Evaluation of HRRR skill by quarter

Steve Weygandt, Curtis Alexander, Stan Benjamin

NOAA ESRL GSD for FAA AWRP MDE

1. Overview

This is an assessment of the HRRR skill by quarter for the past four quarters. The included quarters (grouped by fiscal year):

FY14 Q3 (April-May-June 2014)

FY14 Q4 (July-August-September 2014)

FY15 Q1 (October-November-December 2014)

FY15 Q2 (January-February-March 2015)

It is important to note the following factors that impact the skill performance comparisons:

(1) The NCEP HRRR operational was implementation occurred on Sept. 30 2014 (very end of FY14 Q4). Before that time, there was only a single version of the HRRR, run at GSD. This is the version that was implemented at NCEP (sometimes referred to as HRRRv1).

(2) Given (1), skill differences in the HRRR between FY14 Q3 and FY14 Q4 reflect only the difference in the forecast difficulty between spring and summer. On average, spring convection is more strongly forced and more predictable. The first set of plot comparisons will highlight this difference.

(3) Following the NCEP implementation of the HRRR, skill comparisons can be made between the NCEP operational HRRR (HRRRv1) and the evolving HRRRv2 run at GSD. In general, scores improved with time for the HRRRv2 relative to the HRRRv1 as further improvements were made in the HRRRv2.

2. HRRR skill for FY14 Q3 vs. FY14 Q4

Un-matched HRRR skill score comparisons are shown in Fig. 1 (Critical Success index – CSI) and Fig. 2 (bias). While the cases are not event matched (not possible for this seasonal comparison), the large number of cases in each sample provides high confidence in the differences noted. As noted above, the pre-implementation version of HRRRv1 code was unchanged through the period (with the exception of a few minor modifications that did not affect forecast skill) and is the code version is what was implemented at NCEP in Sept. 2014. Thus, skill differences between the two quarters reflect the differences in the character of the convection and the predictability off it between spring and summer. As expected skill is higher

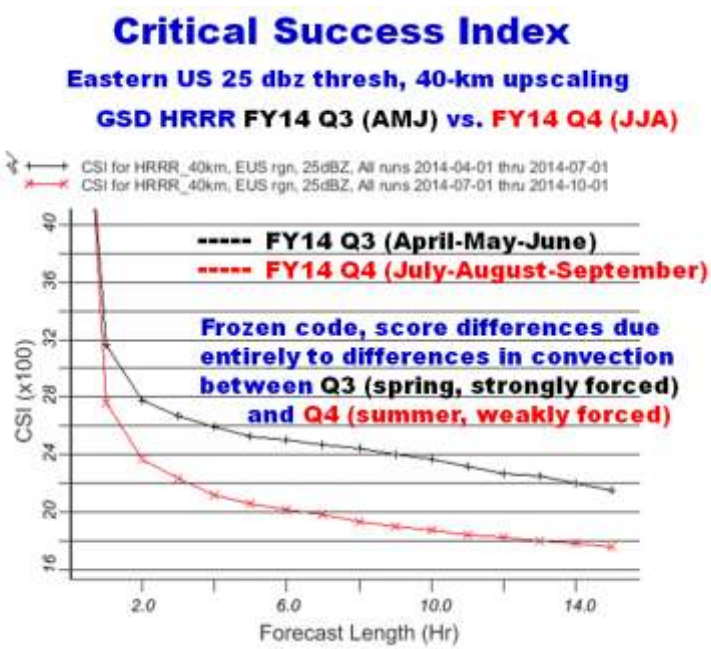

Fig 1. Comparison of HRRR (version 1) CSI scores for FY14 Q3 (AMJ, black curve) vs. FY14 Q4 (JAS, red curve). Scores are for 25 dbz reflectivity threshold up-scaled to 40-km grid over the Eastern U.S.

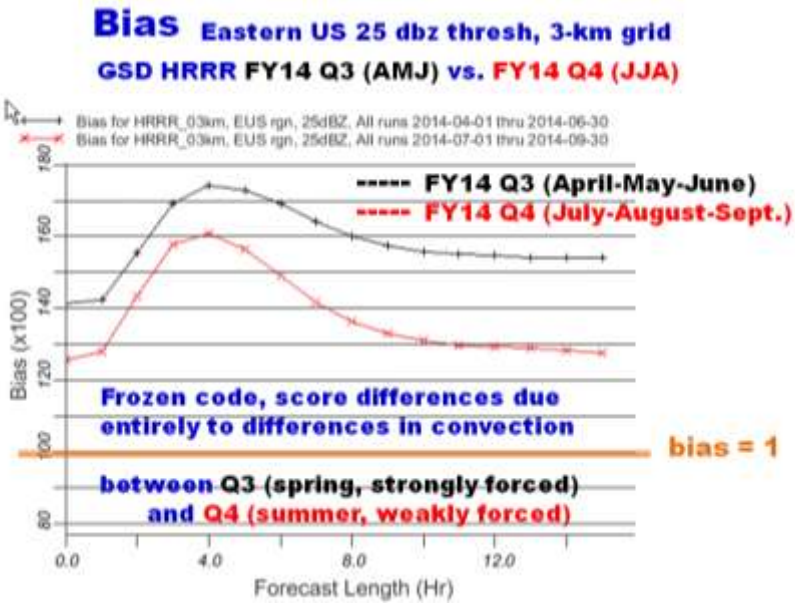

Fig 2. Comparison of HRRR (version 1) bias scores for FY14 Q3 (AMJ, black curve) vs. FY14 Q4 (JAS, red curve). Scores are for 25 dbz reflectivity threshold on the native 3-km grid over the Eastern U.S.

during the spring months (FY14 Q3, black curve), when convection is more strongly forced by travelling synoptic and mesoscale waves and associated fronts and boundaries. In contrast, In contrast, during summer months (FY14 Q4), there are fewer large-scale waves and convection is smaller-scale and more scattered. This makes the convection less predictable, as seen in the lower CSI scores throughout the 15-h forecast period. The bias differences (somewhat more

significant overprediction of the 25 dbz area in the springtime) are associated with a tendency of the model to overpredict larger convective systems, especially over the plains. This bias has been improved over the past several years.

2. HRRR skill for FY15 Q1 vs. FY15 Q2

Following the NCEP operational implementation of the HRRR and completion of the 2014 warm season evaluation (for which the GSD HRRR code version was frozen), GSD scientists began making changes and updates to further improve the HRRR for the 2015 warm season evaluation and the HRRRv2 operational implementation at NCEP (expected in Q4 FY15). The changes include a mix of planned upgrades to improved versions of model and assimilation components (including assimilation of new observations) and specific changes motivated by investigation of shortcomings during the 2014 warm season evaluation. Specific among the latter category was a broad set of changes to address an afternoon and evening warm, dry bias in the HRRRv1 (NCEP operational HRRR). The overall change package is summarized in the MDE monthly reports. As testing of improvements was completed in retrospective cycles involving the RAP and the HRRR, changes were promoted into the GSD real-time experimental HRRR-x. Some of the changes were incorporated during the fall 2014 (FY15 Q1) with the bulk going in early in 2015 (FY15 Q2). Results presented here will focus on just CSI and illustrate that:

(1) the GSD HRRR-v2 does perform better then the NCEP HRRR-oper

(2) the improvement of in the HRRR-v2 relative to the HRRR-oper is greater for FY15 Q2 than FY Q1, reflecting the more positive impact as more of the upgrades were introduced into the GSD HRRR-v2.

(3) the predictability of convection was less for FY15 Q2 (JFM 2015) than FY15 Q1 (OND 2014). This is expected, given the atmosphere is the most stable for the fall period (OND), so the 25 dbz reflectivity contour is most associated with more predictable larger-scale weather systems.

These results are illustrated in Fig. 3, a four way comparison between:

HRRR-v2 FY15 Q1 (OND 2014) --- black curve
HRRR-v2 FY15 Q2 (JFM 2015) --- blue curve
HRRR-oper FY15 Q1 (OND 2014) --- yellow curve
HRRR-oper FY15 Q2 (JFM 2015) --- red curve

Starting with the third point above, we can see from Fig. 3 the following:
(3) Compare the yellow curve (HRRR-oper, OND) to the red curve (HRRR-oper JFM). The higher skill score for the yellow curve for a model version that is frozen (HRRR-oper) indicates that the predictability was lower for the January-February-March period than the October-November-December period.

(1) Comparison of the black curve (HRRR-v2 OND) with the yellow curve (HRRR-oper OND) and the blue curve (HRRR-v2 JFM) with the red curve (HRRR-oper JFM) reveals that in both quarters, the evolving GSD HRRR-v2 skill exceeded the HRRR-oper skill. Note, this is to be expected as improvements are made to for the next HRRR implementation and is not a poor reflection on the operational HRRR.

(2) Noting the difference between the black and yellow curves (OND) is less than the difference between blue and red curve (JFM), we can see that the improvement of the HRRR-v2 relative to the HRRR-oper is greater for FY15 Q2 than FY15 Q1. This is consistent with fact that additional improvements were added into the GSD real-time experimental HRRR-v2 prototype for the FY15 Q2 comparison. This greater improvement can also be seen in that the blue curve (HRRR-v2 JFM) nearly matches the black curve (HRRR-v2 OND), which means that the improvement in skill in the HRRR-v2 from OND to JFM nearly counteracts the decreased predictability of the JFM period compared to the OND period.

All of these results, along with the more recent verification and the case examples indicate that the RAP/HRRR change package for 2015 has resulted in further significant improvements in HRRR skill. This version will be implemented at NCEP in Q4 2015.

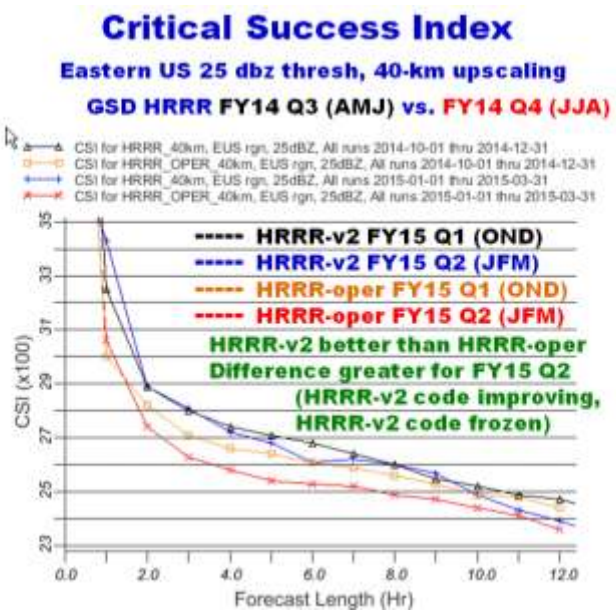

Fig 3. Comparison of HRRR CSI scores for HRRR-v2 FY15 Q1 (OND, black curve), HRRR-v2 FY15 Q2 (JFM, blue curve), HRRR-oper FY15 Q1 (OND, yellow curve), HRRR-oper FY15 Q2 (JFM, red curve). Scores are for 25 dbz reflectivity threshold up-scaled to 40-km grid over the Eastern U.S