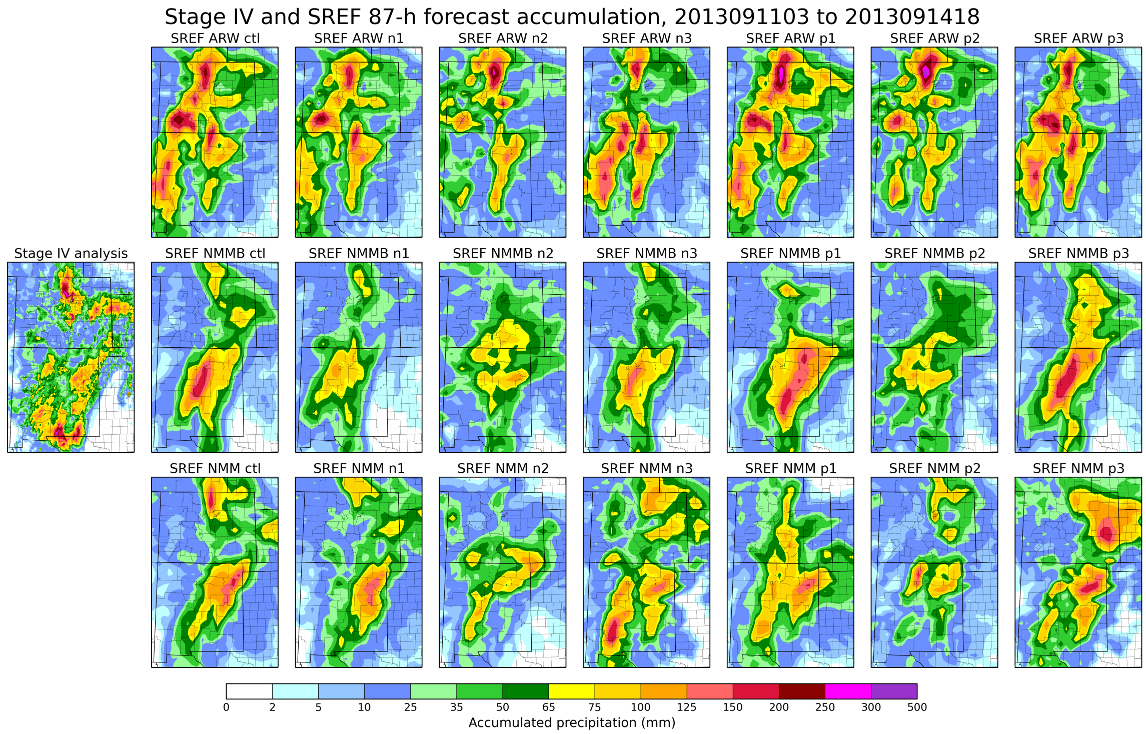

Figure 5: Stamp maps of analyzed and accumulated precipitation forecasts from the NCEP SREF system, initialized at 03 UTC 11 Sep 2013. Individual panels show the different member forecasts. The top row shows the member forecasts that used the WRF/ARW forecast model. The middle row shows member forecasts that used the WRF/NMMB forecast model. The bottom row shows member forecasts that used the WRF/NMM model. "ctl," "n1", and so on are the perturbation number.

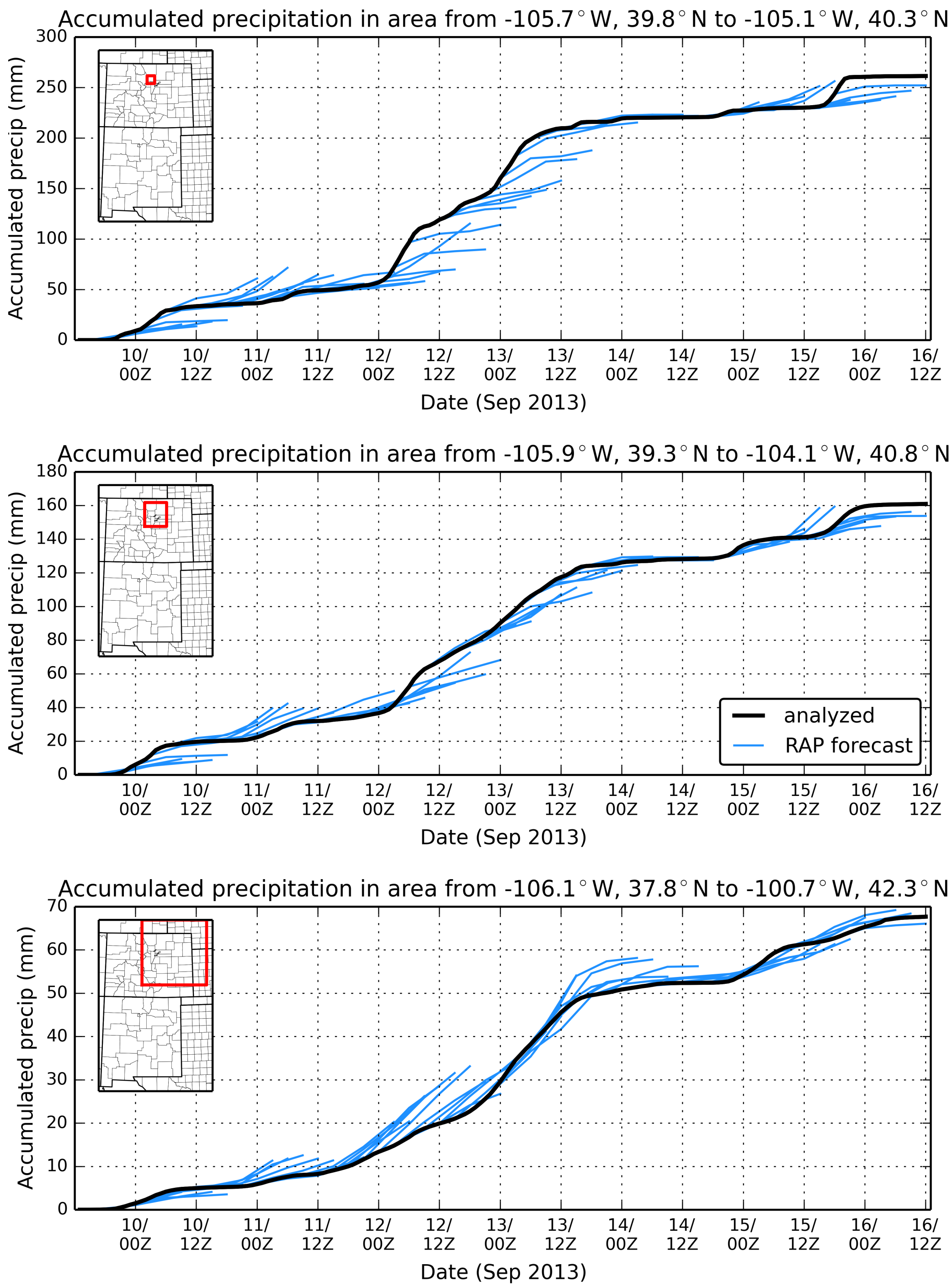

Figure 6: Plume diagrams as in Fig. 2, but for RAP forecasts initialized every 3 hours, plotted over the period from 12 UTC 9 Sep 2013 to 12 UTC 16 Sep 2013. Each RAP forecast extends to +18 h lead time.