

A Features-Based Assessment of the Evolution of Warm Season Precipitation Forecasts from the HRRR Model over Three Years of Development

JANICE L. BYTHEWAY^a AND CHRISTIAN D. KUMMEROW

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

CURTIS ALEXANDER

NOAA/Earth System Research Laboratory/Global Systems Division, Boulder, Colorado

(Manuscript received 21 April 2017, in final form 9 August 2017)

ABSTRACT

The High Resolution Rapid Refresh (HRRR) model has been the National Weather Service's (NWS) operational rapid update model since 2014. The HRRR has undergone continual development, including updates to the Weather Research and Forecasting (WRF) Model core, the data assimilation system, and the various physics packages in order to better represent atmospheric processes, with updated operational versions of the model being implemented approximately every spring. Given the model's intent for use in convective precipitation forecasting, it is of interest to examine how forecasts of warm season precipitation have changed as a result of the continued model upgrades. A features-based assessment is performed on the first 6 h of HRRR quantitative precipitation forecasts (QPFs) from the 2013, 2014, and 2015 versions of the model over the U.S. central plains in an effort to understand how specific aspects of QPF performance have evolved as a result of continued model development. Significant bias changes were found with respect to precipitation intensity. Model upgrades that increased boundary layer stability and reduced the strength of the latent heating perturbations in the data assimilation were found to reduce southward biases in convective initiation, reduce the tendency for the model to overestimate heavy rainfall, and improve the representation of convective initiation.

1. Introduction

The High Resolution Rapid Refresh (HRRR) model is an hourly updated storm-resolving model that was designed to provide rapid update model guidance on convective storms in order to improve severe weather forecasting, air traffic management, aviation hazards forecasting, and dissemination of severe weather warnings (ESRL 2015). The HRRR became the National Weather Service's (NWS) operational rapid update forecast model in September 2014 and has undergone continuous development since before its transition to operational status. Upgraded, experimental versions of the model are run concurrently with the operational

version in order to test updated model performance and measure improvements resulting from updates.

Bytheway and Kummerow (2015, hereafter BK15) performed a detailed features-based assessment of quantitative precipitation forecasts (QPFs) of long-lived warm season convective storms over the central United States using the 2013 experimental version of the HRRR. Bias statistics and composite features indicated that the model produced smaller, more intense storms than observed, particularly early in the forecast. Additionally, although the center of mass of precipitating features was well placed with respect to the observed precipitation, model features were often produced somewhat farther south.

Since BK15, there have been number of specific areas where the HRRR has been updated, including changes to the assimilation and microphysical schemes, to potentially reduce spinup and better represent initial precipitation fields early in the forecast. Changes were also introduced to reduce premature erosion of the cap in southern

^a Current affiliation: Cooperative Institute for Research in Environmental Sciences, Boulder, Colorado.

Corresponding author: Janice L. Bytheway, janice.bytheway@noaa.gov

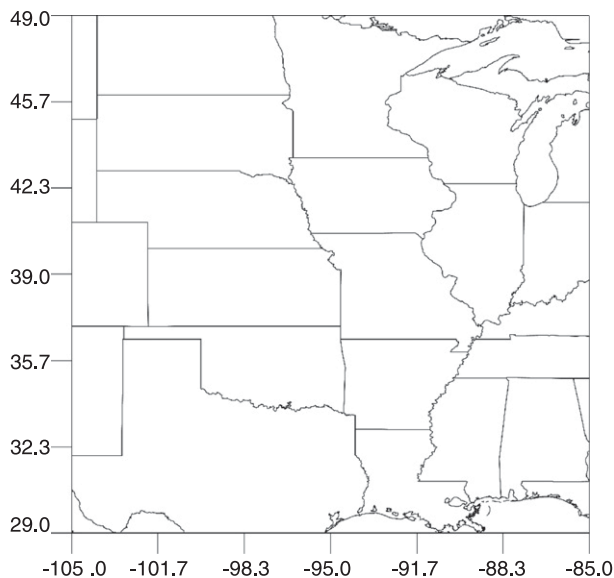


FIG. 1. Map of the study domain.

portions of the domain, with the intent of reducing the southward bias in precipitation production and improving the representation of the onset of convection.

With several years of experimental HRRR output now available, this study replicates the methodology of BK15 over the same domain (85°–105°W, 29°–49°N; shown in Fig. 1) in order to determine whether changes to the model had the expected outcomes, looking specifically for improvements to the problems mentioned above. In particular, answers to the following questions are sought:

- Have biases in intensity and areal extent of precipitation been reduced?
- Has spinup time been reduced?
- Has the slight southward bias been reduced or eliminated?
- Has the tendency for overprediction of heavy rain been reduced?
- How well does the HRRR capture the onset and development of convection?

The remainder of this manuscript will be structured as follows. Descriptions of the HRRR model and the National Centers for Environmental Prediction (NCEP) Stage IV multisensor precipitation product used as a reference are given in section 2. A brief review of the features-based methodology will be given in section 3. Section 4 will provide assessment results that indicate changes in model performance through multiple years of development and will attempt to relate these results to specific changes that were made in the model. Concluding remarks will be presented in section 5.

2. Data

a. High Resolution Rapid Refresh model

The HRRR model is an hourly updated convection-allowing model with hourly forecasts produced over the continental United States (CONUS) at 3-km horizontal resolution with 40 vertical pressure levels. The HRRR domain is within that of the 13-km Rapid Refresh mesoscale model, which also provides the boundary conditions (Benjamin et al. 2013). Full details of the HRRR model can be found in Benjamin et al. (2016). Forecasts from the 2013, 2014, and 2015 experimental versions of the model were evaluated, with the model output files obtained from National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) servers. While the 2016 experimental version of the model became available during the course of this research, the updates made were not expected to have a significant effect on precipitation forecasts (S. Benjamin 2016, personal communication).

While a maximum number of cases for study is desirable, local data storage for the large model output files was limited and, therefore, only a portion of the available forecast runs were used in this study. Specifically, in 2013, forecasts initialized every other hour were used, with odd- (even-) numbered hours obtained on odd- (even-) numbered Julian days. For 2014 and 2015, data were collected at Global Precipitation Measurement (GPM) core satellite overpass times as well as the 4h leading up to the overpass. The GPM dataset is being used to evaluate the HRRR in a separate research study (Bytheway and Kummerow 2017, manuscript submitted to *J. Geophys. Res.*). This provided an adequate sampling of the diurnal cycle and a large number of cases for study in each year (1108, 1207, and 744 matched feature pairs in 2013, 2014, and 2015, respectively; the lower number in 2015 is a result of several missing days of model output).

The HRRR has undergone continuous development and improvement, with new experimental versions of the model released yearly. Some of the changes expected to have an impact on QPFs are listed in Table 1 and described below. Early versions of the model were found to have a warm, dry bias during the day in the warm season, which was traced to a positive bias in incoming solar radiation due to the inability of the model to represent subgrid-scale clouds (e.g., shallow, fair-weather cumulus). As a result, the boundary layer scheme was altered to include a subgrid-scale cumulus parameterization that had a significant impact on incoming sensible heat flux (Benjamin et al. 2016). Additionally, the Thompson microphysics scheme (Thompson et al. 2008) was updated to account for aerosols (Thompson and Eidhammer 2014).

TABLE 1. Differences between the 2013, 2014, and 2015 HRRR that are expected to affect QPF quality.

HRRR version	Assimilation	Microphysics	WRF version	PBL
2013	GSI 3DVAR with radiances	Thompson et al. (2008) with enhancements	3.2.1+	Mellor–Yamada–Janjić (MYJ)
2014	GSI with hybrid ensemble–variational (0.5/0.5)	Same as 2013, with minor adjustments	3.4.1+	Mellor–Yamada–Nakanishi–Niino (MYNN)
2015	GSI with hybrid ensemble–variational (0.75/0.25)	Thompson and Eidhammer (2014)	3.6.1+	MYNN 2015

The addition of aerosol “awareness” resulted in both increased reflection of incoming shortwave radiation and increased cloudiness. These changes served to increase stability at the top of the mixed layer and slow boundary layer growth in the 2015 version of the model.

Additional changes to account for the warm dry bias in the warm season included changes to the land surface model both to account for irrigated cropland and reduce the wilting point of transpiration, essentially making it more difficult for the model to allow agricultural land to go dormant and thus increasing the surface latent heat flux. These changes to the model physics in the boundary layer and at the surface resulted in a cooler, moister mixed layer, which in turn reduced the daytime warm bias by 2°–3°C and reduced the high biases in convective initiation and production (Benjamin et al. 2016).

In addition to physics changes, many changes to the model data assimilation (DA) system were made from 2013 to 2015. The treatment of mesonet and METAR surface data was altered in the assimilation process to produce pseudoinnovations (differences between the observations and background). Using the difference between the observed and background 2-m temperature and dewpoint, these innovations are applied to the vertical temperature and dewpoint profiles in 20-hPa increments through the lowest 75% of the boundary layer. These pseudoinnovations were given increased weight in the assimilation optimization, essentially bringing the profiles at grid points with automated surface stations closer to the observed state and spreading the surface information vertically through the well-mixed layer (Benjamin et al. 2016).

The 2013 HRRR was the first version of the model to assimilate radar reflectivity via diabatic assimilation, a process by which the reflectivity measurement is used to calculate a latent heating perturbation to induce precipitation in the model initial state. The 2013 version of the model was fairly aggressive in adding latent heating in regions with observed reflectivity greater than 35 dBZ, and, as BK15 showed, this resulted in very high biases in extreme rainfall early in the forecast. In the 2014 version

of the model, the threshold for perturbing latent heating was reduced to 28 dBZ, but the strength of the forcing was decreased by a factor of 4, thus increasing the area of influence of the radar observations, but reducing the tendency to induce explosive convection.

Also in 2014, the DA system was upgraded from a 3D variational data assimilation (3DVAR) system to a hybrid system that included both 3DVAR and GFS ensemble Kalman filter (EnKF) systems, each having equal influence on the covariance matrices. This resulted in increased assimilation skill with respect to standard atmospheric observations. The EnKF system weight was increased to 75% in the 2015 version of HRRR.

b. NCEP Stage IV multisensor rainfall product

The NCEP Stage IV multisensor precipitation analysis (Lin and Mitchell 2005; Nelson et al. 2016) is a 4-km hourly precipitation product available over the CONUS. This product serves as the reference precipitation dataset for this study and consists of a mosaic of regional radar analyses produced by individual NWS River Forecast Centers (RFCs) and adjusted to gauge measurements. Each RFC produces an automated version shortly after the end of each hour of accumulation, which is followed several hours later by manual quality control and placement on a national grid at NCEP. Data are available in near-real time; however, they may not contain information from all of the RFCs. Final mosaics become available 12–18 h after the accumulation period and can be obtained from the National Center for Atmospheric Research (NCAR). These final quality-controlled full mosaics are used in the current study. Smalley et al. (2014), Prat and Nelson (2015), and Nelson et al. (2016) provide some discussion of uncertainties in the Stage IV product.

To evaluate whether changes in the HRRR performance over the three years are actually due to model changes or due to differences in the type of storms being forecast, the observed warm season accumulated precipitation for 2013–15 is shown in Fig. 2. The maps in Fig. 2 indicate that the 2013 warm season precipitation

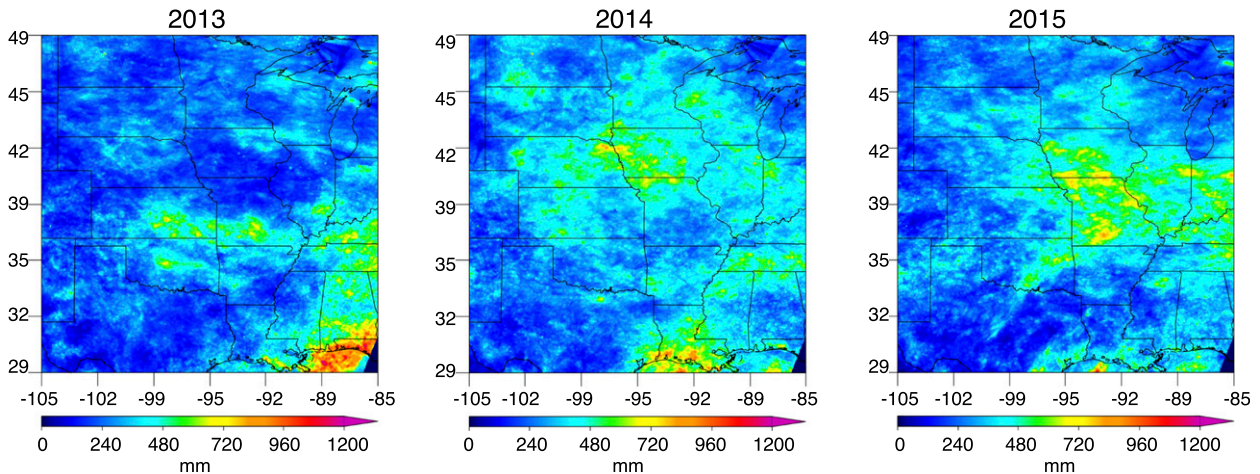


FIG. 2. Seasonal accumulated rainfall from the Stage IV product for June–August (JJA) 2013–15.

was concentrated more over the southern half of the domain, with large amounts of rainfall along the Gulf Coast and a secondary maximum over the southern Great Plains states of Oklahoma, Kansas, and Missouri. The accumulated rainfall distributions in 2014 and 2015 are more similar to each other, with rainfall maxima extending farther north into Iowa and Nebraska, with

the Gulf Coast maximum still present, but in smaller magnitude than in 2013.

For a more statistical view of rainfall over the three-year period of study, Fig. 3 displays PDFs of the storm area, maximum rain rate, mean rain rate, and total rainfall for identified observed features over the three-year period, for both the northern and southern halves

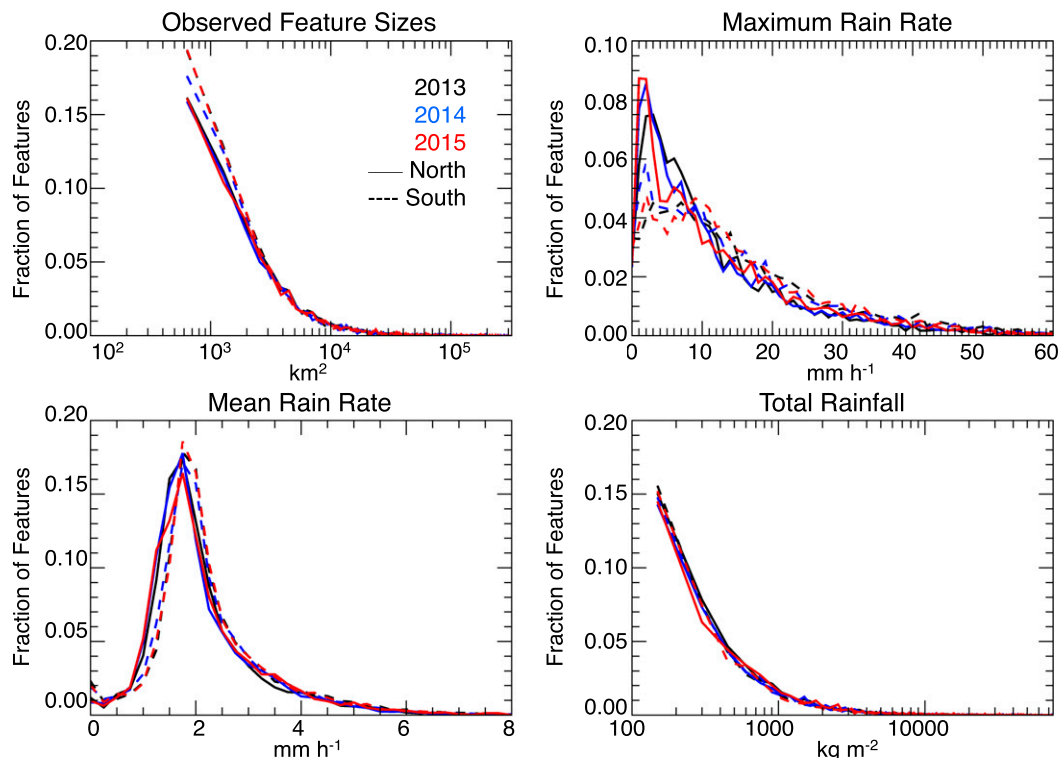


FIG. 3. Distribution of observed feature sizes, maximum rain rate, mean rain rate, and total rainfall from the Stage IV product for the 2013 (black), 2014 (blue), and 2015 (red) warm seasons (JJA) in the northern (solid) and southern (dashed) halves of the study domain.

of the domain. Separating the domain into northern and southern subdomains allows for an examination of whether the different patterns of accumulated rainfall in 2013 are a result of different types of precipitating features. The year-to-year differences in the distributions are generally small, on the order of 1%–2%. The distributions of the mean and maximum intensity are similar in both subdomains in all three years. The average maximum rain rate is $12.48 \pm 0.47 \text{ mm h}^{-1}$ ($15.88 \pm 0.49 \text{ mm h}^{-1}$) in the northern (southern) domain, and the average mean rain rate is $2.35 \pm 0.1 \text{ mm h}^{-1}$ in both domains. Therefore, any significant changes in the forecast intensity biases can be assumed to be due to model changes and not to differences in the types of convective storms occurring. With respect to precipitating area, storms in the northern part of the domain were on average somewhat larger in 2014 and 2015 (5807, 7354, and 8184 km² in 2013, 2014, and 2015, respectively), but similar in the southern domain over the three years examined ($\sim 5800 \text{ km}^2$). Given that the precipitating area of the examined storms spans several orders of magnitude and that the overall size distributions are very similar, the differences in the northern subdomain can also be considered relatively small. Thus, bias changes with respect to precipitating area are likely due to model changes, though impacts from natural variability cannot be completely ruled out.

3. Methodology

The assessment in this study follows the methodology of BK15, wherein features are identified in both the Stage IV and HRRR precipitating fields in a manner similar to the Method for Object-Based Diagnostic Evaluation (MODE; Davis et al. 2006, 2009). For full details of the feature identification, tracking, and matching algorithms the reader is referred to BK15.

The Stage IV and HRRR hourly precipitation fields are linearly averaged onto a 0.05° ($\sim 5 \text{ km}$) grid, which places the fields on the same frame of reference while maintaining the majority of the heterogeneity within the fields. A 15-km smoothing is then applied to the re-gridded fields, and a precipitating feature is defined as any contiguous area within the smoothed field exceeding a given rain rate (1 mm h^{-1}) with a maximum rain rate exceeding 10 mm h^{-1} (to ensure likely convection). Features are identified through each of the first 6 h of the forecast, and a database of statistical properties for each feature is created.

Using the database of statistical properties, both observed and forecast precipitating features are tracked through consecutive forecast hours. If an observed feature lasts for 6 h or more, a matching forecast feature is

sought at the beginning of the forecast. A match is defined as those forecast features that are either collocated with the observed feature or are nearby and statistically similar. If multiple forecast features are found to potentially match an observed feature, then the feature with the highest probability of detection (if collocated) or with the most similar total rainfall to the observed is selected for evaluation. Matching is done only at the first forecast hour, and calculated biases track how well the forecast features' behavior follows that of the observed features. If no matches are found, it is considered a missed forecast.

4. Results

a. Significant changes in model biases in 6-h forecasts

As in BK15, distributions of the bias in the feature mean, maximum, and total rainfall, as well as areal extent, were created for each forecast hour for all three years. Because BK15 indicated that these distributions were most useful early in the forecast, only the first 6 h will be examined here. Significant changes between the three years, as determined by the use of a Student's *t* test, were mostly found to affect the feature mean and maximum rainfall. Significant improvements between the 2013 and 2014 versions of the HRRR were also found with respect to total rainfall during the first forecast hour (FH).

Figure 4 displays the bias distribution for maximum rainfall intensity at forecast hours 1 (top) and 6 (bottom) for the three years of interest, with solid vertical lines indicating the distribution median, dashed vertical lines indicating the interquartile range (IQR), and dotted vertical lines indicating the 10th and 90th percentiles. While the median bias in maximum rainfall intensity remains approximately 20% at FH1 in both 2013 and 2014, both the 75th and 90th percentiles decrease in the 2014 HRRR, with the 75th (90th) percentile decreasing from 90% (160%) in 2013 to 70% (150%) in 2014. The 75th and 90th percentiles at FH1 do not change significantly from 2014 to 2015, though the median bias in maximum intensity at FH1 decreases to just under 10% in 2015, which is a significant change over the 2013 version of the model.

At FH6, median biases in maximum rainfall intensity are again similar in 2013 and 2014, with a 20% decrease in the 75th percentile between the two years, and an overall wider distribution than that at FH1, consistent with BK15's results. In fact, the distribution has widened sufficiently such that the 90th percentile exceeds +200% bias. From 2014 to 2015, the distribution continues shifting to the left, with the 2015 median bias decreasing

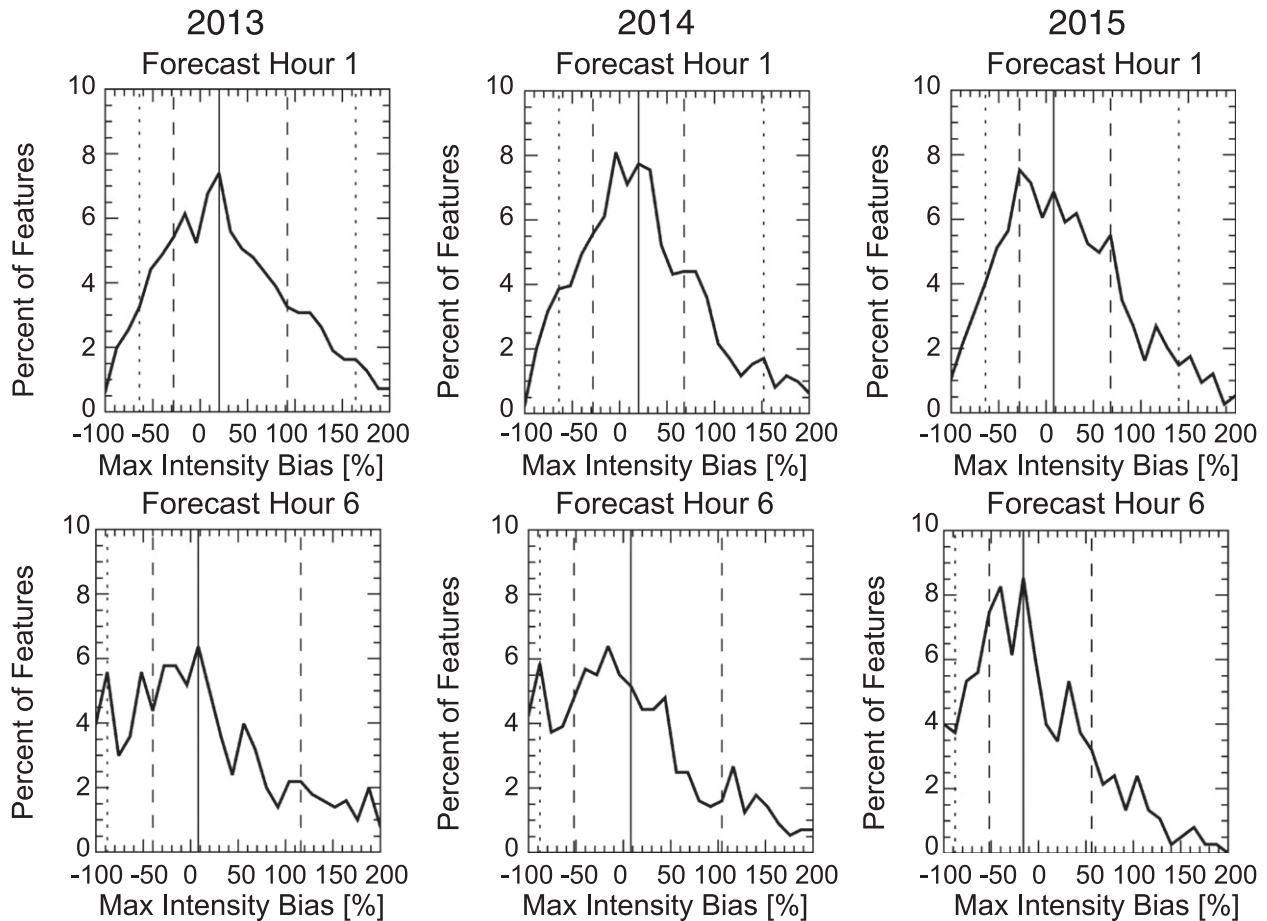


FIG. 4. Distribution of maximum intensity bias for (left) 2013, (center) 2014, and (right) 2015 at (top) FH1 and (bottom) FH6. Solid vertical lines represent the median of the distribution, while dashed vertical lines represent the IQR, and dotted vertical lines indicate the 10th and 90th percentiles.

from $\sim +10\%$ to $\sim -18\%$. The most obvious change at FH6 is the decrease in IQR, with the 75th percentile decreasing from $+105\%$ in 2014 to $+57\%$ in 2015.

Both DA and physics changes likely contribute to the significant improvement in the representation of precipitation intensity. The 2014 DA changes that lessened the imposed latent heat perturbation are expected to have the largest impact early in the forecast period, initially producing less intense storms than previous versions of the model. The improvement at FH6 is likely related more to the modifications made to decrease incoming shortwave radiation and stabilize the boundary layer, since physics changes are expected to have impacts lasting longer into the forecast period than changes to DA techniques.

Mean bias distributions differ significantly at all forecast hours between the 2013 HRRR and the next two versions (excepting FH6 in 2015), while changes to the mean bias distribution between 2014 and 2015 are

only significant at FH1, shown in Fig. 5. At FH1, the median bias in mean rainfall decreases from $+20\%$ in 2013 to $+9\%$ in 2014 and 2015. As in Fig. 4, the mean bias distributions show reduced skewness toward positive biases from year to year. The 75th percentile decreases from $+69\%$ in 2013 to $+57\%$ in 2014 and to $+42\%$ in 2015. While the 10th percentile remains stationary during all three years, the 90th percentile decreases from $+115\%$ in 2013 to $+90\%$ in 2014 and to $+80\%$ in 2015.

Overall, changes in both the median intensity biases and the reduction in the skewness of the bias distributions toward large positive biases can likely be attributed to both the changes to the model physics that increased the boundary layer stability as well as the alterations to the radar DA, with the DA upgrades having more impact early in the forecast and the effects of physics changes carrying forward longer through the forecast period.

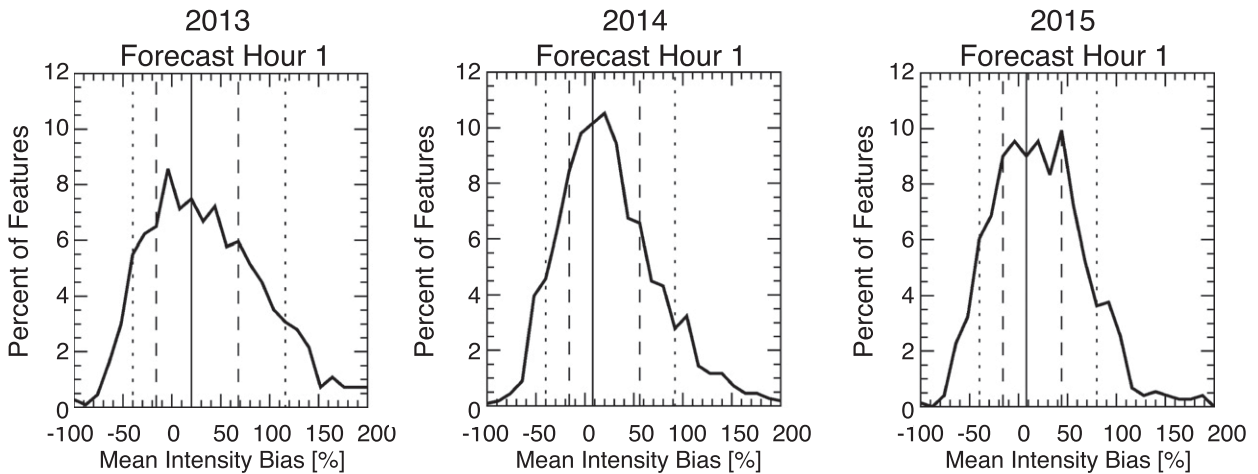


FIG. 5. As in Fig. 4, but for mean rainfall bias in the central U.S. domain at FH1 only.

b. Model spinup time

BK15 showed that in the early hours of the forecast, the best bias statistics were obtained at FH3. Additionally, when comparing FH1 with the observations from either the hour of initialization or the hour when assimilation began (i.e., a 1–2-h lag in observations to the forecast valid time), the lagged biases were smaller than those for the valid hour, particularly with respect to areal extent. By forecast hour 3, there was no longer a need to compare the forecast with lagged observations in order to obtain relatively low (generally within $\pm 10\%$) median biases in feature mean, maximum, and total rainfall, as well as areal extent. These results indicated the need for several hours of spinup prior to the HRRR producing the most accurate forecasts, and that in some cases the assimilated observations were still heavily influencing the precipitation forecast in the first few hours of the model run. In particular, it seemed that the model produced storms with

similar size to the assimilated storms with considerably more intense rainfall. Given that the HRRR is designed for short-term forecasting of convection, it is desirable to have as short a spinup time as possible.

Upgrades to the HRRR DA system between 2013 and 2015, particularly those reducing the strength of the latent heating perturbation induced in regions of observed reflectivity in the 2014 version, were inferred in the previous section to improve intensity biases in the early hours of the forecast. Figure 6 shows the median bias values for feature areal extent, mean rainfall, maximum intensity, and total rainfall for the first 6 h of the forecast in 2013 (black \times 's), 2014 (blue triangles), and 2015 (red stars). Given that the BK15 results generally found median bias values within $\pm 10\%$ by forecast hour 3, here we will define the model as having spun up at the hour when the median bias of a given property reaches $\pm 10\%$ (shown by vertical gray lines in Fig. 6), or when the absolute value of the median bias reaches its

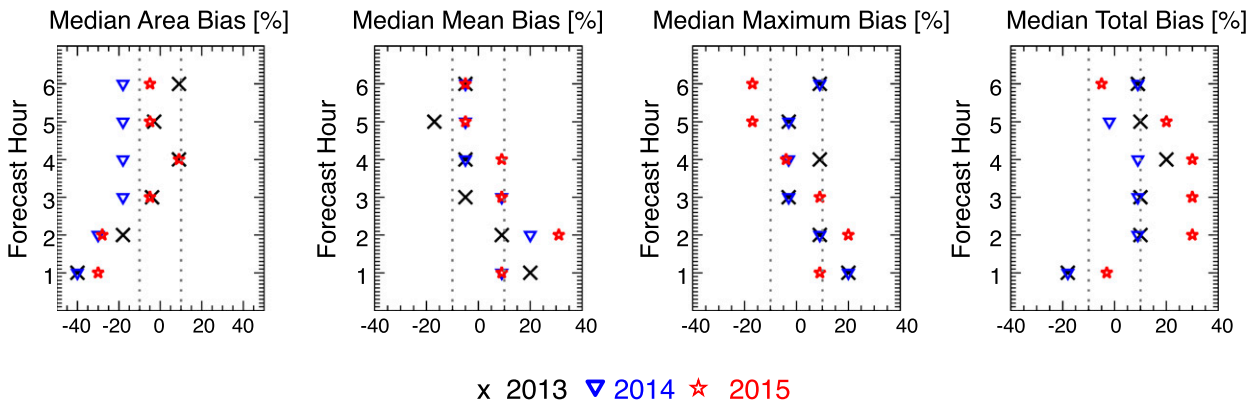


FIG. 6. For each of the first 6 h of a forecast, shown are the median values of the distribution of area, mean, maximum, and total rainfall biases in 2013 (black \times 's), 2014 (blue triangles), and 2015 (red stars). Dotted gray lines outline the $\pm 10\%$ threshold.

TABLE 2. Number of hours needed for the model to reach $\pm 10\%$ median bias in mean, maximum, and total rainfall and raining area in each year, as well as the average total spinup time.

	Area bias	Mean bias	Max bias	Total bias	Total spinup time
2013	3	2	2	2	2.25
2014	3	1	2	2	2.0
2015	3	1	1	1	1.5

minimum during the 6 h examined. The total spinup time is then defined as the average number of hours needed for each of the four bias values to reach the $\pm 10\%$ (or absolute minimum) threshold, given in Table 2. Using this definition of spinup, the model on average spins up 45 min faster in 2015 than in 2013.

This definition of spinup is quite subjective and derived specifically to compare the behavior of the 2014 and 2015 versions of the model with the results found by BK15 using the 2013 version. A more general interpretation of spinup might seek the point in the forecast when model behavior has “stabilized” or reached a statistical equilibrium, indicating increased dependence on the model’s

physics and parameterizations rather than the assimilated data and boundary conditions. Under this definition, it is difficult to discern from the bias behaviors in Fig. 6 just when that stabilization might occur. For example, while the median bias in the precipitating area remains at $\sim -20\%$ after forecast hour 3 in the 2014 version of the model, the median bias in maximum intensity remains fairly oscillatory through the first six forecast hours in all three versions of the model.

From the point of view of the model developers, the model is considered spun up when the size distribution of the model features matches that of the observed features, that is, an appropriate representation of the scales that the model is able to resolve. To examine model performance from this perspective, the distribution of feature sizes identified in both the model and observations is shown in Figs. 7, 8, and 9 for 2013, 2014, and 2015, respectively. In 2013 (Fig. 7), the model distribution indicates a nearly $+10\%$ bias in the smallest size category, with underprediction of larger features, in particular those between 2000 and 5000 km². The large overprediction of the smallest category remains over 5% until FH4, when the underprediction in the 2000–5000-km² category is also smallest. By 2014 (Fig. 8) and 2015 (Fig. 9), the difference

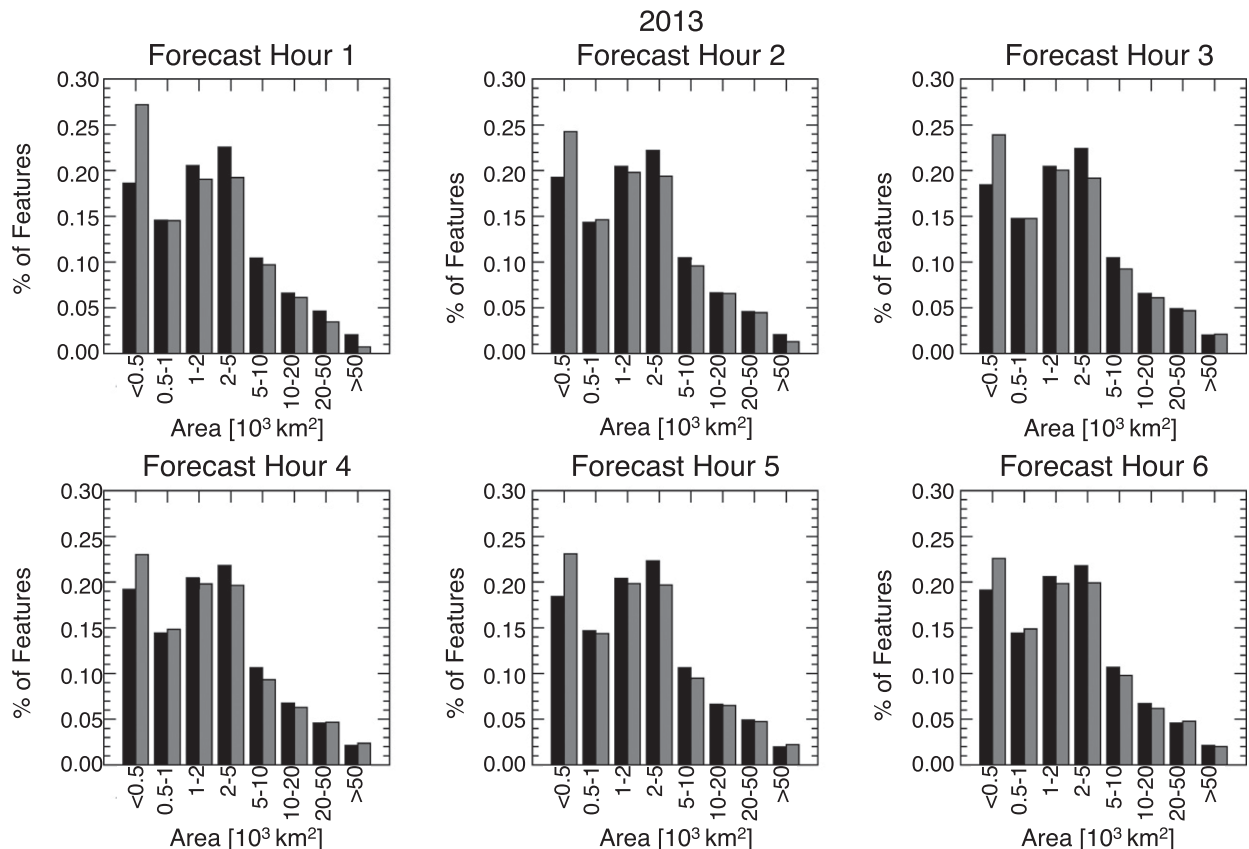


FIG. 7. Distribution of feature sizes from the observations (black) and model (gray) at the first 6 FHs for the 2013 warm season.

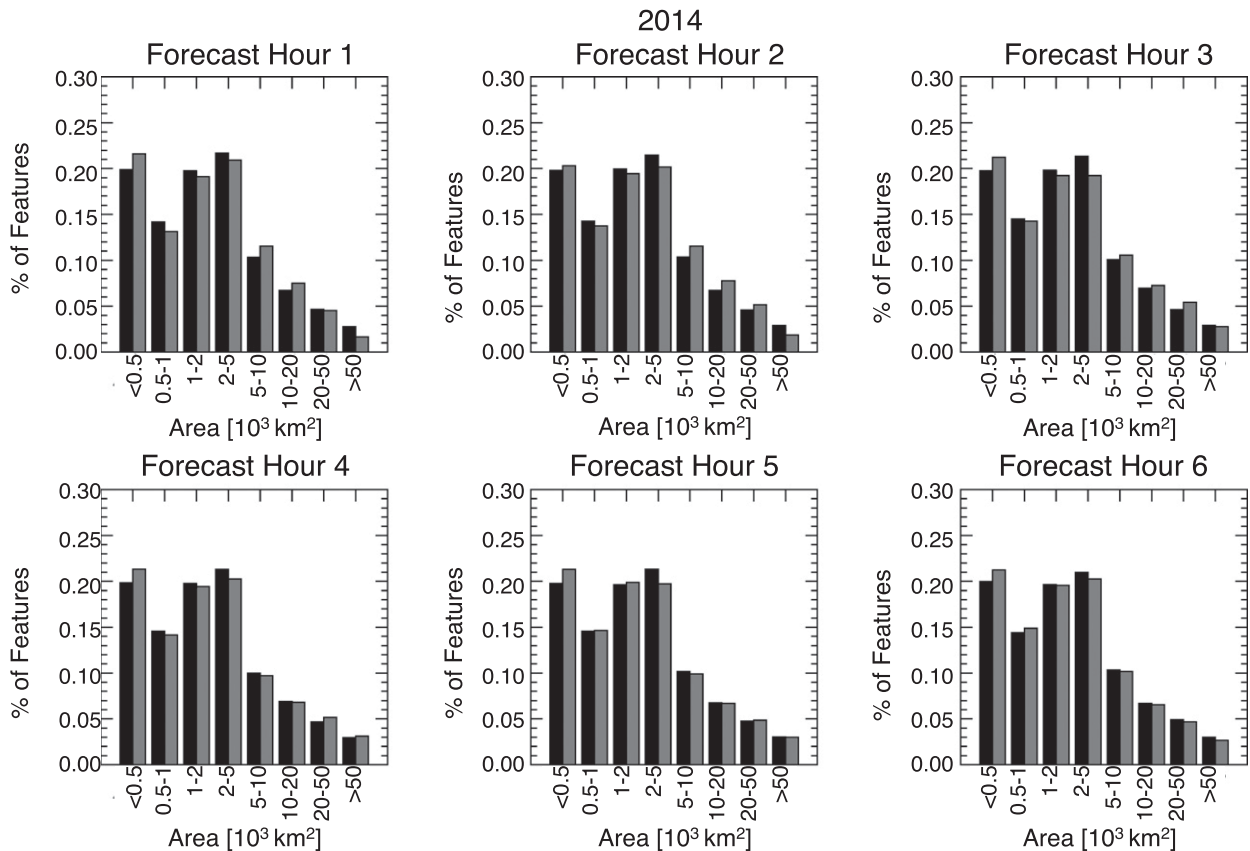


FIG. 8. As in Fig. 7, but for the 2014 warm season.

between the fraction of observed and forecast features in each size category is less than 2% in almost all forecast hours and size categories. Based on this criterion, the model does a better job representing the observed size distribution at early forecast hours in 2014 and 2015 than it does in 2013, indicating a spinup time reduction from 4 h to just 1 h. This definition of spinup does not take into account the model's ability to represent the precipitation processes taking place within the raining area, however (i.e., just because the size is accurately represented does not mean the intensity and distribution of precipitation within the feature is accurately represented as well). Additionally, model spinup will be determined by overall model behavior rather than just its representation of convective precipitation processes. As such, we can say that it appears that model spinup time is decreasing with continued development, but by how much depends on how one defines spinup.

The relatively small changes in overall model spinup given by the median bias definition and the overprediction (underprediction) of features in the smallest (moderate) size category are not surprising given the larger, presumably more mature storms considered in

this study. Because the DA period is fairly short (1 h) and the model output is produced at hourly time steps, the model is given just 1 h to produce mature, possibly intense storms, which is unlikely given the model's current capabilities.

c. Southward biases

Figure 4 in BK15 showed the HRRR producing storms with centers of mass within about 25 km of the centers of mass of observed storms, but with a slight southward bias that increased through the first few hours of the forecast (i.e., storms formed too far to the south, then either did not correct position in ensuing forecast hours or propagated northward too slowly). This bias was attributed to both the generally larger instability of air masses closer to the warm, moist Gulf of Mexico and rapid erosion of the cap by the model. Such diurnally driven boundary layer processes are generally more prevalent to the south, where synoptic forcing is not as strong as it tends to be farther to the north, closer to the warm season storm tracks. Figure 10 shows the difference in the fraction of identified features between the radar and the model in 0.5° -latitude bands by time of day

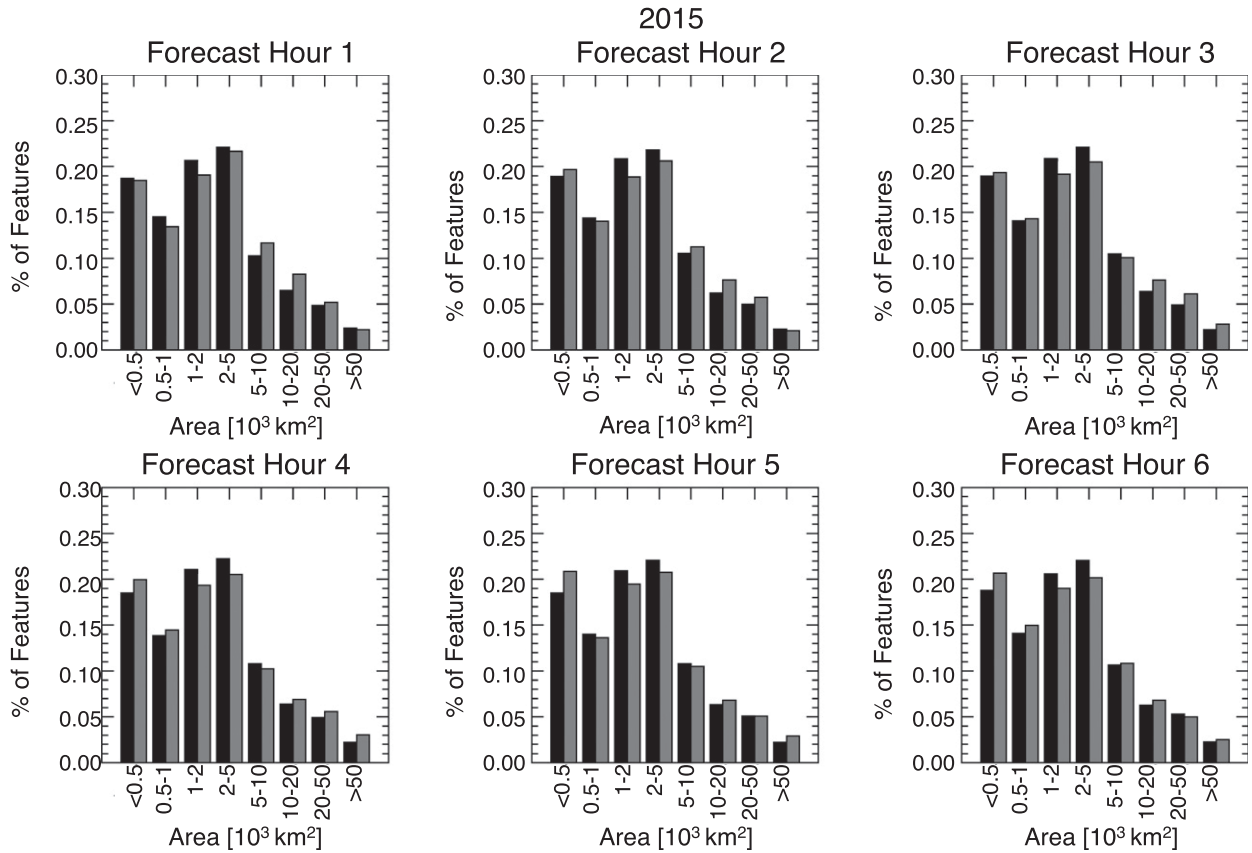


FIG. 9. As in Fig. 7, but for the 2015 warm season.

for the first hour of the forecast (i.e., forecast valid time of 1800 UTC corresponds to 1700 UTC model run) for the eastern domain, calculated as

$$\left(\frac{\sum \text{Radar features in } 0.5^\circ \text{ band}}{\sum \text{All Radar features}} - \frac{\sum \text{Model features in } 0.5^\circ \text{ band}}{\sum \text{All Model features}} \right) \times 100. \quad (1)$$

Blue shading in Fig. 10 indicates a larger fraction of identified forecast features in a latitude band than observed, and appears to start at around 1700 UTC in the southernmost part of the domain in 2013. The overprediction by the model propagates northward over the next several hours; however, the strongest biases are found south of 39°N. Figure 10 also indicates when the model begins to strongly initiate warm season convection, which will be discussed further in section 4e.

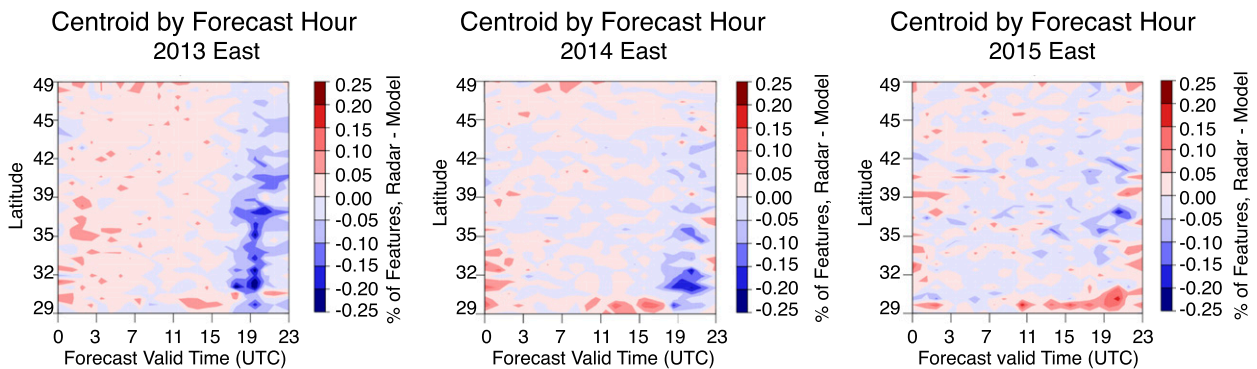


FIG. 10. Difference in the fraction of observed and identified features in 0.5°-latitude bands by time of day for FH1 during the 2013–15 warm seasons.

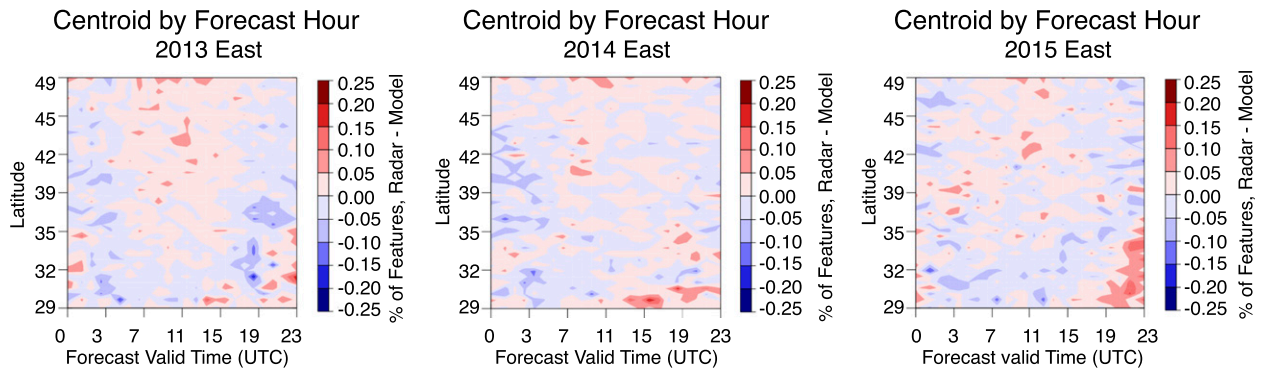


FIG. 11. As in Fig. 10, but for FH3.

In 2014, the overprediction of convection in the southern part of the domain appears to begin about an hour later (~ 1800 UTC), suggesting that updates to the 2014 version of the model delayed convective initiation slightly. The biases are limited mostly to areas south of 35°N , with strongest biases between 29° and 32°N . By 2015, only a small region of overprediction remains between 35° and 39°N , and the tendency for overprediction has been replaced by slight underprediction in the southernmost part of the domain, indicating that the boundary layer physics changes and addition of the subgrid-scale cumulus parameterization made in the 2015 HRRR have in fact significantly reduced the tendency for the model to initiate strong convection in the south.

The tendency for the model to overforecast convection in the southern United States was not only dependent on time of day, but also on forecast hour. Figure 11 shows the same data as Fig. 10, but for the third hour of the forecast run. Overall, the biases are smaller in magnitude, likely as a result of the model becoming more stable during spinup. In 2013, the excess initiation of convection was limited to regions south of 39°N , with some small biases being introduced during the overnight hours (0000–0500 UTC). The slight underprediction seen in FH1 of the 2015 HRRR becomes evident in FH3 in the 2014 version. Combined with the FH1 results shown in Fig. 10, this suggests that the 2014 version of the model created a large number of convective features early in the forecast that either 1) died off rapidly as a result of weaker latent heat forcing during assimilation or 2) coalesced into a smaller number of larger features. By 2015, the underprediction in the southern half of the domain seen in FH1 has started to increase toward the north with time, as the storms not produced at FH1 either fail to materialize later in the forecast or those that are produced either do not propagate northward or dissipate.

d. Overproduction of extreme rainfall

While not direct results of BK15, high biases in maximum rainfall, excessively strong convective cores shown in composite features, and results from the developers' in-house validation work suggested that the HRRR was significantly overpredicting very heavy rainfall. The combination of reduced latent heat forcing in the DA along with changes to the model physics are one possible way to reduce this overprediction. Figure 12 shows the difference in the fraction of identified features with maximum rain rates exceeding 5, 10, 25, 50, 75, and 100 mm h^{-1} [calculated in the same manner as Eq. (1)] in the first 6 h of the forecast for the three years of study.

The top panel in Fig. 12 shows that the 2013 HRRR usually underforecasted the fraction of features with rain rates exceeding 5 and 10 mm h^{-1} . This was also indicated in BK15 by a low probability of precipitation for features exceeding 5 mm h^{-1} in the absence of very large amounts of atmospheric moisture. For features with rain rates exceeding $25\text{--}75\text{ mm h}^{-1}$, however, the 2013 HRRR had a tendency to overpredict during all of the first 6 h of the forecast. This suggests that the 2013 HRRR needed some catalyst to produce moderate to heavy rainfall, but once that requirement was met, it was able to produce and sustain more intense rainfall than observed.

By 2014, the pattern at $5\text{--}10\text{ mm h}^{-1}$ reversed, with the HRRR overpredicting the fraction of features exceeding these thresholds by a significant margin, particularly at forecast hour 3. The tendency to overpredict the fraction of features with maximum rain rate exceeding 5 mm h^{-1} increases over the first 3 h of the forecast and then decreases through hours 4–6. This is possibly related to the assimilation of lighter rainfall induced when the reflectivity threshold for assimilation was reduced from 35 to 28 dBZ. In 2014, FH6 has only very small biases at all thresholds, a likely result of physics changes, as DA impacts are typically expected to be strongest

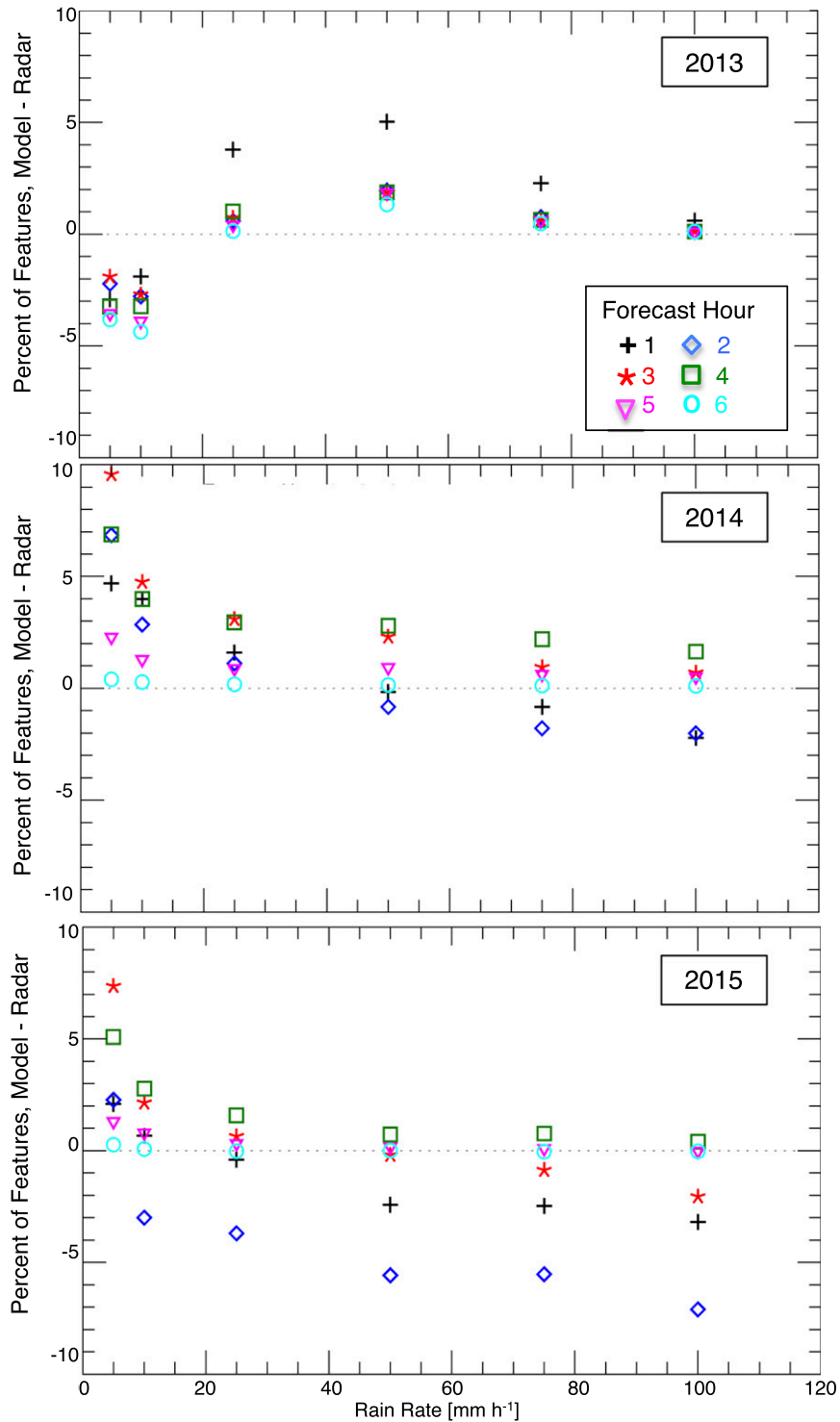


FIG. 12. Difference in the fraction of identified features from the model and observations with maximum rain rate exceeding given thresholds in the first 6 FHs for (top) 2013, (middle) 2014, and (bottom) 2015. Colored symbols indicate the hour of the forecast period.

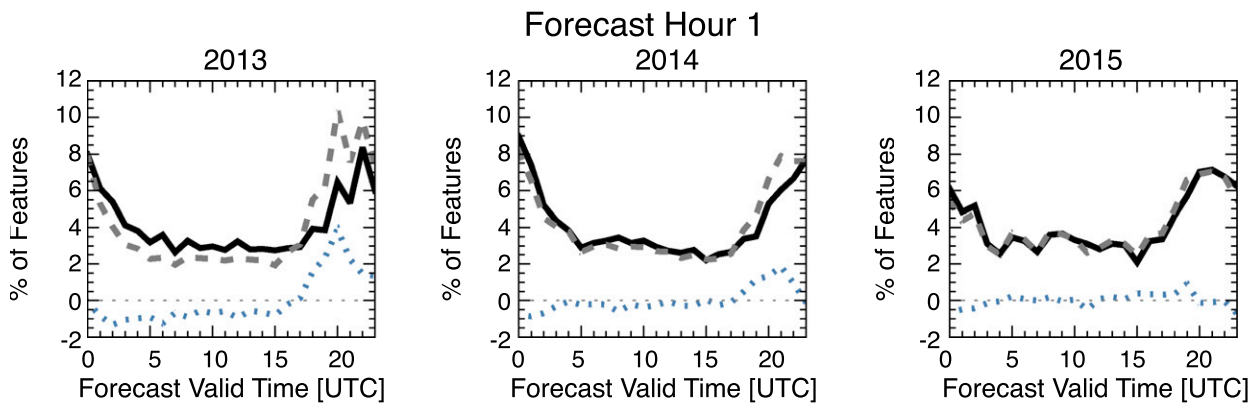


FIG. 13. The fraction of storms produced by the HRRR at FH1 for each forecast valid time compared with the fraction of observed storms identified for each hour. Black solid lines represent the observations, gray dashed lines the model, and dotted blue shows the model – radar.

early in the forecast. DA changes were more likely responsible for the decreasing overprediction of heavy rainfall with increasing thresholds in the first three forecast hours and underprediction of features exceeding 50 mm h^{-1} at FH1 and FH2.

By 2015, only the fraction of features with maximum rainfall exceeding 5 mm h^{-1} are overforecast during all of the first six forecast hours, while the tendency to underpredict the fraction of features with very heavy rainfall in the early forecast hours that became evident in 2014 is now apparent through the first 3 h of the forecast, particularly for those features with maximum rainfall exceeding 75 mm h^{-1} . The largest underprediction is seen at forecast hour 2 for all features with maximum rainfall exceeding 10 mm h^{-1} . Clearly, the 2014 changes made to the DA combined with the increased boundary layer stability in 2015 have acted to dampen the production of convective rainfall early in the forecast, perhaps somewhat too much. It is hypothesized that, with the increased stability afforded by the physics and assimilation upgrades, the model is exhausting much of the energy needed to initiate and sustain these storms early in the forecast, resulting in underforecast heavy rainfall in the next hour, with recovery taking place thereafter. Updates to the physics continue to produce improved forecasts at longer forecast hours, as the biases for forecast hours 5–6 are nearly zero for features exceeding 25 mm h^{-1} , and the overprediction of features with heavy rainfall at FH4 is significantly reduced from 2014.

e. Onset of convection

While not studied in BK15, an evaluation of how the HRRR handles the onset of convection is of interest to help us better understand model processes and how model upgrades have affected forecasts of convective

precipitation. We already have some indication from Fig. 10 that the 2013 HRRR initiated convection around 1700 UTC, and that initiation started closer to 1800 UTC in the 2014 version of the model. Cai and Dumais (2015) found that the 2010 version of the HRRR tended to initiate convection somewhat later than observed with a 4-h lead time. Figure 13 shows the fraction of features identified at each valid forecast hour in the model (dashed gray), radar (solid black), and the difference between them (blue dotted), for all three years at FH1. Figure 14 displays the same for FH3.

In forecast hour 1, the 2013 HRRR starts increasing the fraction of predicted features at approximately 1700 UTC, consistent with the results from Fig. 10, but, contrary to the results of Cai and Dumais (2015), approximately 2 h before the increase in the fraction of observed features. The conflicting results are likely due to the Cai and Dumais (2015) study examining a different version of the model, at a different lead time, and across a different domain, as well as the previous study looking at the absolute number of identified features rather than the fraction of total features identified at each forecast valid time. The 2013 HRRR also dissipates existing convection too quickly, or fails to produce convection during the overnight and morning hours. This is not too surprising: without the forcing and instability caused by daytime heating, long-lived convective features that were not already in existence at sunset would be difficult to produce.

By 2014, the model still initiates convection too early, although it appears to be slightly more in line with the radar results. Overall, the fractional difference between the model and the radar is smaller than in 2013, and the overnight to early morning hours (0400–1600 UTC) are very closely in line with the observations. While the tendency to dissipate convection too early or to fail to

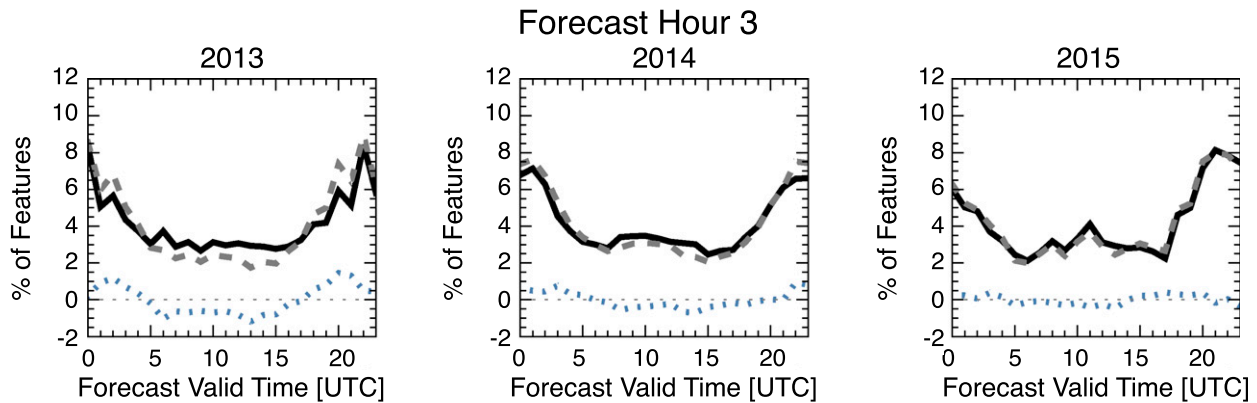


FIG. 14. As in Fig. 13, but for FH3.

form evening convection remains, it is limited to 2200–0300 UTC. Given that these changes are found in FH1, they are likely a result of the DA changes. Whereas the 2013 version rapidly intensified assimilated storms within the first forecast hour, the 2014 version seems to have developed assimilated rainfall into convective storms more slowly.

The 2015 HRRR shows significant improvements in capturing the onset of afternoon convection. The tendency to dissipate or underproduce convection during the evening/overnight hours is also reduced. Since major changes to the DA were made in the 2014 version and physics changes were primarily made in the 2015 version, the continued improvement in convective onset is likely attributable to the improved representation of boundary layer stability and subgrid cloud processes.

As with the southward bias, Fig. 14 indicates that the issues with convective onset are generally more evident in FH1, further indicating a continued need for model spinup. By FH3, the 2013 HRRR still ramps up convective production a bit early, but falls more in line with the observations by 2000 UTC. The lack of overnight convection in the absence of strong forcing is still evident in the same magnitude as at FH1. The 2014 HRRR slightly overpredicts the fraction of storms occurring between 1900 and 0300 UTC at FH3, while convective onset appears well represented at 3-h lead time in the 2015 version of the model.

In addition to the desire to gain an improved understanding of how the HRRR represented convective initiation, there was also some anecdotal evidence that the timing of the storm production was dependent on the size of the features of interest. In particular, it was posited that smaller storms were generally produced too early, while larger, more mature storms were too slow to form. Figure 15 shows the fraction of features by forecast valid time at FH1 for features smaller (larger) than

750 km² (5000 km²). These size thresholds represent approximately the smallest and largest 25% of features.

Figure 15 shows that the smallest ~25% of forecast features do in fact form well before the smallest observed features in 2013 and 2014. In 2015, the timing is much improved. Despite the improved timing, all versions of the HRRR create a much larger fraction of its smallest identified features in the afternoon hours than are observed.

With respect to the largest 25% of features, Fig. 15 shows that the 2013 HRRR produces these more mature features about an hour too early. By 2014 and 2015, the model times the initial onset of larger features more appropriately, but is not creating a large enough fraction of them.

One hypothesis as to why the model initiates a large number of small features but lags behind in the larger features relates to the spinup discussed previously. Many mesoscale convective features in the United States start out as a number of smaller features that later coalesce into a larger system. Current model capabilities are not able to represent this process in the 1 h between initialization and the first forecast, though there has been improvement through updated DA and physics. Figure 16 shows the difference in the number of forecast and observed features identified at various size thresholds for forecasts valid at 1800 UTC at a variety of lead times. As in Fig. 7, the 2013 HRRR significantly overpredicts the smallest features at all forecast lead times, but most appreciably at FH1. The overestimation of the number of features decreases in each increasing size category, with the overestimation of the number of features at 3- and 5-h lead times surpassing that at FH1 in larger size bins. In the 2014 and 2015 versions of the model, the level of overprediction of the smallest category of features decreases significantly in the first forecast hour. In 2014 all size bins show improvement, with

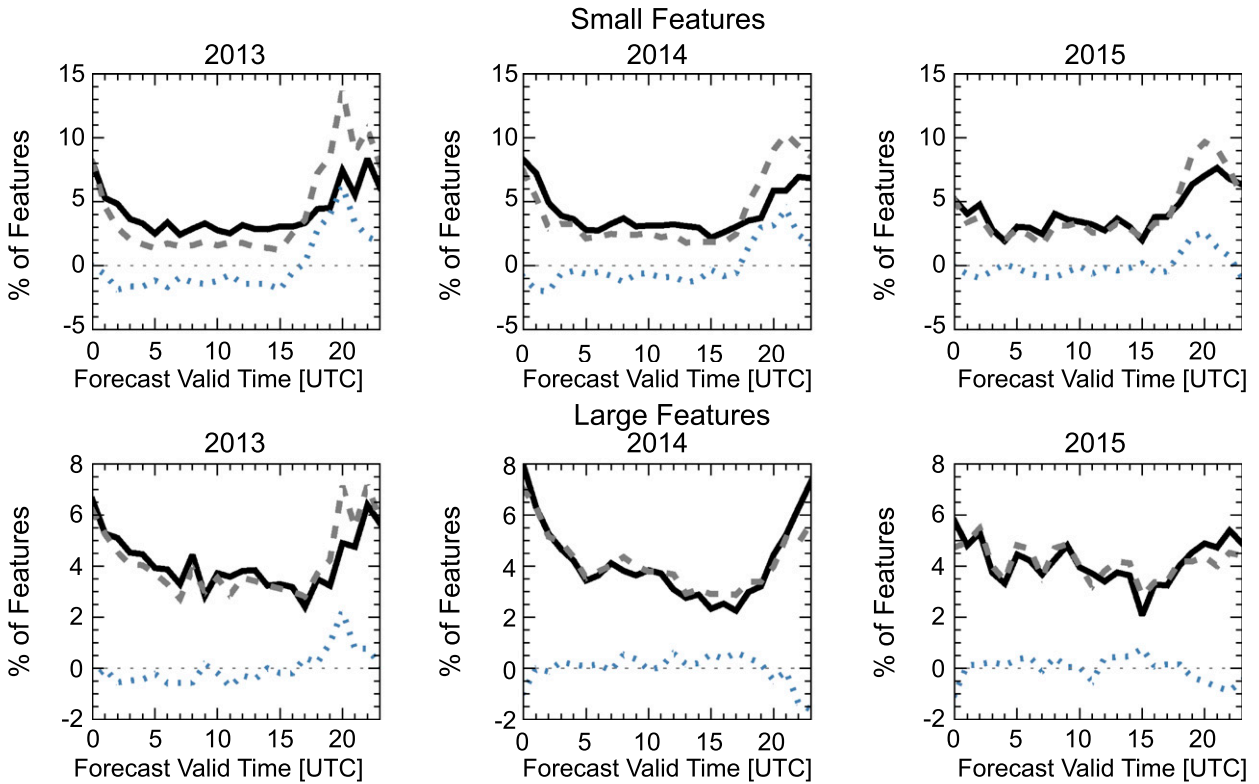


FIG. 15. As in Fig. 13, but for the (top) smallest ~25% and (bottom) largest ~25% of features.

some underestimation of the number of features in the 500–1000-km² bin at FH3. This is possibly an indication that the model is now growing small features into larger features at too rapid a pace.

In 2015, the tendency to overforecast the number of small features at FH1 continues to decrease, with the model now underestimating the number of features in the 500–1000- and 2000–5000-km² bins. In the 500–1000-km² bin, the 1-h lead time now appears to have approximately the same level of underprediction as the 2014 HRRR at

FH3, which could potentially be related to the slight decrease in spinup discussed in section 4b.

5. Conclusions

The HRRR has undergone several years of development at ESRL, with the goal of consistently providing improved forecasts. Several changes made to the model physics, including to the representation of the surface and boundary layer, had the effect of reducing instability,

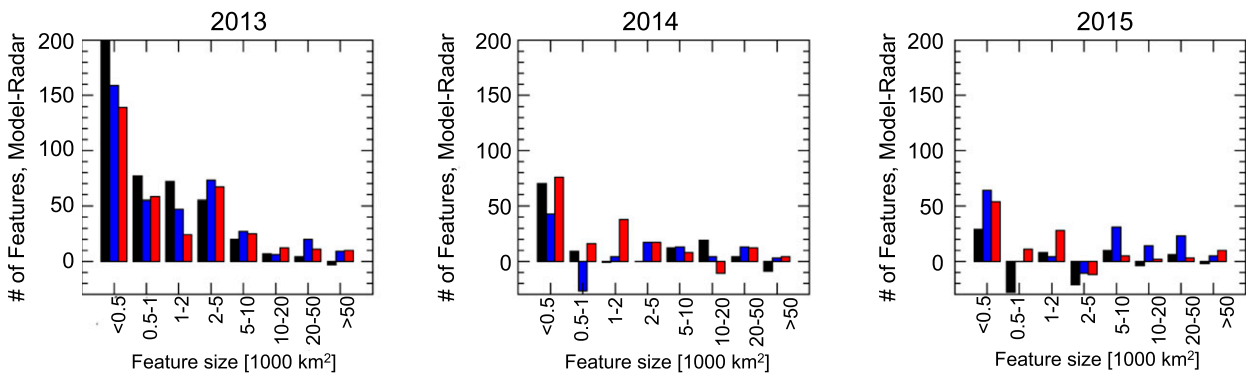


FIG. 16. Difference in the number of features identified by the model and radar in various size bins at forecast lead times of 1 (black), 3 (blue), and 5 (red) h for forecasts valid at 1800 UTC.

particularly in the southern United States where boundary layer forcing tends to produce more warm season convection than synoptically forced disturbances. Changes to the DA also improved the initial state representation, while reducing the model's ability to produce volatile convection early in the forecast. Overall, changes to the HRRR model have resulted in improved forecasts of central U.S. warm season convection and have addressed many of the issues presented in section 1. Generally, changes to the HRRR model have resulted in environments that are less conducive to premature convective initiation and less conducive to creating explosive convection. This has led to overall improvements in the bias distributions for mean, maximum, and total rainfall and areal coverage over the three years of development, with the most significant improvements related to biases in the maximum and mean rainfall intensity. Additionally, these changes have led to improved representation of convective initiation and reduced the tendency to overforecast very heavy rainfall in early forecast hours.

The southward bias discovered in BK15 was shown to be dependent on the time of the forecast, and not an issue with precipitation placement overall. The southward bias was found to correspond with the onset of convection, beginning at around 1700 UTC near the Gulf Coast, and migrating northward over the next several hours. This tendency for the HRRR to position convective rainfall south of its observed position was strongest in the 2013 version of the model and in the first hour of the forecast, with improvements such that the 2015 version of the model actually appears to have a bit of a northward bias at forecast hour 1.

While the forecast representation of storm sizes indicates significantly reduced spinup times in 2014 and 2015 (Figs. 8 and 9), the median biases of raining area and intensity indicate the continued need for model spinup. The continuing need for spinup is also evident in the tendency for the bias in the forecast fraction of heavy rainfall to worsen between forecast hours 1 and 2 in 2014 and 2015 (Fig. 12), the better representation of centroid placement and convective onset seen in the third forecast hour (Figs. 11 and 14), and the reduced difference between the number of predicted features in various size categories with increasing lead time (Fig. 16).

Based on the results of this study, upgrades made to the HRRR model in 2014 and 2015 have had the intended impact on convective precipitation forecasts, but there are still many biases remaining to be addressed with continued future updates.

Acknowledgments. This research was supported by Task 1 funds for Education received from NOAA by the

Cooperative Institute for Research in the Atmosphere (CIARA). Stage IV data were provided by NCAR/EOL under sponsorship of the National Science Foundation (<http://data.eol.ucar.edu/>).

REFERENCES

- Benjamin, S., and Coauthors, 2013: Data assimilation and model updates in the 2013 Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) analysis and forecast systems. *NCEP/EMC Meeting*, Washington, DC, NCEP/EMC/Model Evaluation Group, http://ruc.noaa.gov/pdf/NCEP_HRRR_RAPv2_6jun2013-Benj-noglob.pdf.
- , and Coauthors, 2016: A North American hour assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, doi:10.1175/MWR-D-15-0242.1.
- Bytheway, J. L., and C. D. Kummerow, 2015: Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central US. *J. Adv. Model. Earth Syst.*, **7**, 1248–1264, doi:10.1002/2015MS000497.
- Cai, H. and R. E. Dumais Jr., 2015: Object-based evaluation of a numerical weather prediction model's performance through forecast storm characteristic analysis. *Wea. Forecasting*, **30**, 1451–1468, doi:10.1175/WAF-D-15-0008.1.
- Davis, C., B. G. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, doi:10.1175/MWR3145.1.
- , —, —, and J. Halley-Gotway, 2009: The Method for Object-Based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Wea. Forecasting*, **24**, 1252–1267, doi:10.1175/2009WAF2222241.1.
- ESRL, 2015: High-Resolution Rapid Refresh (HRRR). Earth System Research Laboratory, <https://rapidrefresh.noaa.gov/hrrr/>.
- Lin, Y., and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., P1.2, https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.
- Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. *Wea. Forecasting*, **31**, 371–394, doi:10.1175/WAF-D-14-00112.1.
- Prat, O. P., and B. R. Nelson, 2015: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012). *Hydrol. Earth Syst. Sci.*, **19**, 2037–2056, doi:10.5194/hess-19-2037-2015.
- Smalley, M., T. L'Ecuyer, M. Lebsock, and J. Haynes, 2014: A comparison of precipitation occurrence from the NCEP Stage IV QPE product and the CloudSat cloud profiling radar. *J. Hydrometeorol.*, **15**, 444–458, doi:10.1175/JHM-D-13-048.1.
- Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, doi:10.1175/JAS-D-13-0305.1.
- , P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.