WRF Core Test

Ceiling and Visibility Performance for the ARW and NMM Dynamical Cores

14 July 2006

Jennifer Simard, Paul Herzegh and Richard Bateman

NCAR/RAL, FAA/AWRP National Ceiling & Visibility PDT

Executive Summary:

This study compares the performance of the ARW and NMM dynamical cores for WRF forecasts of ceiling and visibility out to 12 hours. Study results are based upon bulk performance statistics (*e.g.*, CSI, POD, FAR) derived from four-season verification. In all cases the ceiling and visibility forecast values were derived from the NCEP post-processor.

Overall, our evaluation found no significant systematic difference in ceiling and visibility forecast skill between the two cores. Phase 1 results (wherein Ferrier microphysics were used for both cores) showed a very slight performance advantage in the ARW over the NMM for ceiling, and to the contrary a very slight advantage in the NMM over the ARW for visibility. Phase 2 results (utilizing NCAR microphysics for both cores) yielded no evidence of ceiling or visibility forecast performance advantage in either core.

Ceiling and visibility forecast skill improved noticeably in Phase 2 over Phase 1 for both cores. Thus, a secondary result of this study is that the Phase 2 physics package significantly benefits ceiling and visibility forecast applications. Forecast performance differences between the two cores were diminished in Phase 2 over Phase 1.

---

1. Introduction

This study utilizes a statistical approach to compare the ceiling and visibility forecast skill performance of the WRF model using the ARW and NMM dynamical cores. The approach allows for efficient processing of the datasets available and provides insight into season-averaged performance characteristics. In future work, more detailed characterization of possible differences may be explored through a case study approach, and through selection of geographic or time-of-day analysis domains. Extension from the current 12-hour limit to 24-hour forecasts would also be desirable.

2. Model Output Post-Processing

Both cloud base height (a proxy for ceiling) and visibility are products of model post-processing. The NCEP post-processor was used to prepare both fields. Early plans to extend the study to include products from RUC-equivalent post-processor algorithms were unattainable due to the heavy tasking faced by DTC staff. We expect that future work will close that gap.

Because the output field for cloud base height included both height and a numerical data flag value (-5000 denoting cloud-free conditions), the interpolation of cloud base height to the final RUC 13 km grid used in the analysis phase had to both preserve the flag value and avoid use of the flag value in any interpolation of cloud base height. This was accomplished through use of a nearest-neighbor interpolation scheme productively applied by Meral Demirtas of DTC.

3. Verification Processing

The verification process used is a close derivative of that used within AWRP as part of the AWTT evaluation of ceiling and visibility products. Output fields of ceiling (actually cloud base height) and surface visibility were interpolated to the horizontal locations of approximately 1,690 METAR sites within the ConUS using a nearest-neighbor scheme. Each 3h, 6h, 9h and 12h forecast of ceiling and visibility was verified against the METAR observation at the forecast valid time.

Verification addressed the forecast occurrence of ceiling < 1000 ft agl and visibility < 3 statute miles. These thresholds define IFR (Instrument Flight Rules) conditions, which are of major importance to aviation. A two-by-two contingency matrix holds all verification results organized as

- Hits: Forecast yes / Observed yes
- Misses: Forecast no / Observed yes
- False Alarms: Forecast yes / Observed no
- Null Events: Forecast no / Observed no.

Performance statistics derived from contingency matrix elements included CSI (critical success index), POD (probability of detection), FAR (false alarm ratio), frequency bias (ratio of forecasted to observed events), Heidke Skill Score and Peirce Skill Score. A reasonable assessment of forecast skill can be drawn from examination of CSI, POD, FAR and frequency bias. These values will be presented below.

4. Data Examined

The datasets available for examination were organized into two groups according to model physics used. Phase 1 data result from use of Ferrier microphysics, while Phase 2 data show results from the use of NCAR microphysics.

Ceiling: Due to limited time available to complete model post-processing, Phase 1 and Phase 2 ceiling data from the following two-week periods were verified:

- Fall: 15 Nov – 30 Nov 05
- Winter: 15 Jan – 30 Jan 06
- Spring: 25 March – 9 April 06
- Summer: 15 July – 30 July 05

Spring and summer results (though yielding somewhat fewer events) followed the same trends as fall and winter results. For brevity, only fall results will be shown in the figures below.

Visibility: Phase 1 data were obtained for the following full-month periods:

- Fall: 1-30 Nov 05
- Winter: 15 Jan – 15 Feb 06
- Spring: 25 March – 25 April 06
- Summer: 15 July – 15 August 06

To accommodate post-processing limitations, Phase 2 data were examined over the same set of two-week periods outlined for ceiling above.

5. Ceiling Results

Since the results for the fall season are characteristic of those found overall, we use fall data in this section and the next to illustrate our findings.

As shown in Fig. 1, CSI (a useful metric to estimate skill in the correspondence between 'yes' forecasts and 'yes' events) and POD for the ARW very slightly exceed those for the NMM during Phase 1. During Phase 2 this difference disappears.

Results for FAR shown in Fig. 2 are consistent with the results above. Slightly lower values of FAR (a lower false alarm ratio) are shown for the ARW in Phase 1, but in Phase 2 this slight difference disappears.

As shown by the frequency bias values in Fig. 2, both cores have a strong tendency to overforecast low ceilings in both Phases. Overforecasting is somewhat lower in magnitude during Phase 2, as is the frequency bias difference between the ARW and NMM. These tendencies illustrate that frequency bias characteristics are strongly affected by model physics, and thus are not a direct indicator of dynamical core performance.

6. Visibility Results

As shown in Fig. 3, CSI and POD for the ARW very slightly exceed those for the NMM during Phase 1. During Phase 2 this difference disappears.

Results for FAR shown in Fig. 4 indicate a slightly lower false alarm ratio for the ARW during both Phases. This is not likely a significant performance indicator, as frequency bias for the ARW is significantly lower than that for NMM in both Phases. With a tendency to overforecast comes a corresponding tendency toward false alarms.

Frequency bias values much less than unity in Fig. 4 illustrate a strong tendency by both cores to underforecast visibilities < 3 miles. This tendency is somewhat greater in the ARW, but is significantly diminished in Phase 2. As mentioned in Sec. 5, frequency bias characteristics can be strongly affected by model physics, and thus should not be taken as a direct indicator of dynamical core performance.

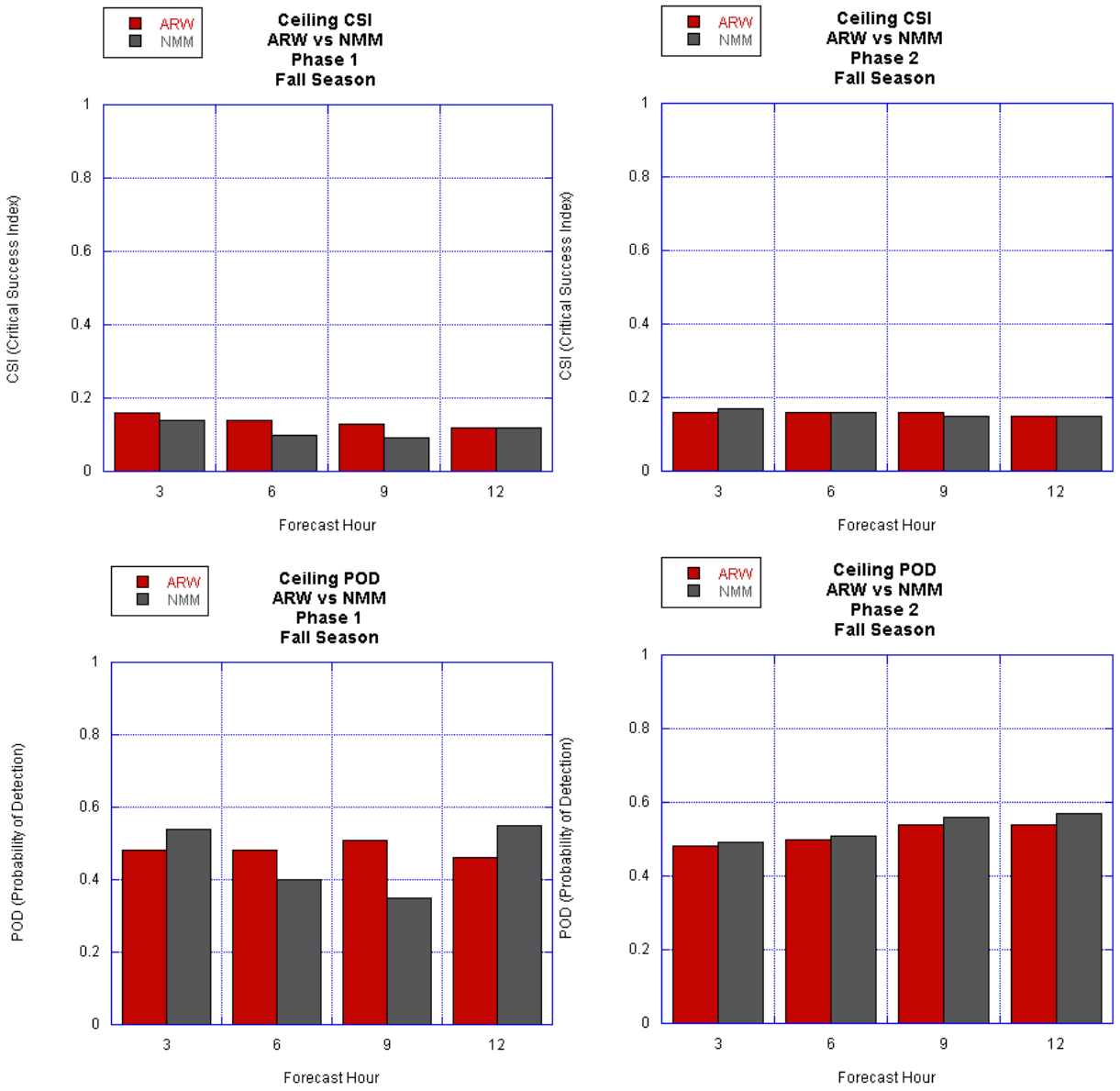

Fig. 1. CSI and POD for fall season forecasts of ceiling < 1000' agl for Phase 1 (left) and Phase 2 (right). While CSI and POD values for ARW slightly exceed those for NMM in Phase 1, no significant advantage is held by either core during Phase 2. Note that Phase 2 performance generally exceeds that for Phase 1.

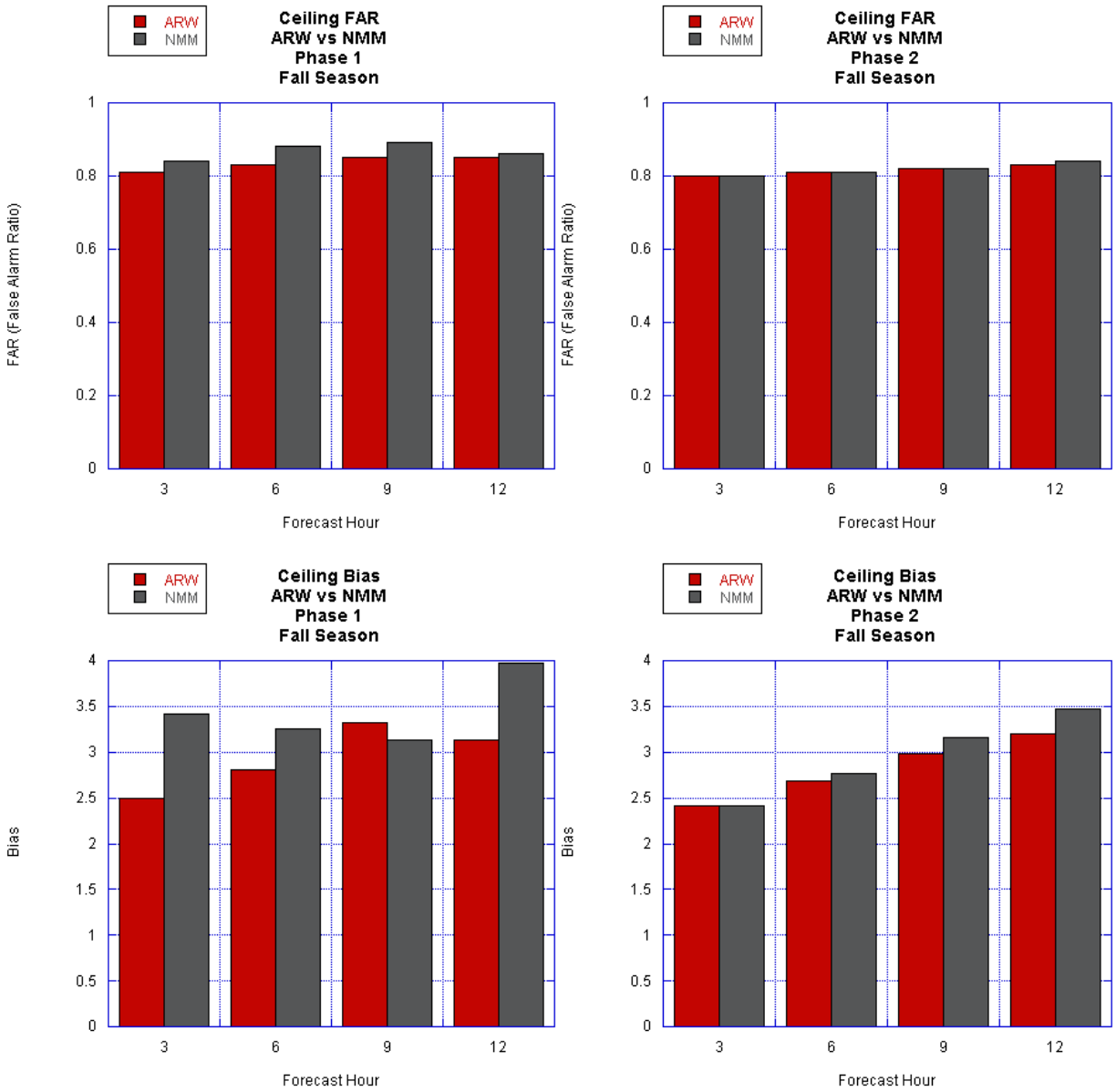

Fig. 2. FAR and frequency bias for fall season forecasts of ceiling < 1000' agl for Phase 1 (left) and Phase 2 (right). While FAR values for ARW are slightly better (lower) than those for NMM in Phase 1, no significant advantage is held by either core during Phase 2. Note that Phase 2 performance generally exceeds that for Phase 1.

Frequency bias values much greater than unity illustrate a strong tendency by both cores to overforecast ceilings < 1000' agl. Overforecasting is somewhat lower in magnitude during Phase 2, as is the frequency bias difference between the ARW and NMM. These tendencies illustrate that frequency bias characteristics are strongly affected by model physics, and thus are not a direct indicator of dynamical core performance.

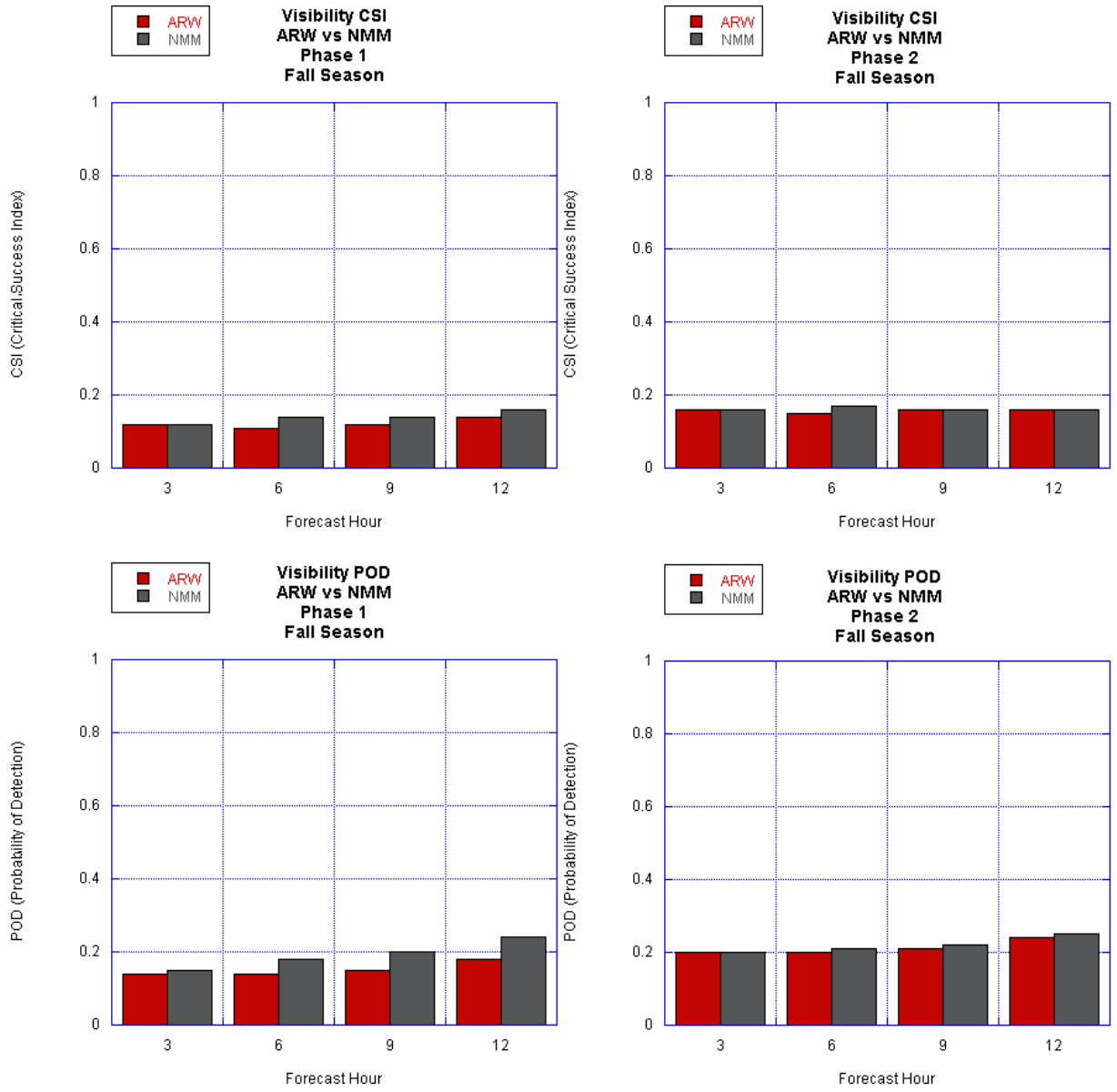

Fig. 3. CSI and POD for fall season forecasts of visibility < 3 miles for Phase 1 (left) and Phase 2 (right). While CSI and POD values for NMM very slightly exceed those for ARW in Phase 1, no significant advantage is held by either core during Phase 2. As seen for ceiling in Fig. 1, Phase 2 performance slightly exceeds that for Phase 1.

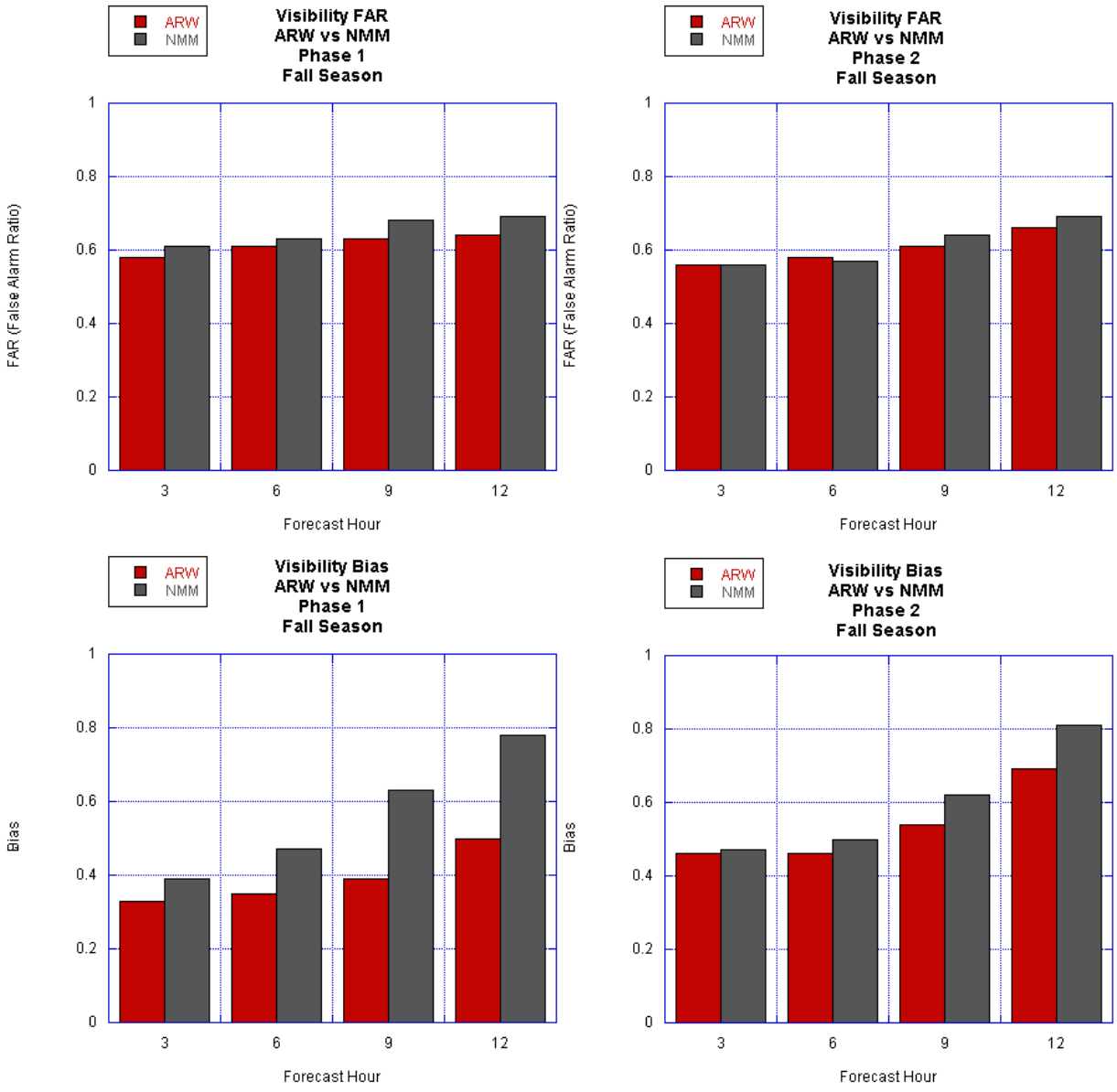

Fig. 4. FAR and frequency bias for fall season forecasts of visibility < 3 miles for Phase 1 (left) and Phase 2 (right). While FAR values for ARW are slightly better (lower) than those for NMM in Phase 1, this slight ARW advantage is seen only at 9 and 12 h during Phase 2. Note that Phase 2 performance generally exceeds that for Phase 1. Frequency bias values much greater than unity illustrate a strong tendency by both cores to overforecast ceilings < 1000' agl.

Frequency bias values much less than unity illustrate a strong tendency by both cores to underforecast visibilities < 3 miles. This tendency is somewhat greater in the ARW, but is significantly diminished in Phase 2. This illustrates that frequency bias characteristics can be strongly affected by model physics, and thus are not a direct indicator of dynamical core performance.