The Developmental Testbed Center

2007 13-km Dynamic Core Test Report

Ligia Bernardet

October 10, 2008

Contributors: The Developmental Testbed Center (DTC) 2007 13-km Core Test (CT2007) could not have been accomplished without the participation of many individuals and organizations. From the National Center for Atmospheric Research (NCAR) Research Applications Laboratory, the CT2007 had the participation of Barbara Brown, Randy Bullock, Robert Gall, Eric Gilleland, John Halley-Gotway, Lacey Holland, Louisa Nance, Beth Weekley, Jamie Wolff, and Lara Ziady. From the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory, the following participated: Ligia Bernardet, Mark Govett, Christopher Harrop, Steven Koch, Andrew Loughe, Jeff Smith, and Betsy Weatherhead. Additionally, the following model developers were helpful in revising the CT2007 design and developing code: Michael Duda, Jimy Duhia, Dave Gill, and Joe Klemp from NCAR, and Tom Black, Hui-ya Chuang, Geoff DiMego, Michael Ek, Zavisa Janjic, and Matthew Pyle from the NOAA National Centers for Environmental Prediction.

Index

1. Executive Summary

The main goal of the CT2007 was the compilation and intercomparison of verification statistics for forecasts run with the two dynamic cores of the Weather Research and Forecasting (WRF) Model: the Non-hydrostatic Mesoscale Model (NMM) and the Advanced Research WRF (ARW) cores. To this end, the WRF model was configured and run similarly for both cores, limiting the setup differences to the dynamic solvers. The following results summarize the findings:

- The annual mean of the precipitation bias reveals statistically significant (SS) overprediction by both cores at the lowest thresholds. The seasonal breakdown indicates that most of the overprediction occurs in spring and summer, while the fall has underprediction at the higher thresholds. The magnitude of the bias remains fairly constant throughout the forecast period.
- The precipitation Equitable Threat Score (ETS) is higher for the lower thresholds and decreases in time for all thresholds. The highest ETS occurs in the fall and winter, and the lowest in summer.
- The intercore differences in precipitation bias and ETS are not SS.
- The temperature bias is largest at 100 hPa. At the end of the forecast period, the median reaches 2.1 ^{o}C for the NMM and 2.7 ^{o}C for the ARW, characterizing the largest temperature bias intercore difference. Smaller SS intercore temperature bias differences occur at other levels and lead times, some favoring the ARW (especially at 2-m AGL) and others the NMM (especially at upper levels).
- The two cores produce very similar temperature bias-corrected root mean square errors (BCRMSEs). The largest values occur at 200 hPa at the 60-h lead time, with approximately 2.2 ^{o}C.
- Both cores have positive relative humidity (RH) bias at all times, with the NMM displaying smaller bias at several levels and lead times.
- The RH BCRMSE peaks at 500 hPa with approximately 27% but the intercore differences are not SS.
- Both cores have maximum wind error at 250 hPa, where the approximate values for bias and BCRMSE for the 60-h forecast reach -1.2 and 10.0 ms^{-1}, respectively.
- The ARW produces faster winds than the NMM in the lower troposphere, while the NMM has faster winds in the upper troposphere.
- The ARW wind speed absolute bias is smaller than the NMM's at several levels and lead times. Conversely, the NMM has smaller absolute bias than the ARW at 10-m AGL at all lead times.

- The NMM forecasts have lower wind BCRMSE than the ARW at all lead times at 10-m AGL, 850 hPa and 700 hPa, and at 24- and 48-h lead times at 500 hPa.
- The temperature, RH and wind BCRMSE grow for both cores throughout the forecast period, but the intercore differences do not. In general, the bias for these variables does not grow in time, with the exception of temperature at 700 and 100 hPa and wind at 250 hPa. The bias intercore difference only grows in time for temperature at 100 hPa.

2. Introduction

The Weather Research and Forecasting (WRF) model contains two dynamic cores: the Non-hydrostatic Mesoscale Model (NMM – Janjic 2003, 2004) core, developed at the National Centers for Environmental Prediction (NCEP), and the Advanced Research WRF (ARW – Skamarock et al. 2005) core, developed at the National Center for Atmospheric Research (NCAR). Each dynamic core corresponds to a set of dynamic solvers that operates on a particular grid projection, grid staggering, and vertical coordinate. The WRF model also contains a multitude of physical parameterizations, many of which can be used with both dynamic cores.

The WRF model is currently run operationally at several weather forecasting centers, including NCEP and the Air Force Weather Agency (AFWA). NCEP runs the NMM in its North American Mesoscale (NAM) model application and runs both the ARW and NMM in the High-Resolution Window (HRW) and Short-Range Ensemble Forecasting (SREF) applications. Additionally, NCEP is scheduled to start operations of the WRF Rapid Refresh application in 2009 using the ARW. AFWA runs the ARW over a variety of domains in support of its Department of Defense operations.

Several comparisons of forecasts by the ARW and NMM dynamic cores have been conducted in the past, including the WRF Test Plan (Seaman et al. 2004), the NSSL Spring Program (Kain et al. 2005), the Developmental Testbed Center (DTC) Winter Forecasting Experiment (DWFE, Bernardet et al. 2008) and the Rapid Refresh Core Test (RRCT, Brown et al. 2007). In all of these studies except for the Rapid Refresh Core Test, differences between the configurations using ARW and NMM went beyond the dynamic core to include variations in initial conditions, physics packages etc. Therefore, a formal assessment of the differences between the forecasts by the two cores was not possible until the RRCT. The results of the RRCT indicated that, out to a 24-h lead time, the differences between ARW and NMM forecasts for the configuration employed were quite small. The forecast from each core had strengths and weaknesses, as described in Nance (2006), with the ARW fairing an overall mild advantage.

The similarity in objective verification scores between ARW and NMM obtained during the RRCT led to the idea that research results obtained with one core could be transferred to the other. However, to ascertain this possibility, it was deemed necessary to test whether the similarity in results was limited to the first 24 h or could be carried onto a 60 h forecast. With this motivation, the CT2007 was designed with two goals: 1) to quantify the errors in the forecasts produced by the ARW and NMM dynamic cores of the WRF model in a given configuration, and 2) to quantify the differences between forecasts produced by the two dynamic cores in a given configuration. Both analyses were based on objective forecast verification statistics of the model. An ancillary goal of the CT2007 was to quantify the differences in forecast verification statistics from models run on different platforms. This goal was addressed by running a subset of the forecasts on two platforms. A final goal was to make all results accessible to the community. Therefore, products have been archived and results summarized in this report. It is expected that results from this test will be used as a baseline for further testing.

Since the CT2007 is an extension of the RRCT, many parameters in the end-to-end forecast system, including forecast domain and grid spacing, were not altered from the RRCT. However, the WRF Preprocessing System (WPS) was used instead of the WRF Standard Initialization (SI), the version of all codes was updated, and the source of initial conditions was altered. The need to update the model and preprocessor source code stemmed from the importance of using the latest developments in the physics and dynamics of the model, and from the need to use code that is in the WRF repository to maximize support from the developers. This approach also keeps the results more relevant to the WRF user community.

Model output was ingested into the WRF Postprocessor (WPP) to obtain de-staggered grids interpolated in the vertical to isobaric levels and in the horizontal to a common grid. Postprocessed forecasts were verified using the NCEP Verification System and used to generate images.

The timeline for this test was as follows:

- **March 2007:** CT2007 design document and proposed namelists were sent to WRF developers.
- **May 2007:** NMM developers worked with DTC to change smoothing of topography to be more consistent with ARW.
- **June 2007:** Feedback from WRF developers, recommending the substitution of the NAM model physics package for the Rapid Update Cycle (RUC) physics package, including the unified Noah Land Surface Model (LSM) was received and incorporated in the CT2007 design document.

- **September 2007:** Developers finalized code by implementing the unified Noah LSM in the WRF code repository.
- **September 2007:** Feedback from WRF developers, recommending the use of 58 vertical levels, reproducing the NAM model distribution of levels up to 50 hPa, was received and incorporated in the CT2007 design document.
- **October – November 2007:** Runs were completed and archived.
- **December 2007:** Bug affecting initialization of ARW was found.
- **January/2008:** Reruns were completed and archived.
- **February 2008:** Statistical analysis of results began.
- **September 2008:** Report was completed.

3. Experiment Design

This section describes the components of the forecast system employed in the CT2007. The WPS, WRF model, WPP, NCEP Verification System, graphics generation, and data archival were run as an end-to-end forecast system controlled by the NOAA Earth System Research Laboratory (ESRL) Workflow Manager, with the WRF Portal being used as an interface for some of the runs. The computation of verification scores and their confidence intervals (CIs) was done after the runs were complete.

a) Domains Configuration

Even though the ARW and NMM use different grid projections, making it impossible to exactly match the forecast domains of the two cores, extensive efforts were employed to minimize the differences in area coverage and number of grid points between the configurations. The ARW and NMM domains along with the common grid used for the postprocessing are shown in Fig. 1.

The specifications of each horizontal domain are:

- NMM: 280 x 435 grid points; total 121,800 grid points.
- ARW: 400 x 304 grid points; total 121,600 grid points.
- Postprocessing: 451 x 337 grid points; total 151,987 grid points.

Both dynamic cores use 58 vertical levels, which were distributed similarly in the two dynamic cores. An exact match in vertical levels is not possible because the ARW uses a sigma-pressure vertical

coordinate, while the NMM uses a hybrid system, with sigma-pressure levels below 300 hPa and isobaric levels aloft.

Several additional domains were used as sub-regions for verification. The results of the NCEP Verification System were area-averaged over the entire CONUS (G164), over the Eastern and Western sectors of the CONUS (Grids165 and 166) (Fig. 2) and over the 14 regional domains shown in Fig. 3 to compute regional averages of statistics.

b) Forecast Periods

To address the two main goals of this test, as described in the Section 2, the ARW and NMM dynamic cores were used to forecast 120 cycles divided into the four seasons (Table 1). The runs were initialized every 36 h, therefore alternating 00 and 12 hr cycles.

For the platform intercomparison test, redundant forecasts were computed on the LINUX platform. For the spring season, the same cycles were computed in the LINUX and IBM platforms. For the remaining seasons, only a third of the cycles were computed on LINUX, under the premise that the differences between platforms could be resolved with this subset.

Table 1. Range of dates in each season of the Core Test

Summer:	09 July – 24 August 2005
Fall:	10 October – 23 November 2005
Winter:	10 January – 22 February 2006
Spring:	10 April – 23 May 2006

c) Initial and Boundary Conditions

The Core Test employed initial and boundary conditions from NAM model. For the retrospective periods to be used, the NAM model corresponded to the Eta model.

The time-invariant component of the lower boundary conditions (topography, soil and vegetation type etc.) was generated through the geogrid program of WPS. The use of a unified pre-processor for the

two dynamic cores should minimize the differences in lower boundary conditions. The time-independent lower boundary condition generated on the IBM platform was used on both platforms.

d) Forecast Model

The WRF code used in this test does not correspond to a public release. Instead, a snapshot of the top of the WRF code repository as of August 29, 2007 was used. The choice of code was based on the need to use the latest code developments, especially the unified Noah LSM, which was not available in the public release at that time (WRF v2.2).

The timesteps used in the two dynamic cores are not the same due to differences between the numerical stability requirements for the two dynamic cores. The NMM used a timestep of 30 s, and the ARW a long timestep of 72 s with an acoustic timestep of 18 s. In order to maintain consistency between the two dynamic core configurations, the frequency of physics calls was set identically when possible and similarly otherwise. Calls to radiation were done every 30 minutes for both dynamic cores. Calls to the boundary layer, microphysics and cumulus parameterization in the ARW were done every time step (every 72 s), while in the NMM they were done every other timestep (every 60 s). Model output files were written every three hours.

The physics suite used for both cores is described on Table 2.

Table 2. Physics Suite used with the ARW and NMM cores.

Microphysics	Ferrier
Surface Layer	Janjic
Planetary Boundary Layer	Mellor-Yamada-Janjic
Convection	Betts-Miller-Janjic
Land-Surface Model	Noah
Radiation SW and LW	GFDL

The ARW has several aspects of the numerics that can be configured through a namelist ingested at runtime. The ARW was configured to use the 3rd order Runge-Kutta time integration, 5th-order horizontal advection and 3rd-order vertical advection, with a positive-definite advection scheme for moisture. Additionally, the ARW was configured to employ three-dimensional divergence damping

(coefficient 0.1), external model filter (coefficient 0.01), vertical velocity damping and a 5-km deep diffusive layer at the top of the domain (coefficient 0.02).

e) Postprocessing

The WPP (Chuang et al. 2004) was used to destagger the ARW and NMM forecasts and to interpolate them to a common Lambert-Conformal grid (Fig. 1). Additionally, the WPP was used to generate derived meteorological variables including mean sea level pressure, and to interpolate the forecasts to isobaric surfaces, as required by the plotting and verification programs.

f) Model verification

Model verification partial sums were generated with the NCEP Verification System (Chuang et al. 2004), comprised of the Surface and Upper Air Verification System and the Quantitative Precipitation Forecast (QPF) Verification System.

Through the NCEP Surface and Upper Air Verification System, forecasts were bilinearly interpolated to the location of the observations (METARs, RAOBS, and ACARS) contained in the RUC Prepbufr files. Partial sums (aggregations by geographical region using the mean) were generated for several domains. However, in this report, only the results aggregated over the CONUS will be presented. Additionally, in this report the upper air temperature, relative humidity (RH) and wind analysis will be restricted to comparisons against rawinsondes. Due to the limited availability of METAR and RAOB observations, upper air verification results are computed at 12-h intervals, while surface results are computed at 3-h intervals. Upper air verification starts as early as the initialization time, while surface verification starts at the 3-h lead time, since forecasts of shelter-level variables are not available at the model startup time. Verification metrics were computed for all mandatory levels except 1000 hPa, which is below ground level in many locations. Relative humidity results are displayed only up to 500 hPa, since the mixing ratio is very small aloft and the observational uncertainty is large.

For the precipitation verification, a grid-to-grid comparison, in which the forecasts and the precipitation analyses were first interpolated to the 12-km grid spacing Grid 218 and then compared, was used. Accumulation periods were 24 and 3 h. The observational datasets were the NCEP Stage II analysis for the 3-h accumulations and the River Forecast Center (RFC) analysis for the 24-h accumulation (valid at 12 UTC). In this report, the precipitation verification will be limited to the 24-h accumulations.

For an initial exploratory analysis of the results, the verification partial sums were ingested in a mySQL database, which was queried to compute and display, on a web-interface, time-averaged verification statistics.

The statistical results presented in this report were computed from the partial sums using routines developed by the DTC in the statistical programming languages R (for precipitation) and S-plus (for temperature, relative humidity and winds). The statistics were computed separately for the ARW and NMM models, and also generated for the difference between each ARW and NMM pair. All differences in this report use the convention ARW - NMM. The metrics used are frequency bias and Equitable Threat Score (ETS) for precipitation and bias and bias-corrected root mean square error (BCRMSE) for the continuous variables (temperature, RH and wind). The BCRMSE represents the errors without bias and is defined as the square root of the estimated variance of the error which, when summed to the square of the bias, amounts to the mean square error.

For precipitation, the temporal aggregation (all cycles of the test, or a breakdown by season) was done by summing up the contingency tables for all cycles involved and computing the scores based on the aggregated table. On the other hand, for the temporal aggregation of temperature, RH and winds, the median was used. Since it is a robust statistical measure, the median is a better representation for distributions that contain outliers. In this report, the seasonal breakdown will be presented just for precipitation.

The temporal median was computed using the entire sample for the chosen aggregation period but excluded instances whose partial sums were based on too few rawinsondes or surface observations. Computations of surface results did not include valid times in which less than 1,400 stations were present. Similarly, a cut off number of 14 rawinsondes was used to discard upper air results.

It is important to note that the CT2007 has a limited number of forecast cycles, so there is no guarantee that the mean and median of the sample match those of the population (all possible cycles, over many years). To quantify the uncertainty in the mean and median, CIs were computed. Confidence intervals for precipitation were created using a resampling (bootstrapping) technique, while the CIs for the other variables were computed assuming a normal distribution. Auto-correlation in the time series increases the uncertainty in the results, since it decreases the effective sample size. Since the runs were initialized every 36 h, auto-correlation was present and taken into account in the computation of CIs for

temperature, RH, and winds. For results aggregated separately for the 00 and 12 UTC initialization times, as done in this report for surface variables, the auto-correlation is much smaller, since the cycles are spaced every 72 h. For this reason, the auto-correlation correction was not used for precipitation.

All CIs shown are for the 99% level, indicating that there is only a 1% chance that the population mean or median falls outside the CI. The interpretation of the results in situations when the CI encompasses the value for a perfect forecast merits some explanation. The perfect bias for the continuous variables (temperature, RH, and winds) is zero, while it is one for precipitation, since the precipitation bias is defined as the ratio of the forecasted to the observed area for a given threshold. When the bias mean or median is clearly on one side or the other of the perfect value but the CI encompasses the value for a perfect forecast, it is not possible to say that the model differs significantly from the perfect value, and therefore it is not possible to know if the model is over or underpredicting that quantity. For the same reason, caution is needed when interpreting differences in scores. Even if the mean or median score has a clear sign, the models are not significantly different if the CI encompasses zero. The interpretation of intercore differences for temperature, RH, and wind bias also merits an explanation. When the CI on the median of this difference does not encompass zero, it is possible to say that there is a SS difference between the cores, with one of them being warmer, moister, or windier than the other. However, this is not enough to information to ascertain which core has less error. Knowing that one core is warmer, moister, or windier than the other only translates to knowledge about which core has smaller error if the two cores have biases of the same sign.

g) Graphics Generation and Display

Graphics for the ARW and NMM forecasts computed in each platform were generated using the NCAR Command Language (NCL). The suite of images was comprised of, among others, temperature, dewpoint, horizontal and vertical winds, precipitation, vorticity, composite reflectivity and soundings in Skew-T diagrams for several locations. Additionally, graphics for the intercore differences were generated for selected variables.

All graphics are displayed on the website http://www.dtcenter.org/plots/coretest2007.

h) Data Archival

Input and output from several stages of the end-to-end system have been archived to the NCAR Mass Storage System (MSS) including:

- NAM model files used for initial and boundary conditions.
- Datasets used for forecast verification (RUC Prepbufr and Stage II and RFC precipitation analyses).
- Static files produced by the ***geogrid*** component of WPS.
- Final output of WPS.
- Initial and boundary condition files produced by the ***real*** component of the WRF model.
- Output files produced by WRF.
- Outputs of the WPP, both before and after horizontal interpolation from the native grids to the common postprocessing grid.
- Prepfits and vsdb files output by the NCEP verification system.
- Images produced by NCL.

Additionally, all source codes and executables have been archived.

4. Results

The results of the objective verification of precipitation, temperature, RH, and winds are described below. All results represent spatial averages over the CONUS. The median temperature, RH and wind results were computed aggregating the 00 and 12 UTC cycles for upper air. The surface results were computed separately for the 00 and 12 UTC cycles to highlight the modulation of the errors by the diurnal cycle. Since surface results for 00 and 12 UTC are similar when for any given local time, only the 00 UTC results will be presented in this report.

a) Precipitation accumulated in 24 h

Since the RFC precipitation analysis used for verification is only published at 12 UTC, for the forecast cycles initialized at 12 UTC, verification is available for the 24 and 48-h lead times, while for the forecast cycles initialized at 00 UTC, verification is available for the 36 and 60-h lead times. The 24-h and 60-h results will be presented in this report.

The biases for the ARW and NMM for the 24-h lead time are shown in Fig. 4a for several thresholds. For the annual aggregation, overprediction is noted for both cores at the 0.01 and 0.10-in thresholds and underprediction occurs for the NMM at the 0.75-in threshold. For the other thresholds, the results are not conclusive regarding over or under-prediction. The CIs are noticeably larger for the higher thresholds, reflecting large variability and a smaller sample size.

When the seasonal distribution is examined, statistically significant (SS) bias results are present for additional thresholds (Table 3). Summer and spring have SS overprediction at the lower thresholds, while summer also displays a marked underprediction at the intermediate thresholds. Fall (for both cores) and winter (NMM only) have SS underprediction at high thresholds.

Table 3. Forecasts of 24-h accumulated precipitation for the 24-h lead time for both dynamic cores classified as under or overprediction at all time periods and thresholds. Only SS results are presented.

Threshold (in)	Time period	Prediction	Core
0.01	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
0.10	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
0.50	Annual	Under	NMM
	Summer	Under	Both
0.75	Annual	Under	NMM
	Summer	Under	Both
1.00	Summer	Under	Both
1.50	Fall	Under	NMM
2.0	Fall	Under	Both
	Winter	Under	NMM

At the 60-h lead time (Fig. 4b), the annual aggregation shows overprediction at the 0.01- and 0.10-in thresholds, and the results are inconclusive for higher thresholds. There is not a pronounced change in bias magnitude between the 24- and 60-h forecasts. The CIs are wider than at 24 h, indicating more variability in the sample. The seasonal decomposition is presented in Table 4. The differences among the seasons are more pronounced at the 60-h lead time, with the fall displaying the lowest mean biases at all thresholds and SS underprediction at thresholds above 0.75 in, while winter and spring have the highest mean biases at the mid thresholds, and summer and spring the highest mean biases at the upper thresholds.

The bias differences, as shown in Fig. 5, are small and are not statistically significant for any threshold. The CI for the 2-in threshold in winter could not be accurately computed due to high variability in a small sample.

The ETS for the 24–h lead time (Fig. 6a) displays its maximum aggregated values for the lowest thresholds and is smaller for higher thresholds. The seasonal breakdown indicates that summer has the lowest scores for thresholds up to 0.5 in. The 60-h results (Fig. 6b) indicate a loss of skill with lead time, especially for the summer season. Fall and winter have the highest aggregated forecast skill at all thresholds at both lead times.

Table 4. Same as Table 3, but for the 60-h lead time.

Threshold (in)	**Time period**	**Prediction**	**Core**
0.01	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
	Winter	Over	Both
0.10	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
	Winter	Over	Both
0.25	Spring	Over	Both
	Winter	Over	ARW
0.50	Summer	Under	ARW
1.00	Spring	Over	ARW
	Fall	Under	Both
1.50	Fall	Under	Both
2.0	Fall	Under	Both

The ETS differences between ARW and NMM, as shown in Fig. 7 for the 24- and 60-h lead times, are very small and are not statistically significant for any threshold. As in the bias plot, the CI for the 2-in threshold could not be computed due to high variability in a small sample.

b) Temperature

The vertical distribution of temperature bias for several lead times is shown in Fig. 8. At all lead times, both cores display SS cold biases at 700 hPa, topped by warm biases at 300, 250, and 100 hPa. The largest bias occurs at 100 hPa, where the magnitude surpasses 2.1 ^{o}C at 60 h. It is possible that observational error is partially responsible for the high bias results obtained at 100 hPa.

The intercore differences in bias reveal SS differences at several levels at lead times 24 h and beyond. The ARW has smaller bias between 400 and 200 hPa (results are SS for 200 hPa at 24 h, 400 and 200 hPa at 48 h, and 400 and 300 hPa at 60 h), while the NMM has smaller bias at 850, 700, and 100 hPa at all lead times after initialization. The intercore differences reach as much as 0.5 ^{o}C at 100 hPa.

The evolution of temperature bias with forecast lead time for selected levels is shown in Fig. 9. As discussed before, the sign of the bias varies with level. However, for a given level, the sign of the bias remains virtually unaltered through the length of the forecast. At the surface, as shown in Fig. 9a, there is a warm bias with a prominent diurnal cycle. The bias increases during the night to reach a maximum in the early morning (15 UTC or 9 AM CST) and decreases during the day, to reach a minimum in the mid afternoon (21 UTC or 3 PM CST). At the surface and at 500 hPa, the magnitude of the bias does not increase with lead time, but an increase is noted at 850, 700, and especially at 100 hPa (Figs. 2b, 2c, and 2f). At 250 hPa, the median bias peaks at the 12 h lead time, and decreases thereafter. It is interesting to note that at the surface both cores have positive bias, while at the lowest upper air level (850 hPa), they both have negative biases, especially at the later forecast hours.

The time series of 2-m AGL temperature bias shows a SS ARW advantage at all lead times. Statistically significant temperature bias intercore differences are absent at 500 and 250 hPa. However, they are present at the levels for which the bias increases with time, because typically the bias increases faster for one of the cores. At 850 and 700 hPa, the bias for both cores becomes progressively more negative with time. This progression is more accentuated for the ARW, leading to increasing intercore differences which become SS after initialization, with the NMM displaying smaller magnitude of bias. The median difference reaches 0.3 ^{o}C for the 60-h lead time. At 100 hPa, a level at which both cores have a warm bias, the intercore difference increases with time, with the median reaching a maximum of 0.6 ^{o}C at 60 h.

Both cores have BCRMSEs that decrease with height from 850 to 300 hPa (Fig. 10). Above this level, the errors increase up to 200 hPa. At the 60-h lead time, the BCRMSE reaches a maximum of

approximately 2.2 ^{o}C at 850 and 200 hPa. Intercore differences in BCRMSE temperature do not exceed 0.05^{o} C and cannot be considered statistically significant for any level.

The temperature BCRMSE evolution with lead time (Fig. 11) indicates that the error grows in time for both dynamic cores at all levels. The rate of growth is the largest at 850, 700, 500 and 250 hPa (about 0.8 ^{o}C in 60 h), and smallest at the surface and 100 hPa. The surface BCRMSE plot (Fig. 11a) indicates that, superposed on the overall slight increase in errors over time, a semidiurnal modulation is present, with errors increasing in the early morning and early afternoon. Similar results were obtained for the average of the 12 UTC cycles (not shown). The intercore temperature BCRMSE is only statistically significant at the 30-h lead time, when the ARW has smaller error. No growth of the differences is observed towards the later forecast periods at any level.

c) Relative humidity

The RH vertical profiles displayed in Fig. 12 indicate that the initial RH bias is near zero at 850 hPa, but is positive at upper levels for the ARW. At later times, the RH bias is positive at all lead times and levels. The largest median biases after initialization are found at 500 hPa with approximately 4%. The SS intercore differences point to smaller NMM absolute bias at two levels after initialization: 700 hPa at all times and 500 hPa at 24 h. The ARW does not produce forecasts with smaller absolute RH bias than the NMM at any level or lead time.

The 2-m AGL RH bias time series (Fig. 13a) reveals mostly negative biases alternating with short periods of positive or near-zero biases. A diurnal cycle is noticeable, with the biases displaying the largest negative values in the night and early morning and the near-zero or positive values in the mid afternoon. At upper levels (Figs. 13b-d), the RH displays positive bias at all times after initialization. The median bias increases in the first 24 h of simulation, and remains stable or decreases thereafter. The largest upper level median bias is approximately 4%.

It should be noted that RH and temperature errors are not independent. The temperature bias is positive at 2-m AGL and negative at 850 and 700 hPa. If the mixing ratio forecast were perfect, this would lead to RH biases with negative values at 2-m AGL and positive at 850 and 700 hPa, as obtained here.

The 2-m AGL intercore differences show RH bias SS differences at several lead times. At 03, 06, 24 and 48 h the ARW produces SS moister forecasts, while the NMM is SS moister at 33, 36, 39, 54, 57,

and 60 h. Given the sign of the RH bias, this results in ARW forecasts with smaller absolute bias at 3 and 6 h, and NMM forecasts with smaller absolute bias at the later times. While no SS intercore differences are present at 850 hPa, at higher levels, all SS intercore differences point to NMM forecasts with smaller absolute bias. They are present at all forecast hours at 700 hPa, and at 12 and 24 h at 500 hPa. The largest upper air intercore difference occurs at 700 hPa at the 48 h lead time, with a median of 1.7 %.

The RH BCRMSE vertical profile (Fig. 14) shows that there is minimal variation of the error with height at the initial time. At later times, the median error is smaller at 850 hPa and increases with height, reaching 27 % at 500 hPa at 60 h. The intercore differences are not SS at all levels and lead times.

The times series of median RH BCRMSE at 2-m AGL (Fig. 15) shows a diurnal cycle revealing larger errors at the late afternoon, reaching 15% in the second day. At upper levels the RH BCRMSE grows in time, especially in the first 12 h of the forecast. On the other hand, the intercore differences in RH BCRMSE do not grow with time and are not statistically significant.

d) Winds

The vertical profile of wind speed bias is shown in Fig. 16 for several lead times. The results indicate an initialization with wind speeds that are too low at all levels below 300 hPa and too high at 150 hPa. Later in the forecast, the 850 hPa and 150 biases are reduced to near zero, while the 250 hPa negative bias increases. The largest absolute median biases are noted at 250 hPa at the 60-h lead time, with approximately -1.2 ms^{-1}.

SS intercore differences are not present at initialization but can be seen at later forecast lead times at multiple levels. At the 24-h lead time, SS intercore differences are seen at 850, 700, 150, and 100 hPa, with the ARW being faster in the lower troposphere and the NMM faster in the upper troposphere. At the 48-h and 60-h lead times, SS intercore bias differences occur at the same levels and with the same sign as for the 24-h forecast, with the addition of SS results at 500 hPa, with ARW producing faster winds. Among those SS results, only four can be used to make statements about model performance, and in all four cases the ARW produces smaller absolute bias: 700 hPa at 24-, 48-, and 60-h, and 500 hPa at 48 h.

The time series of wind speed bias at several levels is presented on Fig. 17. At 10-m AGL, both cores have positive bias at all times, with a diurnal oscillation: the bias is exacerbated in the early morning

and minimum in the late afternoon. The maximum positive wind bias values are found at 10-m AGL, with the NMM reaching 0.8 ms^{-1}, and the ARW 1.3 ms^{-1}. At 850 hPa, both cores start off with negative bias but have near-zero bias at later times. At 700 hPa, the NMM has negative bias at all times, but the ARW results differ from zero significantly only at 00 and 12-h lead times. Therefore, both at 850 and 700 hPa the bias improves after initialization, and remains approximately stable after 24 h. At 500 hPa the bias is negative at all times, and remains approximately constant with lead time. At 250 hPa, the bias is near zero in the first 12 h, but progresses to negative values thereafter, reaching the maximum negative bias of all levels at 60 h with approximately -1.2 ms^{-1}. At 100 hPa, the bias is near zero at all times, except for the NMM at 36 h, for which positive bias is registered.

SS intercore differences are present at several levels and lead times. The largest intercore differences are noted at 10-m AGL, where the ARW produces higher positive bias than the NMM, leading to smaller NMM absolute bias at all lead times, with the median reaching a maximum of 0.7 ms^{-1} at the 45-h lead time. The ARW generates faster winds than the NMM at all lead times beyond initialization at the 850 and 700 hPa, while the NMM is faster for the 12-h forecast at 250 hPa and at all lead times beyond initialization at 100 hPa. Based on the intercore differences and on the sign of the bias for both cores, the ARW produces SS smaller absolute bias at 850 hPa for the 12-h lead time, at 700 for all lead times, and at 100 hPa for the 24 and 48-h lead times, while the NMM does not produce superior forecasts at upper levels at any lead time.

The vertical profiles of wind BCRMSE (Fig. 18) indicate that at the initial time the median errors range from 3.0 at 850 hPa to 4.7 ms^{-1} at 150 hPa. At later times, the shape of the error profile changes and the median displays a maximum at 250 or 300 hPa, reaching 10.5 ms^{-1} at 60 h. The intercore differences are SS at 850 and 700 hPa for the 24- and 48-h lead times, indicating that the NMM's errors are approximately 0.3 ms^{-1} lower than the ARW's.

The time series of wind BCRMSE are presented in Fig. 19. At the surface, a very slow error growth is superimposed to a diurnal cycle displaying larger errors in mid afternoon (compatible with results from average of 12 UTC cycles – not shown). At all upper levels, the wind BCRMSE grows in time, with the largest growth (over 6.0 ms^{-1} in 60 h) occurring at 250 hPa, which is the level with the largest errors (reaching 10.5 ms^{-1} at 60 h).

All SS intercore differences for wind BCRMSE favor the NMM. They are present at all lead times beyond initialization at 10-m AGL, 850 and 700 hPa, and at 24-, 36- and 48-h lead times at 500 hPa. The largest intercore BCRMSE magnitude occurs at the surface with 0.4 ms^{-1}.

5. Discussion and final conclusions

The main conclusions of this Test are listed in the Executive Summary and will not be repeated here. Instead, a comparison between the results for the 24-h lead time of the RRCT and CT2007 will be presented. It is important to keep in mind that there are several differences in configuration between the two Core Tests:

- The SI was used in the RRCT and WPS in the CT2007;
- Different versions of WRF and WPP were employed;
- Different datasets were used for initialization (RUC for the RRCT and NAM model for the CT2007);
- The RRCT was run for two physics packages (Phase 1 and Phase 2 physics), while the CT2007 was run using a single physics suite, more similar to the Phase 1 of the RRCT;
- Different cycles were used for each season.

Moreover, the methodology used to present the RRCT results in Nance (2006) differs somewhat from the one used in this report, so the comparison must be done with caution. The temperature, RH, and winds results for the RRCT were summarized using the mean, while the median is used in Section 4 of this report. Additionally, the RRCT employed RMSE, while BCRMSE was used in the CT2007 report. Finally, 100-hPa results were only examined in the CT2007.

For this comparison, plots of the CT2007 annual mean RMSE over the CONUS domain at the 24-h lead time were computed. The CT2007 bias medians (Figs. 8, 12, and 16) will be used to compare against the CT2007 bias means since, for the annual aggregation over the whole CONUS, the means and medians are fairly similar (not shown).

The main findings for the RRCT, quoted from the Executive Summary in Nance (2006), are listed below in italic type, with a comment on whether each result is valid in the CT2007. For a description of the criteria for “concern” and “serious concern”, see Nance (2006). It should be noted that the RRCT main findings were those upheld for both physics packages and most seasons and domains, while the CT2007 results are mostly based on the CONUS results for the entire test period.

1. *The ARW wind vector RMSE is less than the NMM wind vector RMSE in the 300-150 hPa layer. The magnitude of these differences also meets the criteria for concern.*
   - The CT2007 results are substantially different. Statistically Significant wind RMSE mean differences only occur below 400 hPa and without exception favor the NMM (Fig. 20). These differences reach the level of concern at 850 and 700 hPa.
2. *Differences between the wind speed bias for the two dynamical cores indicate the upper level wind in the ARW forecasts are generally weaker than those in the NMM forecasts, whereas the lower level winds in the ARW forecasts tend to be stronger than those in the NMM forecasts.*
   - Confirmed in the CT2007 (Fig. 16).
3. *The ARW temperature RMSE is less than the NMM temperature RMSE in the 400-200 hPa layer. Although the differences in the layer show consistency across physics packages and observational data type, the magnitude of these differences decreases during the second 12 hours of the forecast. The magnitude of the differences in this layer generally does not exceed the threshold for concern.*
   - The CT2007 results differ substantially in that the temperature RMSE mean differences at the 400-200 hPa layer are not SS (Fig. 21). SS temperature RMSE are present and favor the NMM at levels 700 and 100 hPa, the latter reaching the level of serious concern.
4. *Differences between the temperature biases for the two dynamical cores indicate the ARW forecasts are generally colder than the NMM forecasts. Both cores exhibit a negative temperature bias at lower levels that transitions to a positive bias at upper levels. This vertical structure combined with the colder temperatures in the ARW forecast leads to the magnitude of the temperature bias being smaller for the NMM at lower levels and smaller for the ARW at upper levels.*
   - Confirmed in the CT2007 (Fig. 8).
5. *The relative humidity RMSE differences did not exhibit any SS signature that was consistent for both physics packages. On the other hand, a number of the differences for phase 2 and one difference for phase 1 that were found to be SS did exceed the thresholds for concern and serious concern. The ARW RH RMSE was smaller than that of the NMM for all of these cases.*
   - In the CT2007, the RH RMSE intercore differences are not SS (Fig. 22).
6. *Differences between the RH biases for the two dynamical cores indicate the ARW forecasts are generally associated with higher values of RH than the NMM forecasts. This tendency for higher RH in the ARW forecasts can at least partially be explained by the tendency for the temperatures to be colder in the ARW forecasts.*

- Confirmed in CT2007 (Fig. 12).

7. *Very few of the differences between the QPF verification measures for the two dynamical cores are SS and show consistency for the two physics packages. Only the bias differences at the lowest thresholds produce consistent SS results. The bias differences at the lowest thresholds indicate the NMM produces less overestimation of the areal coverage than the ARW. On the other hand, all of these differences do not exceed the threshold for concern.*
    - Mostly confirmed in CT2007, bias and ETS intercore differences are not SS (Figs. 4 and 5).

This list underlines substantial differences between the results for the two Core Tests, which can be attributed to the disparity in version of the code and initialization procedures between the two Core Tests.

As mentioned before, this report presented just a subset of the CT2007 results. Future work involves incorporating verification against aircraft observations, exploring the full seasonal and geographical decomposition of the results for all variables, quantifying the differences between results obtained in the LINUX and IBM platforms, and presenting the temperature, RH, and wind results using box plots, to go beyond the median in characterizing the distribution.

References

Bernardet, L., L. Nance, M. Demirtas, S. Koch, E. Szoke, T. Fowler, A. Loughe, J. L. Mahoney, H-Y Chuang, M. Pyle, and R. Gall, 2008: The Developmental Testbed Center and its Winter Forecasting Experiment. *Bull. Amer. Meteor*. Soc., **89**, 611-627.

Brown, J. M., S. Benjamin, T. G. Smirnova, G. A. Grell, L. R. Bernardet, L. B. Nance, R. S. Collander, and C. W. Harrop, 2007: Rapid-Refresh Core Test: aspects of WRF-NMM and WRF-ARW forecast performance relevant to the Rapid-Refresh application. *22st Conference on Wea. Anal. Forec*., Park City, UT, Amer. Meteor.Soc.

Chuang, H-Y, G. DiMego, M. Baldwin, and WRF DTC team, 2004: NCEP's WRF Post-processor and verification systems. *5th WRF / 14th MM5 Users' Workshop*, Boulder, CO. [Available online at http://www.mmm.ucar.edu/mm5/workshop/ws04/Session7/Chuang.Hui-Ya.pdf].

Janjic, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteorol. Atmos. Phys.* **82**, 271–285.

_____, 2004: The NCEP WRF core. *20th Conf. on Num. Wea. Pred*., Seattle, WA. [Available online at http://ams.confex.com/ams/pdfpapers/70036.pdf].

Kain, J., S. Weiss, G. Carbin, M. Baldwin, D. Bright, J. Hart, and J. Levit, 2005. Comparison of different WRF configurations in a severe weather forecasting environment. The 2005 SPC/NSSL Spring Program. 6th WRF / 15th MM5 Users' Workshop, 27-30 June, Boulder, CO.

Nance, L., 2006: Weather Research and Forecasting Core Test: WRF DTC Report. [Available from Developmental Testbed Center, NCAR, P.O. Box 3000, Boulder, CO, 80307-3000 or online at http://ruc.fsl.noaa.gov/coretest2/WRF DTC_report.pdf.]

Seaman, N., R. Gall, L. Nance, S. Koch, L. Bernardet, G. DiMego, Z. Janjic, J. Powers, W. Skamarock, and F. Olson. 2004: The WRF process: Streamlining the transition of new science from research into operations. *5th WRF User's Workshop, Boulder, CO.* [Available online at http://www.mmm.ucar.edu/mm5/ workshop/ws04/Session7/Seaman.Nelson.pdf].

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2, NCAR Tech Note, NCAR/TN–468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO, 80307]. Available on-line at: http://box.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf)

List of Figures

17. Annual median wind speed bias (ms^{-1}) as a function of forecast lead time for a) 2-m AGL (00 UTC cycles only), b) 850 hPa, c) 700 hPa, d) 500 hPa, e) 250 hPa, and f) 100 hPa. ARW is shown in blue, NMM in red, and their difference in green.
18. Same as Fig. 16 except for wind BCRMSE.
19. Same as Fig. 17 except for wind BCRMSE.
20. Vertical profile of annual mean wind RMSE (ms^{-1}) at the 24-h leadtime for a) ARW, NMM and ARW-NMM results, b) detail of ARW-NMM results, indicating the values of concern (between yellow and red lines) and serious concern (magnitude greater than red line). ARW is shown in blue, NMM in red and the ARW-NMM difference, in green.
21. Same as Fig. 20 except for temperature RMSE (^{o}C).
22. Same as Fig. 20 except for RH RMSE (%).

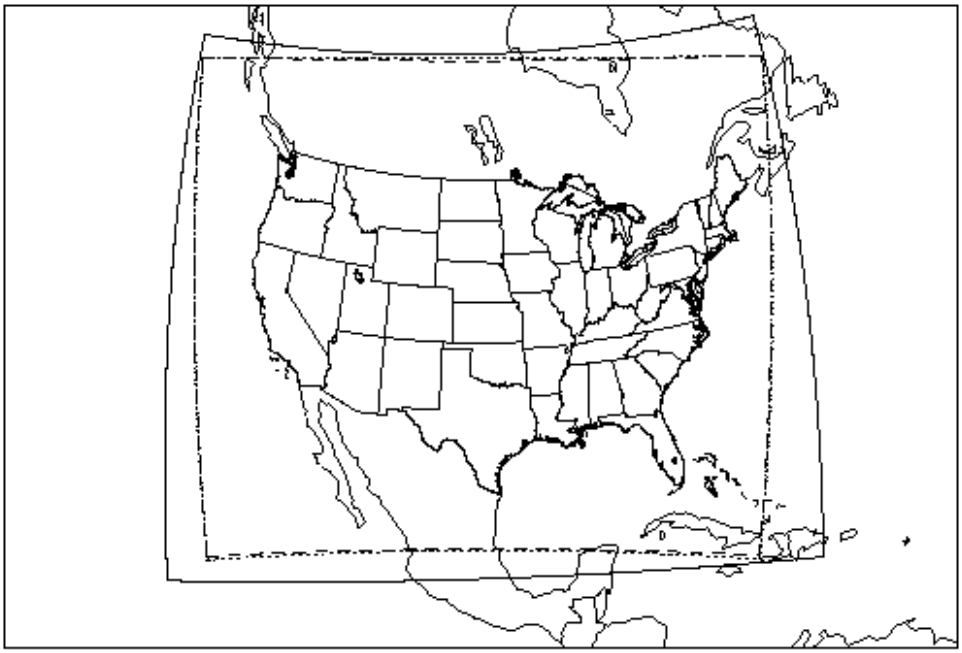

Figure 1. Map showing the boundaries of the computational domains used for the ARW (dashed line) and the NMM (dotted). The solid line shows the boundaries of the domain used for postprocessing.

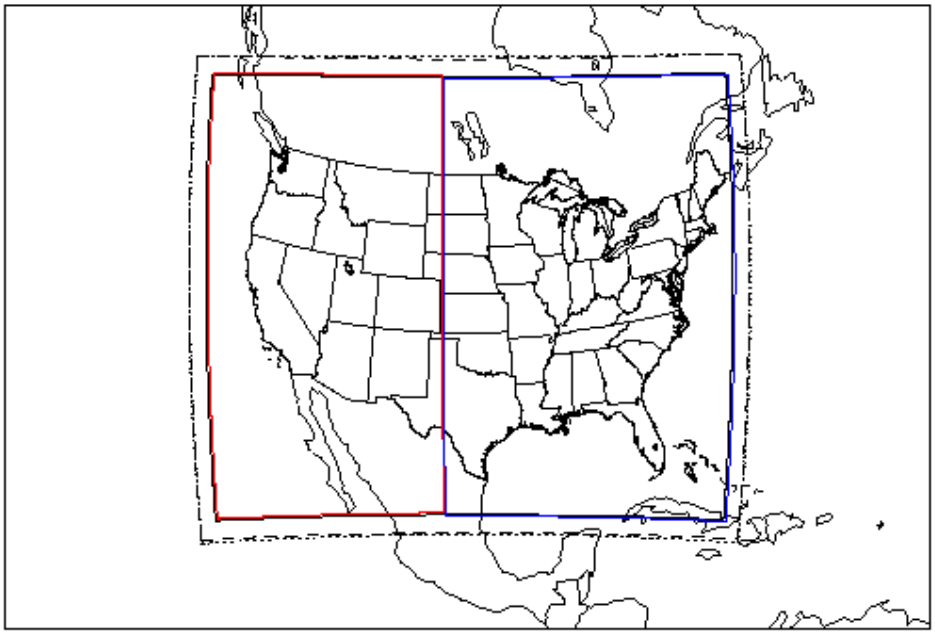

Figure 2. Map showing the boundaries of the verification domains: CONUS (solid black), West (solid red), and East (solid blue). The computational domains used for the ARW (dashed line) and the NMM (dotted) are shown for reference.

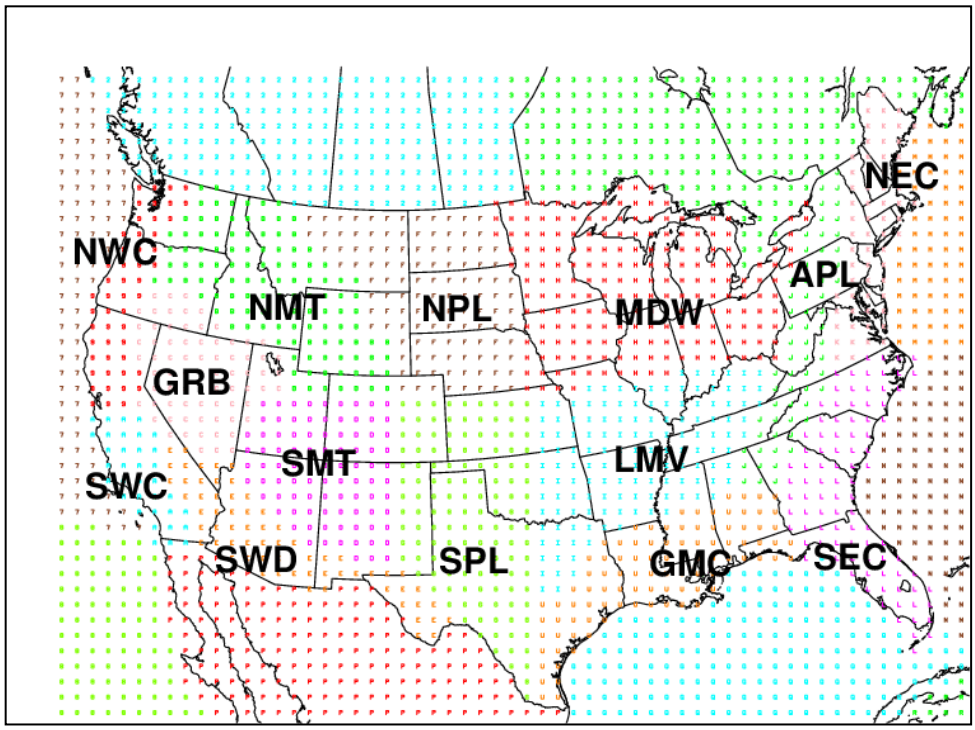

Figure 3. Map showing the locations of the 14 regional verification domains.

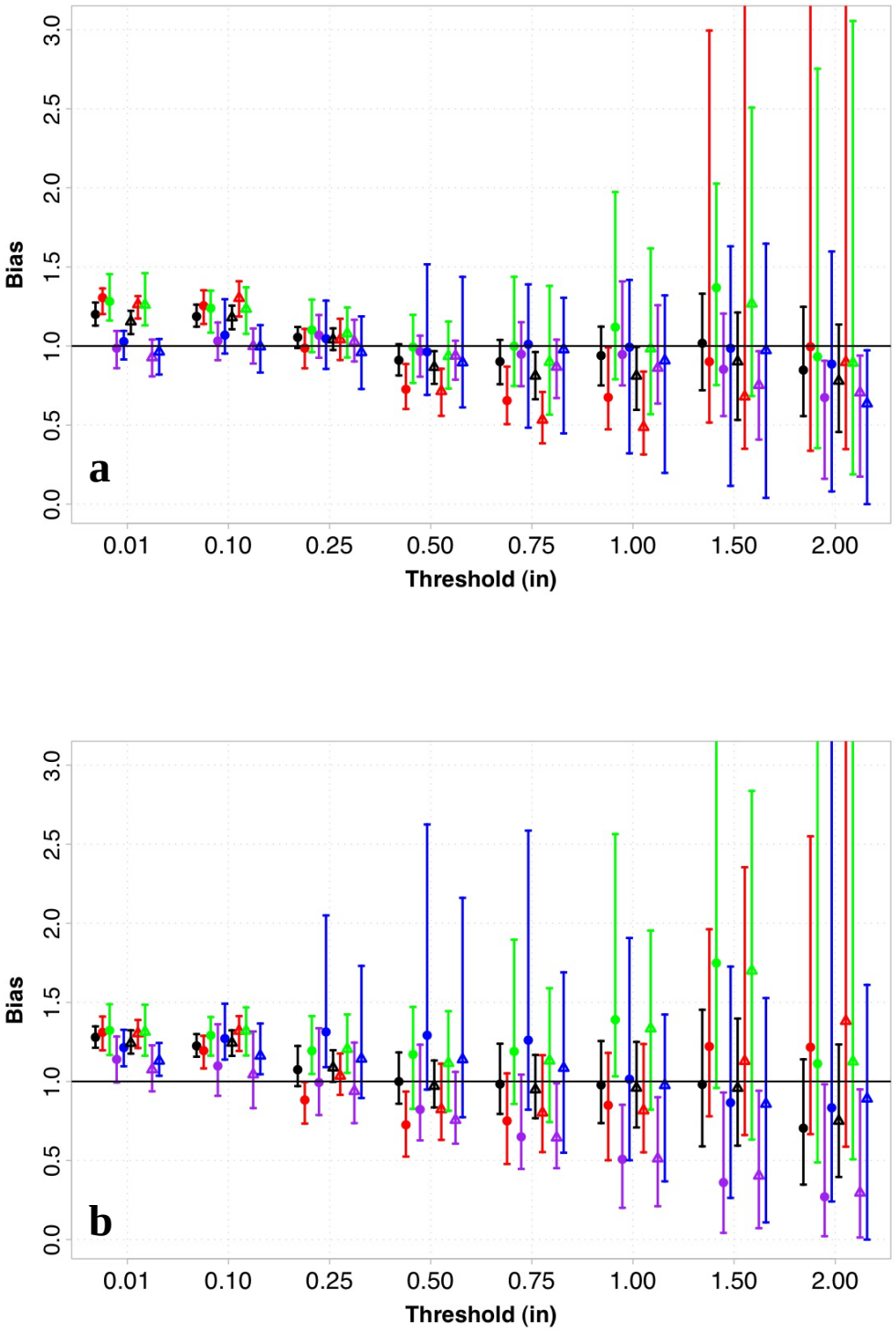

Figure 4. Bias for 24-h accumulated precipitation at the a) 24-h and b) 60-h lead times. ARW/NMM are circles/triangles. Annual mean in black, summer in red, spring in green, fall in purple and winter in blue. Vertical bars represent the 99% CIs.

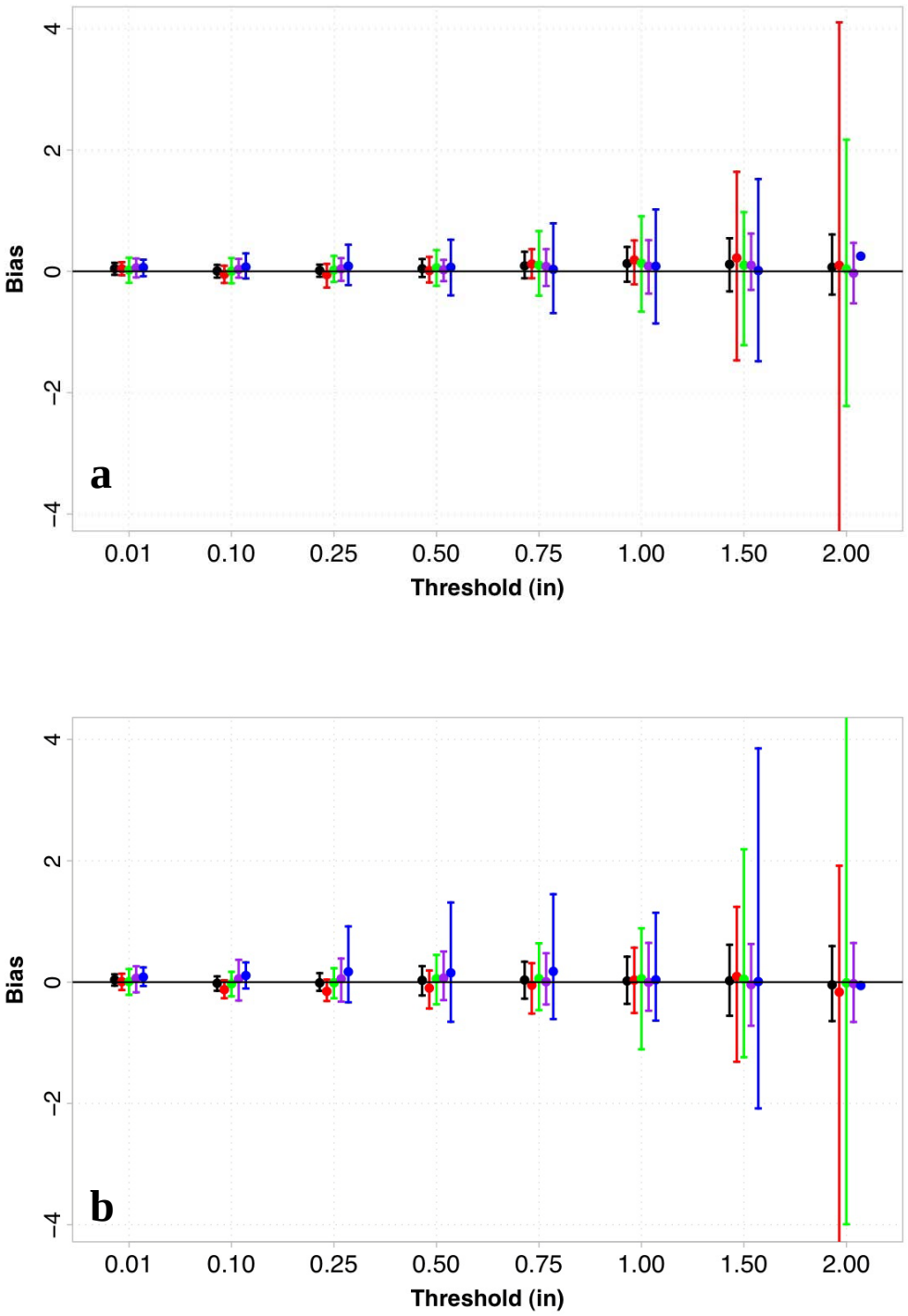

Figure 5. ARW-NMM difference in bias for 24-h accumulated precipitation at the a) 24-h and b) 60-h lead times. Annual mean in black, summer in red, spring in green, fall in purple and winter in blue.Vertical bars represent the 99% CIs.

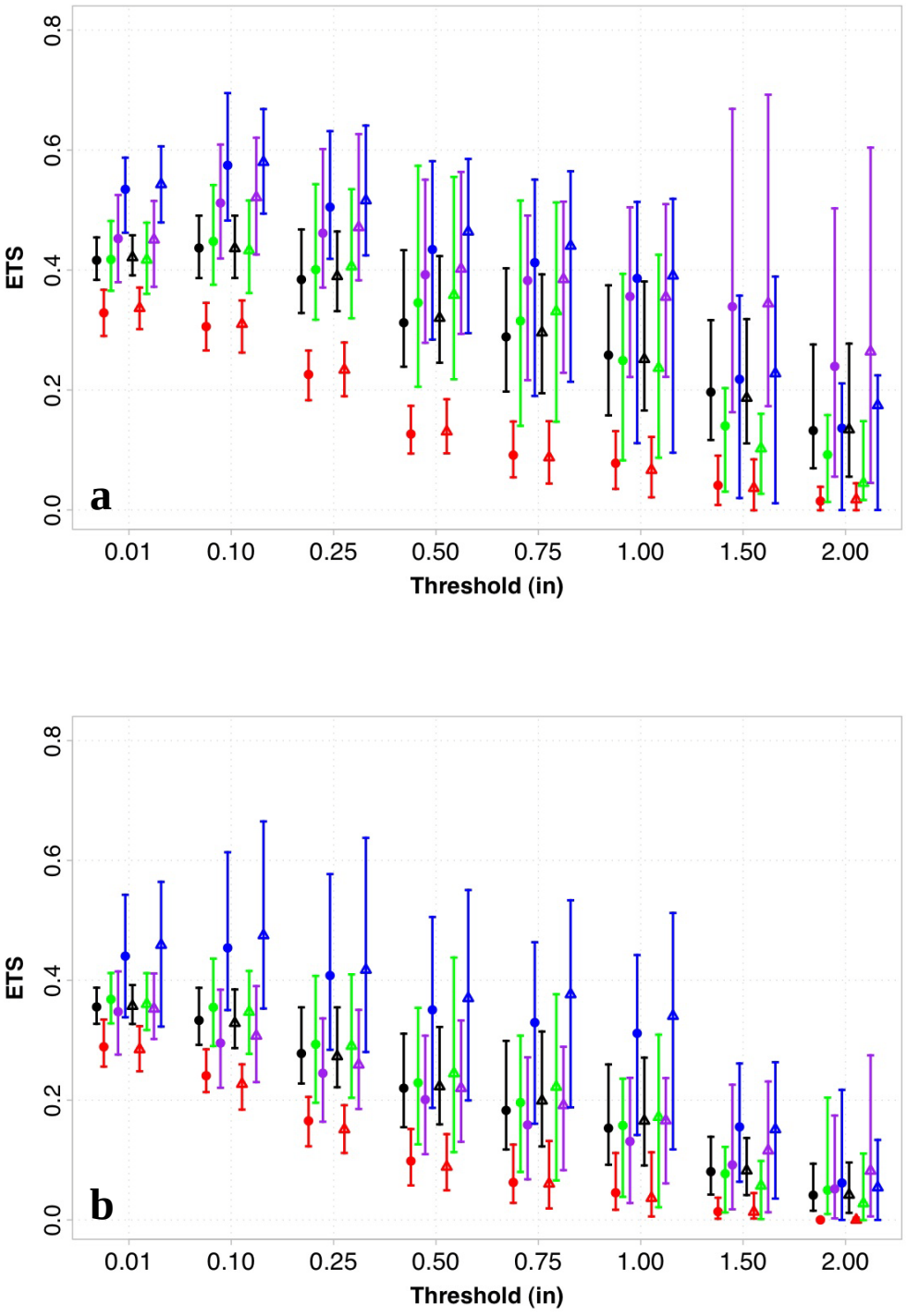

Figure 6. Same as Fig. 4, except for the ETS.

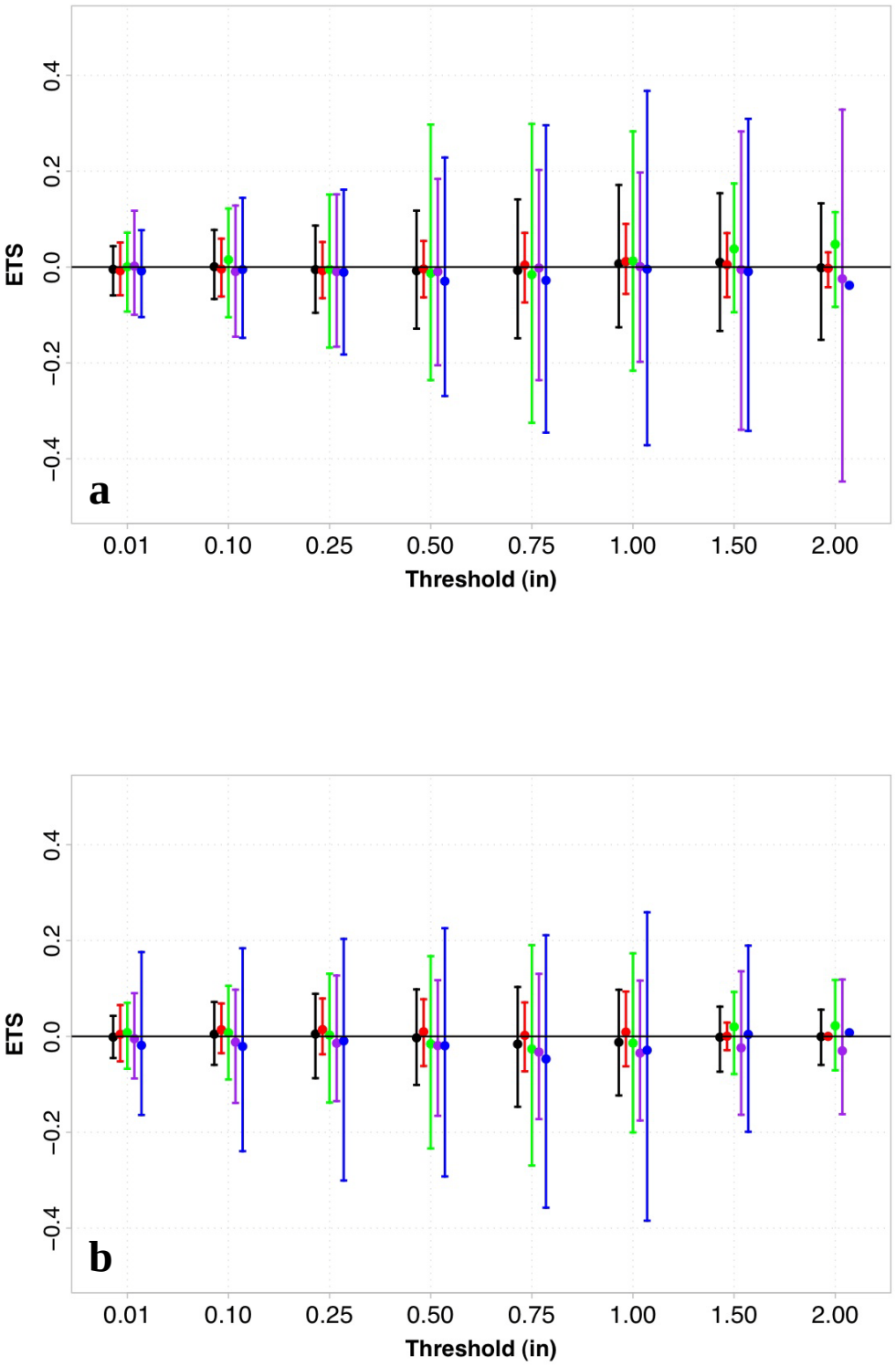

Figure 7. Same as Fig. 5, except for the ETS.

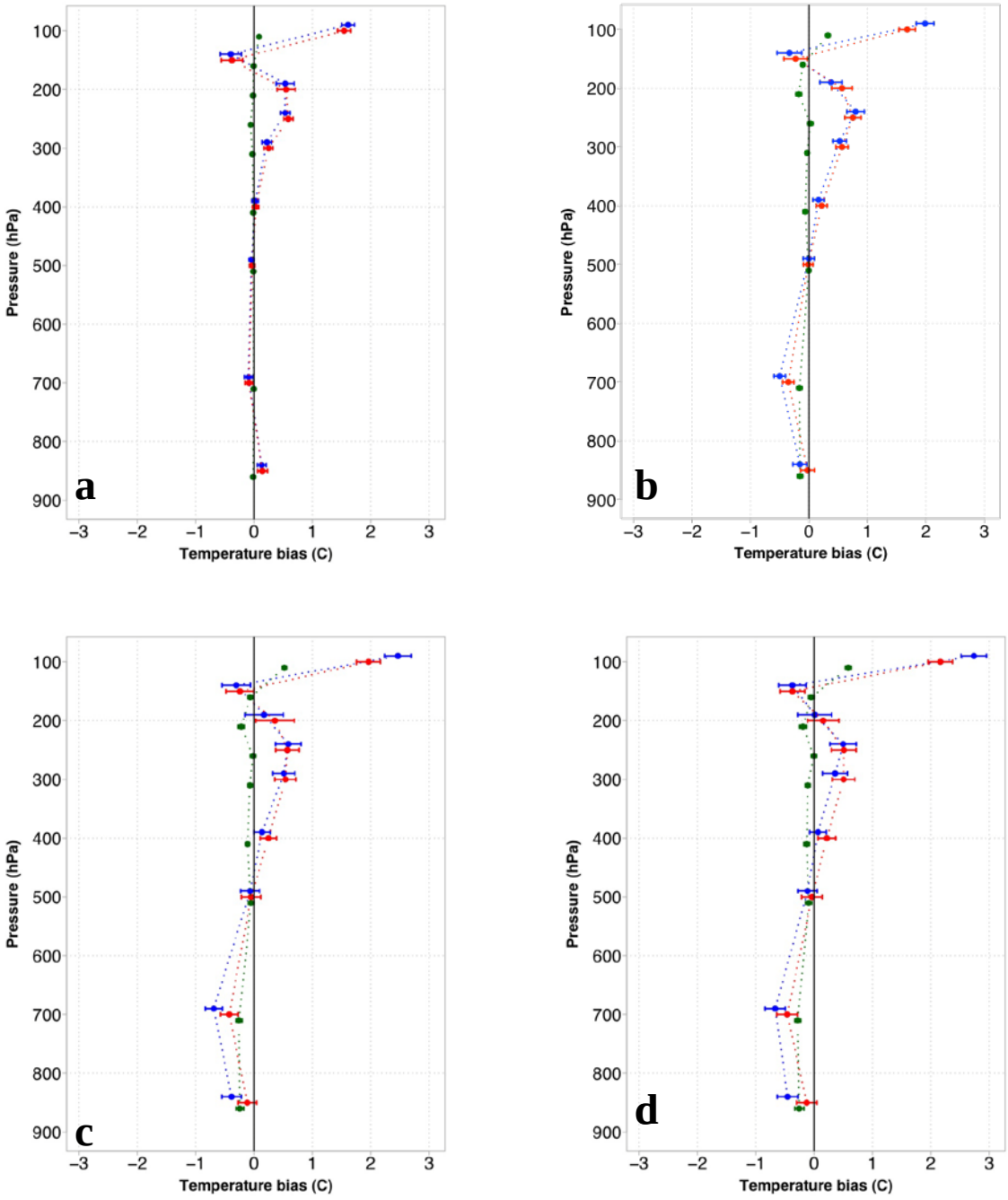

Figure 8. Vertical profile of annual median temperature bias (C) at the a) 00-h, b) 24-h, c) 48-h, and d) 60-h leadtimes. ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Horizontal bars represent the 99% CIs.

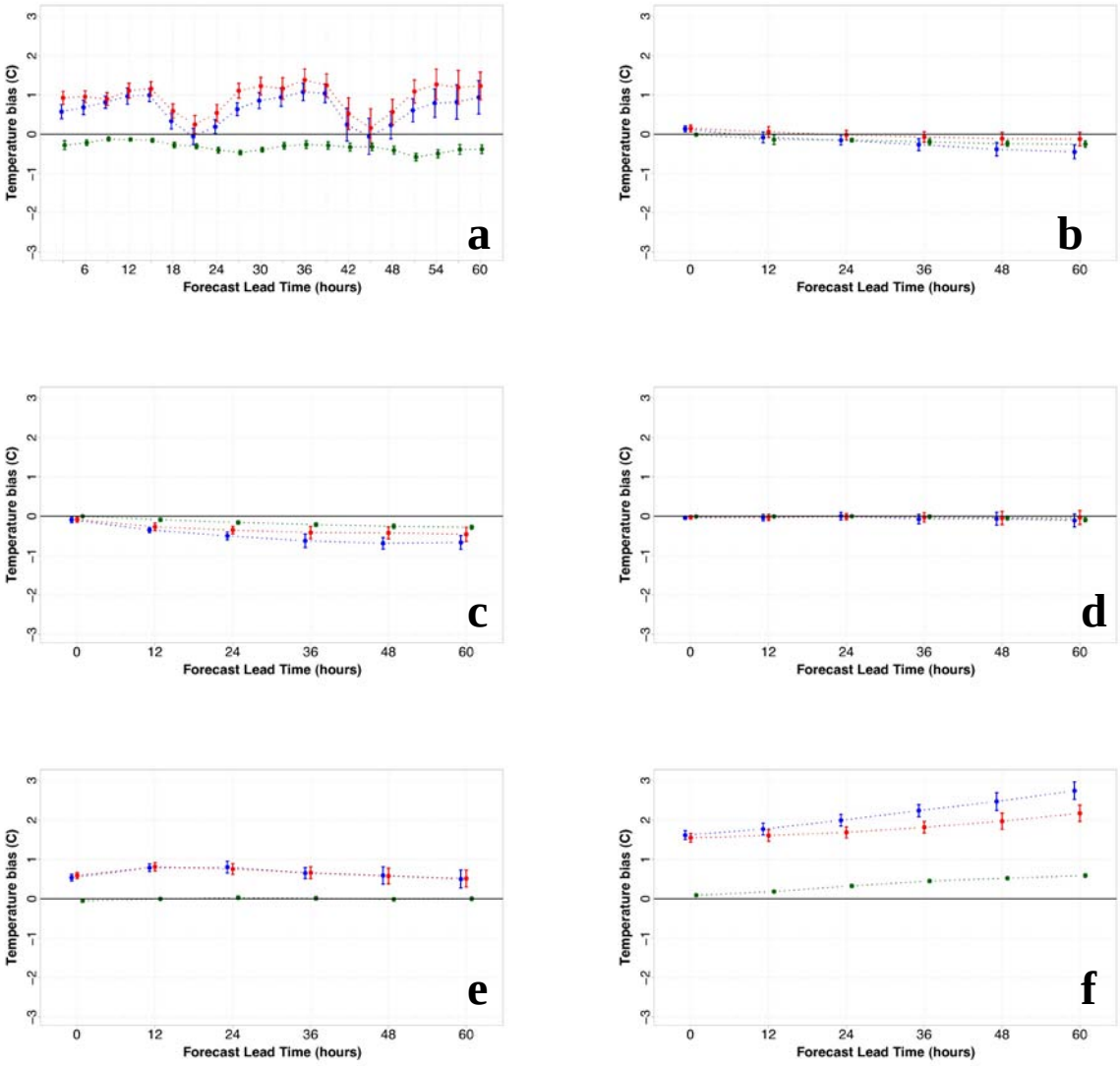

Figure 9. Annual median temperature bias (C) as a function of forecast leadtime for a) 2-m AGL (00 UTC cycles only), b) 850 hPa, c) 700 hPa, d) 500 hPa, e) 250 hPa, and f) 100 hPa. ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Vertical bars represent the 99% CIs.

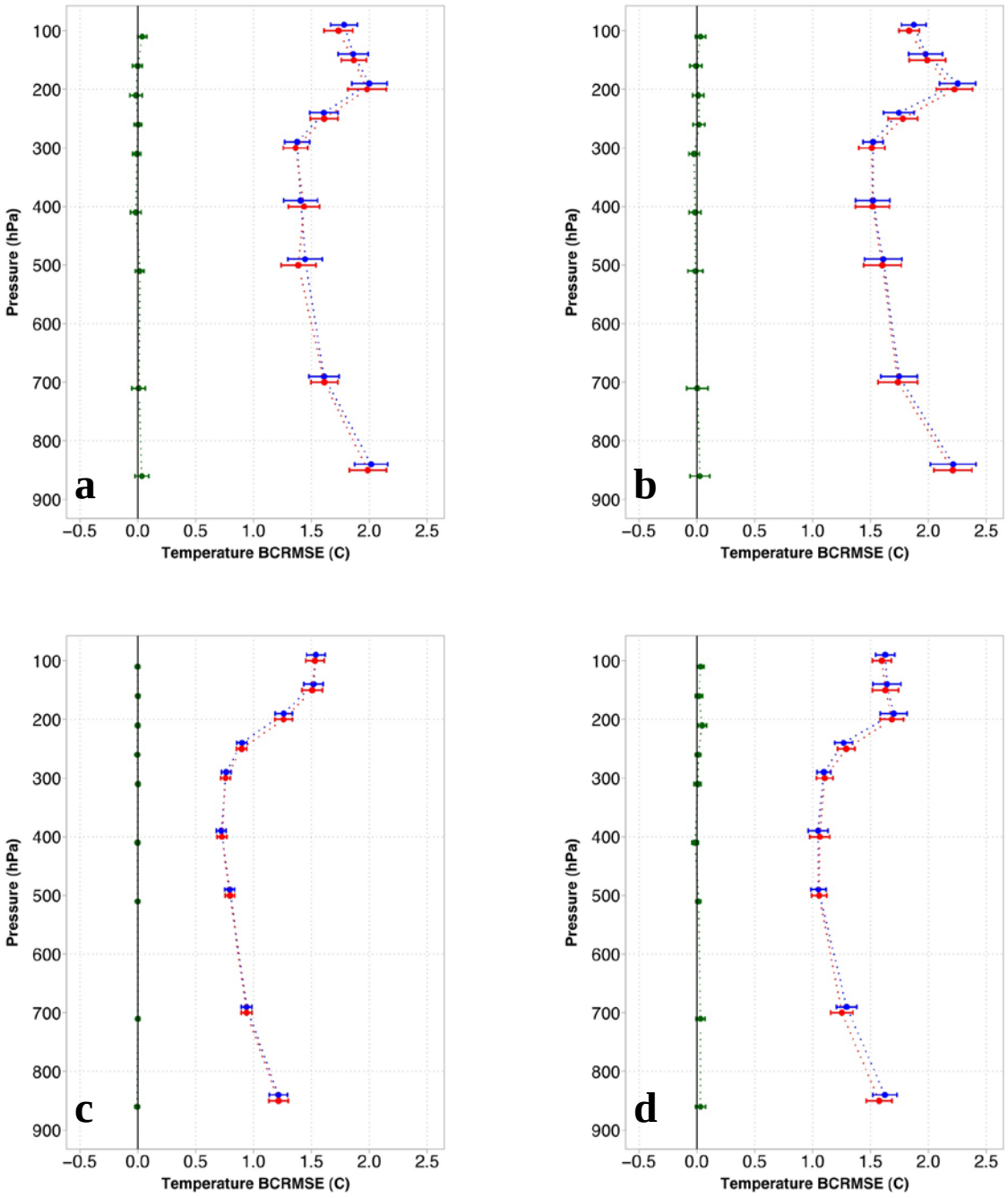

Figure 10. Same as Fig. 8 except for BCRMSE.

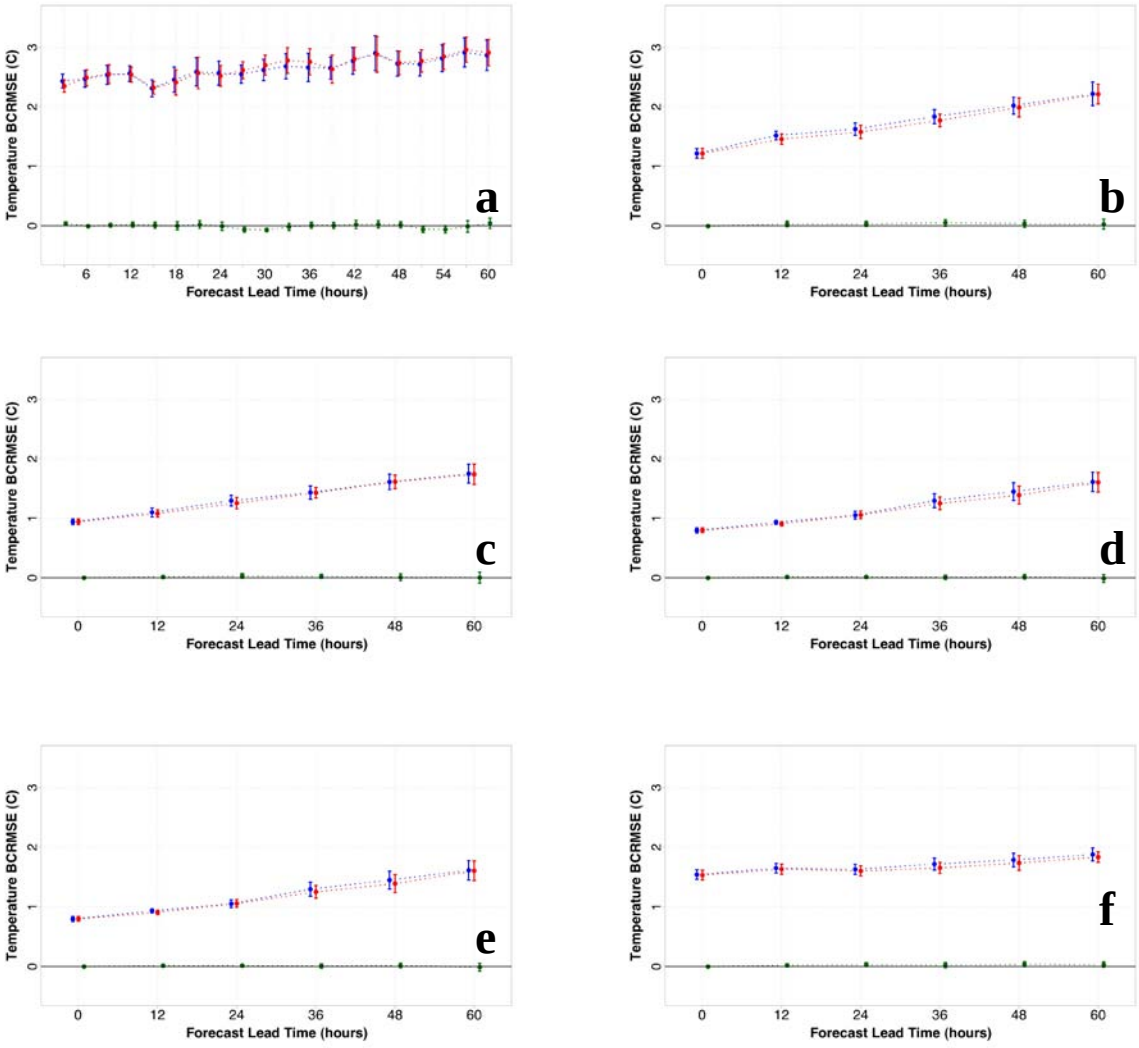

Figure 11. Same as Fig. 9 except for BCRMSE.

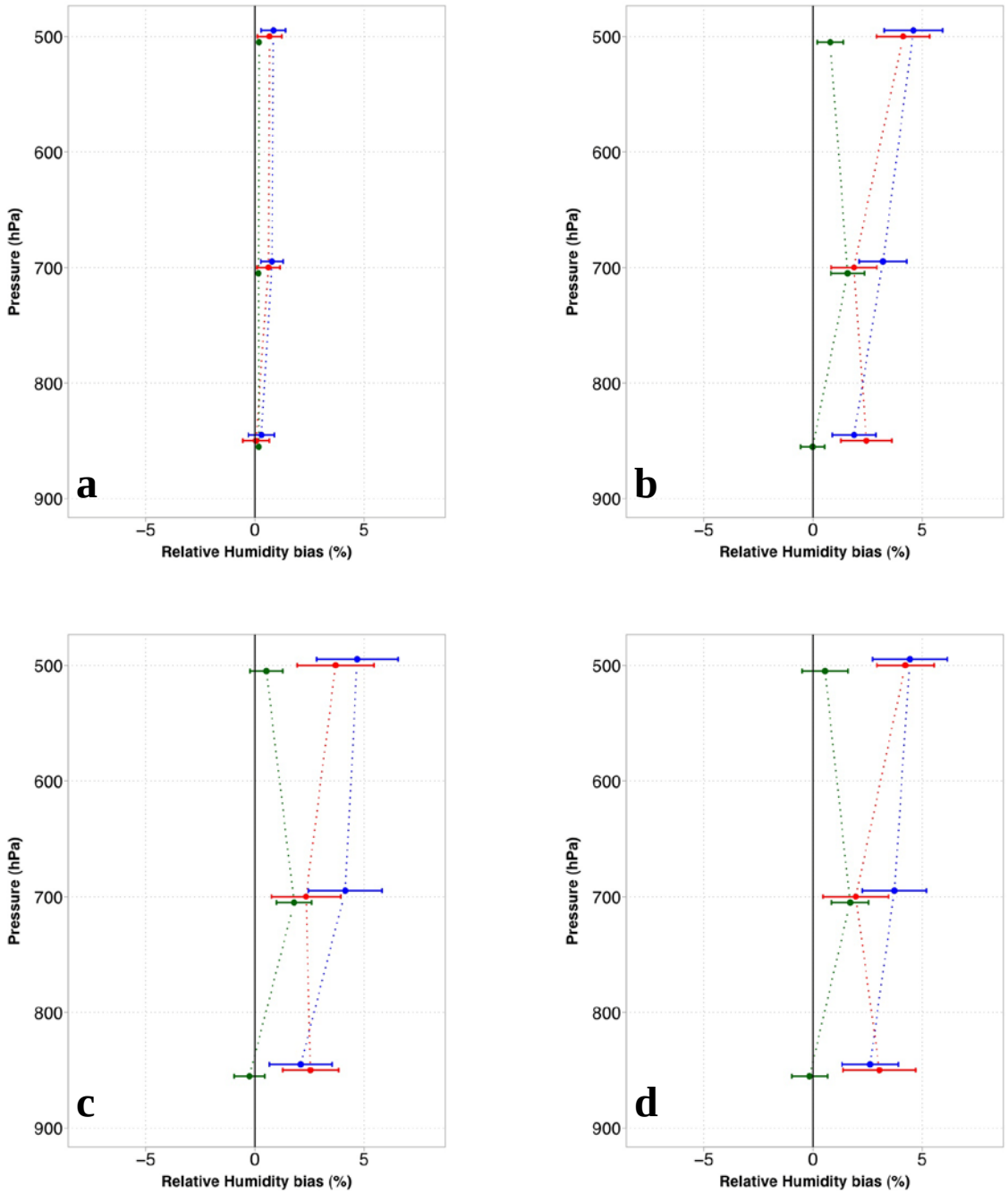

Figure 12. Vertical profile of annual median relative humidity bias (%) at the a) 00-h, b) 24-h, c) 48-h, and d) 60-h leadtimes. ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Horizontal bars represent the 99% CIs.

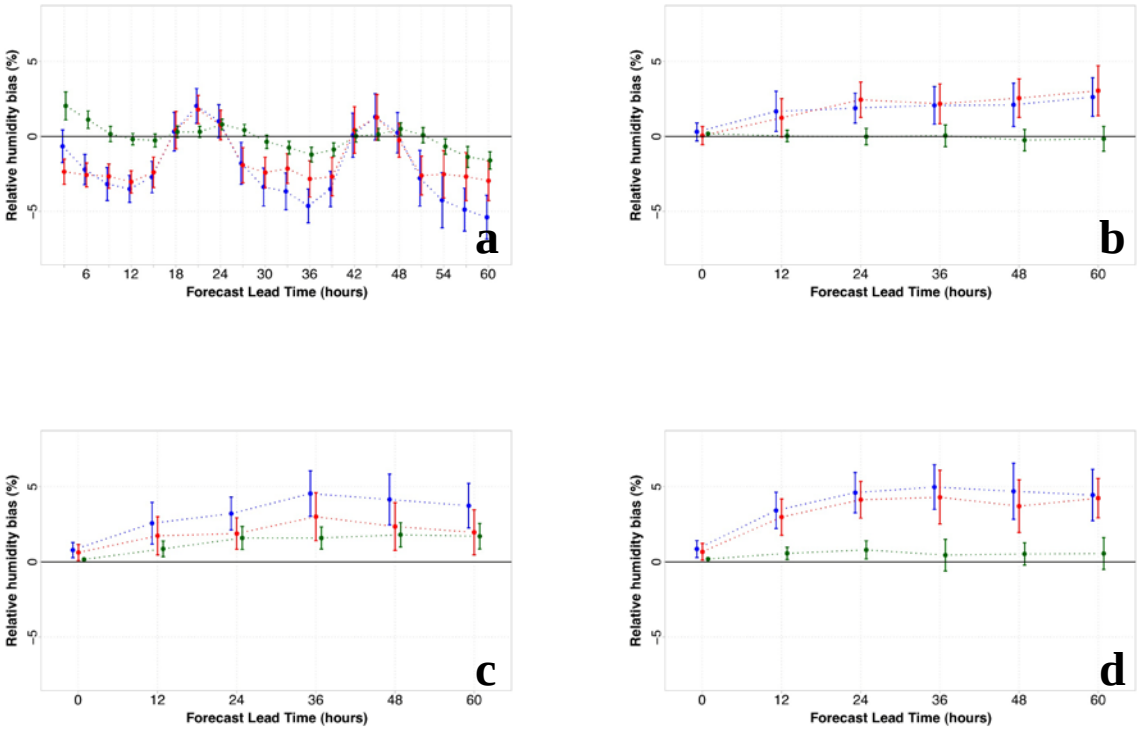

Figure 13. Annual median relative humidity bias (%) as a function of forecast leadtime for a) 2-m AGL (00 UTC cycles only), b) 850 hPa, c) 700 hPa, d) 500 hPa. ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Vertical bars represent the 99% CIs.

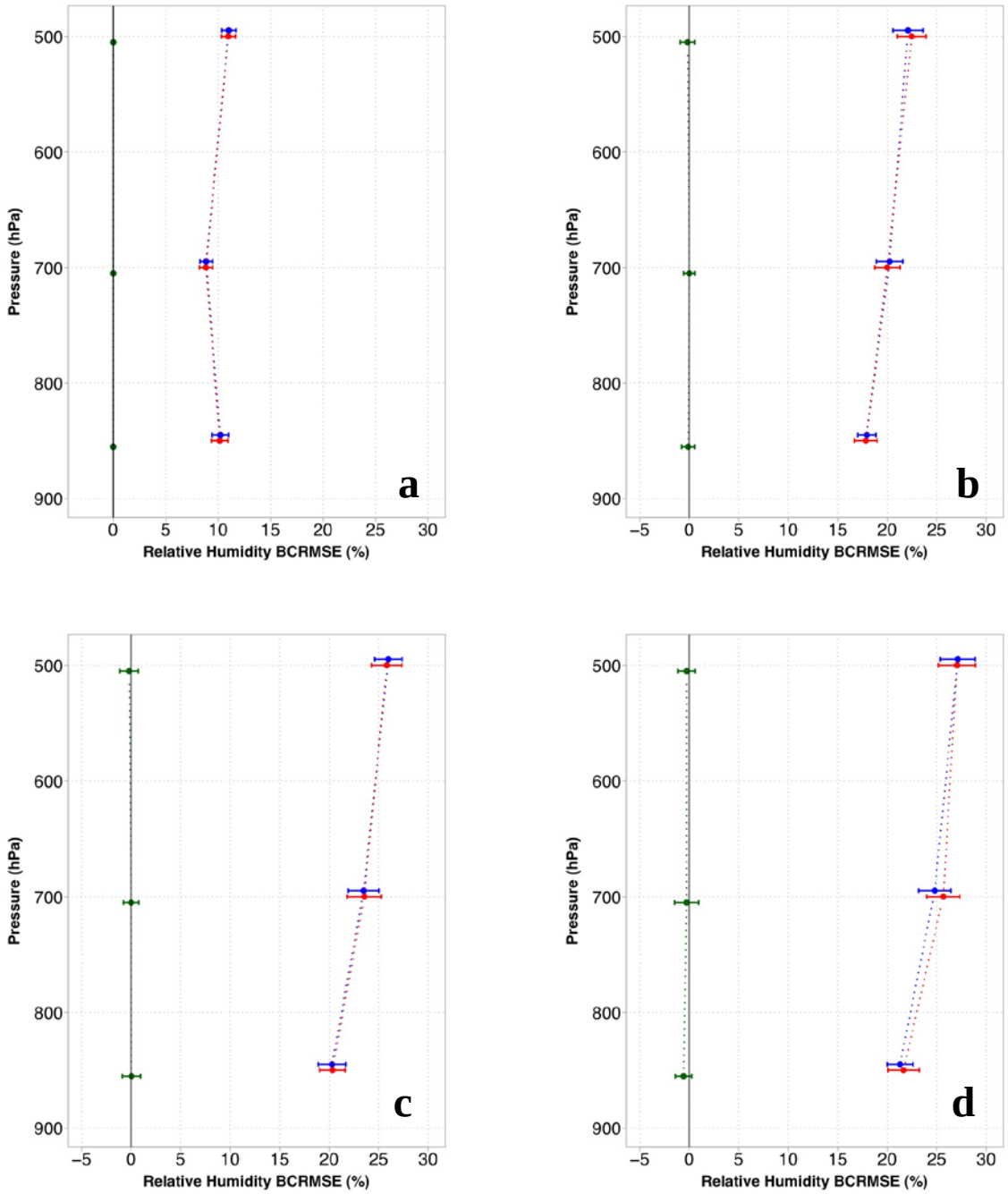

Figure 14. Same as Fig. 12 except for BCRMSE.

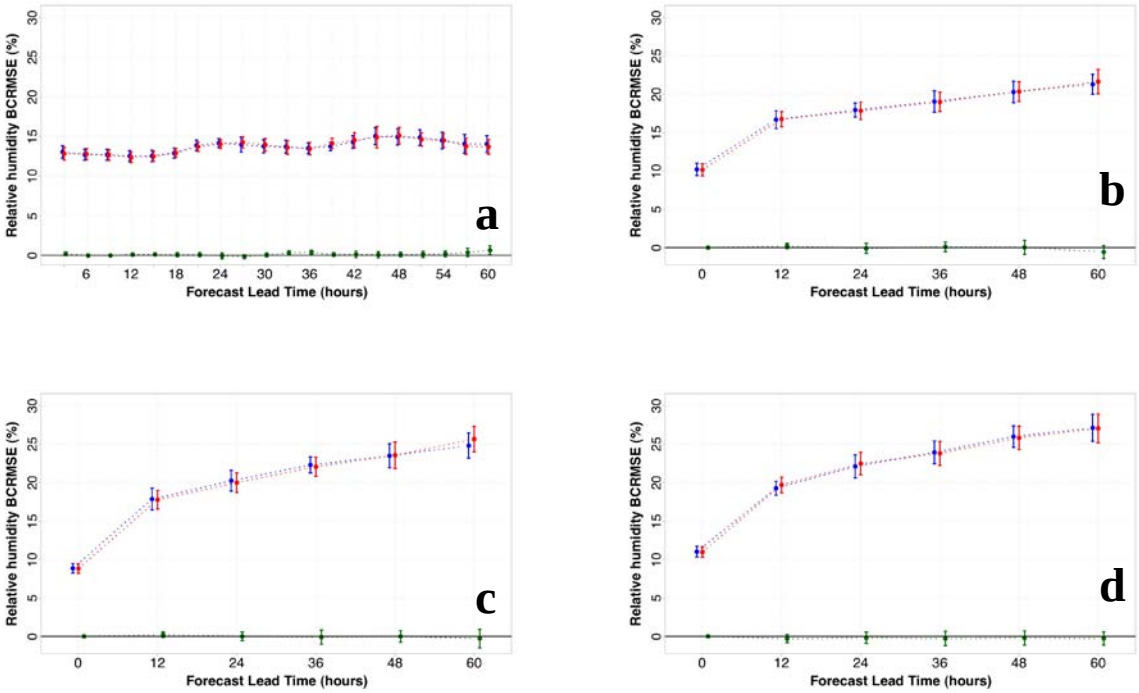

Figure 15. Same as Fig. 13 except for BCRMSE.

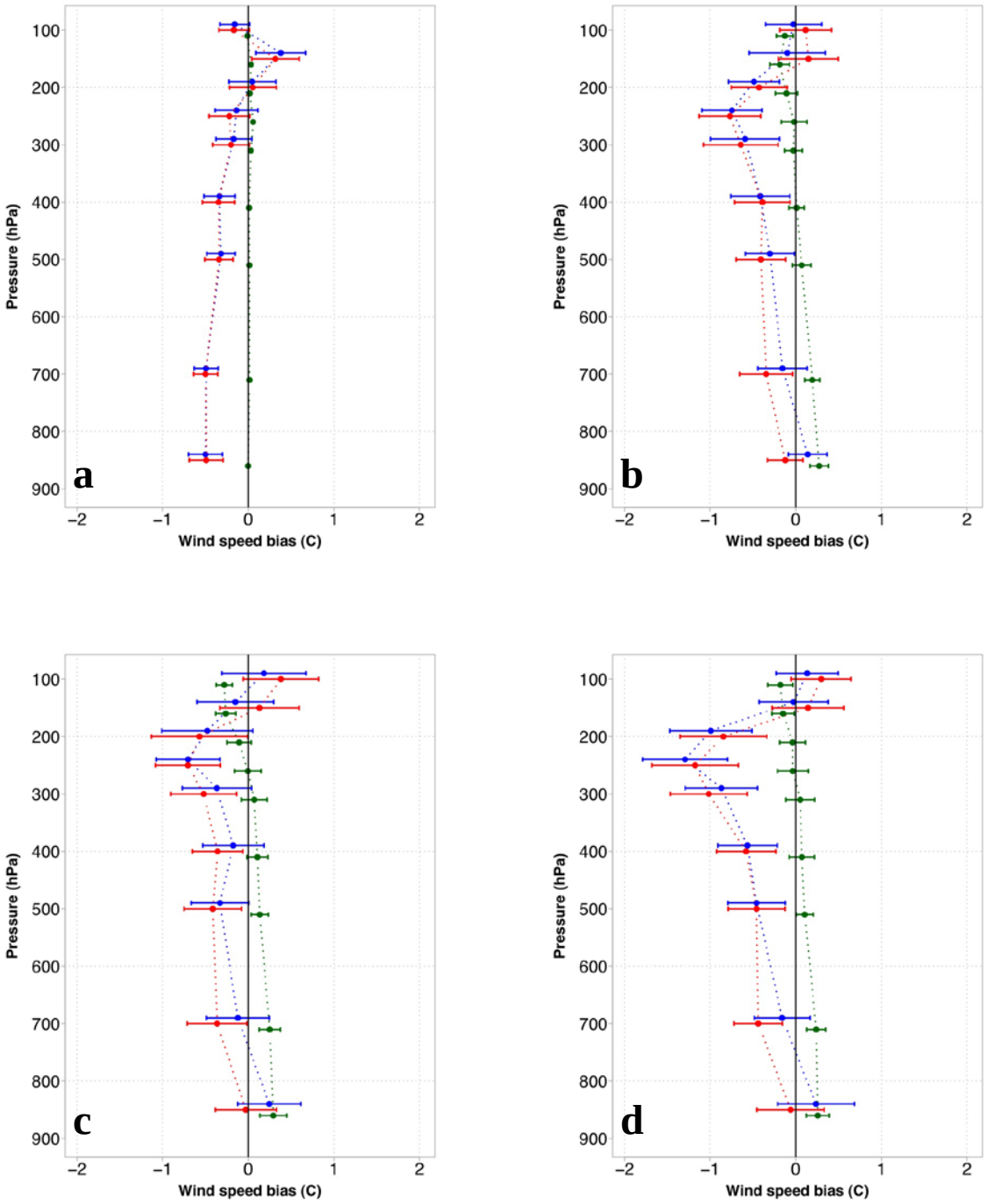

Figure 16. Vertical profile of annual median wind speed bias (m/s) at the a) 00-h, b) 24-h, c) 48-h, and d) 60-h leadtimes. ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Horizontal bars represent the 99% CIs.

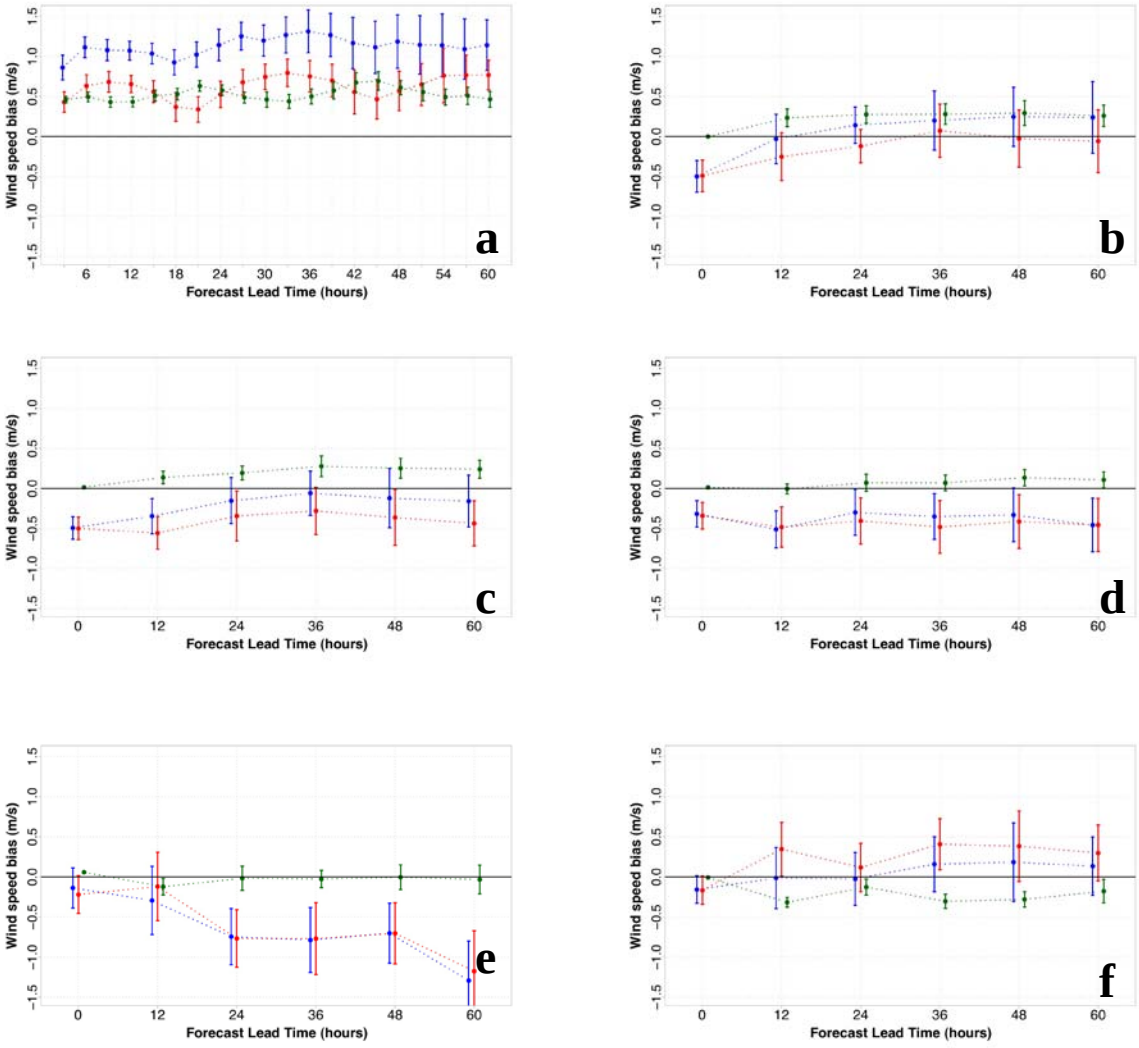

Figure 17. Annual median wind speed bias (ms^{-1}) as a function of forecast leadtime for a) 2-m AGL (00 UTC cycles only), b) 850 hPa, c) 700 hPa, d) 500 hPa, e) 250 hPa, and f) 100 hPa. ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Vertical bars represent the 99% CIs.

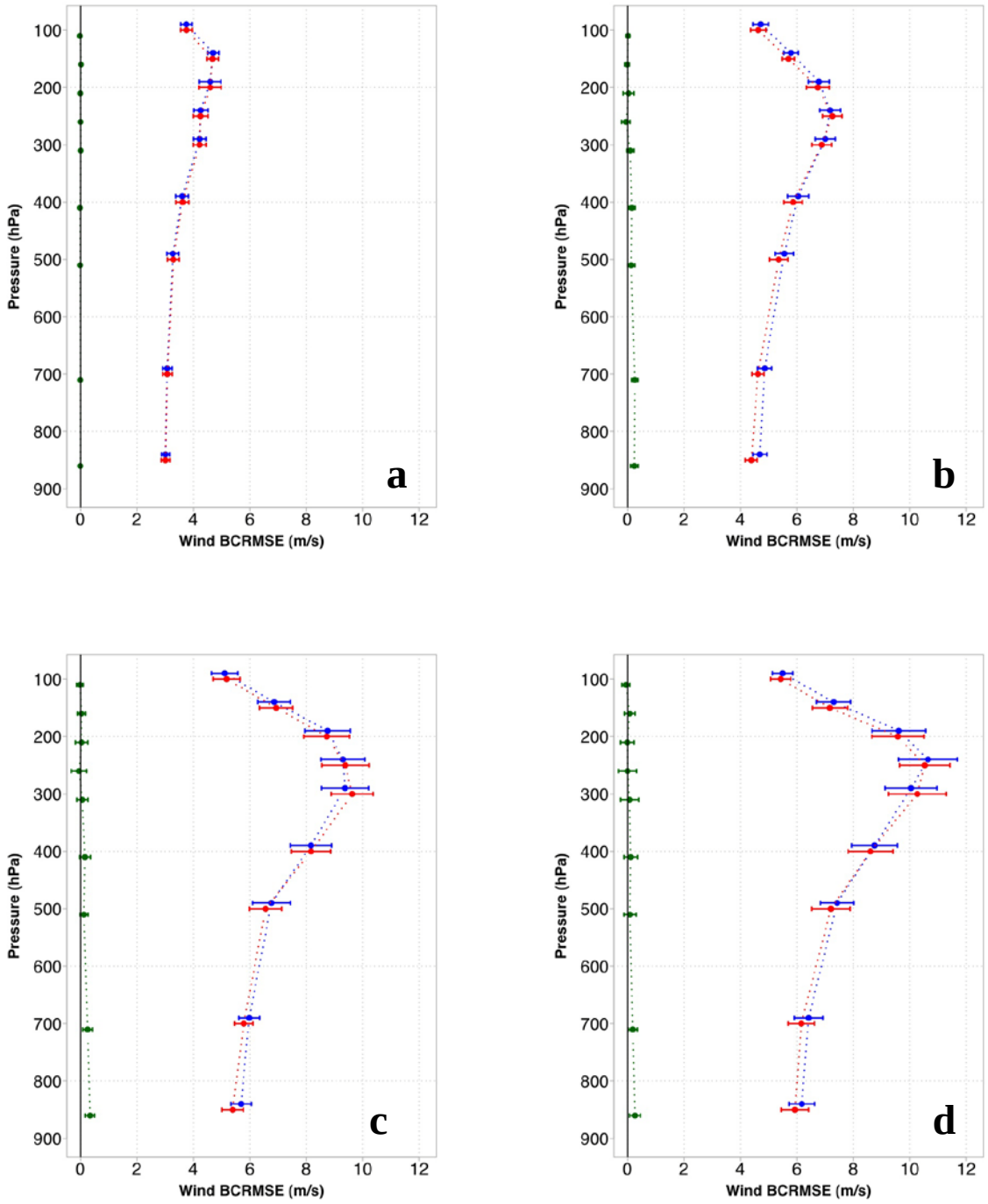

Figure 18. Same as Fig. 13 except for BCRMSE.

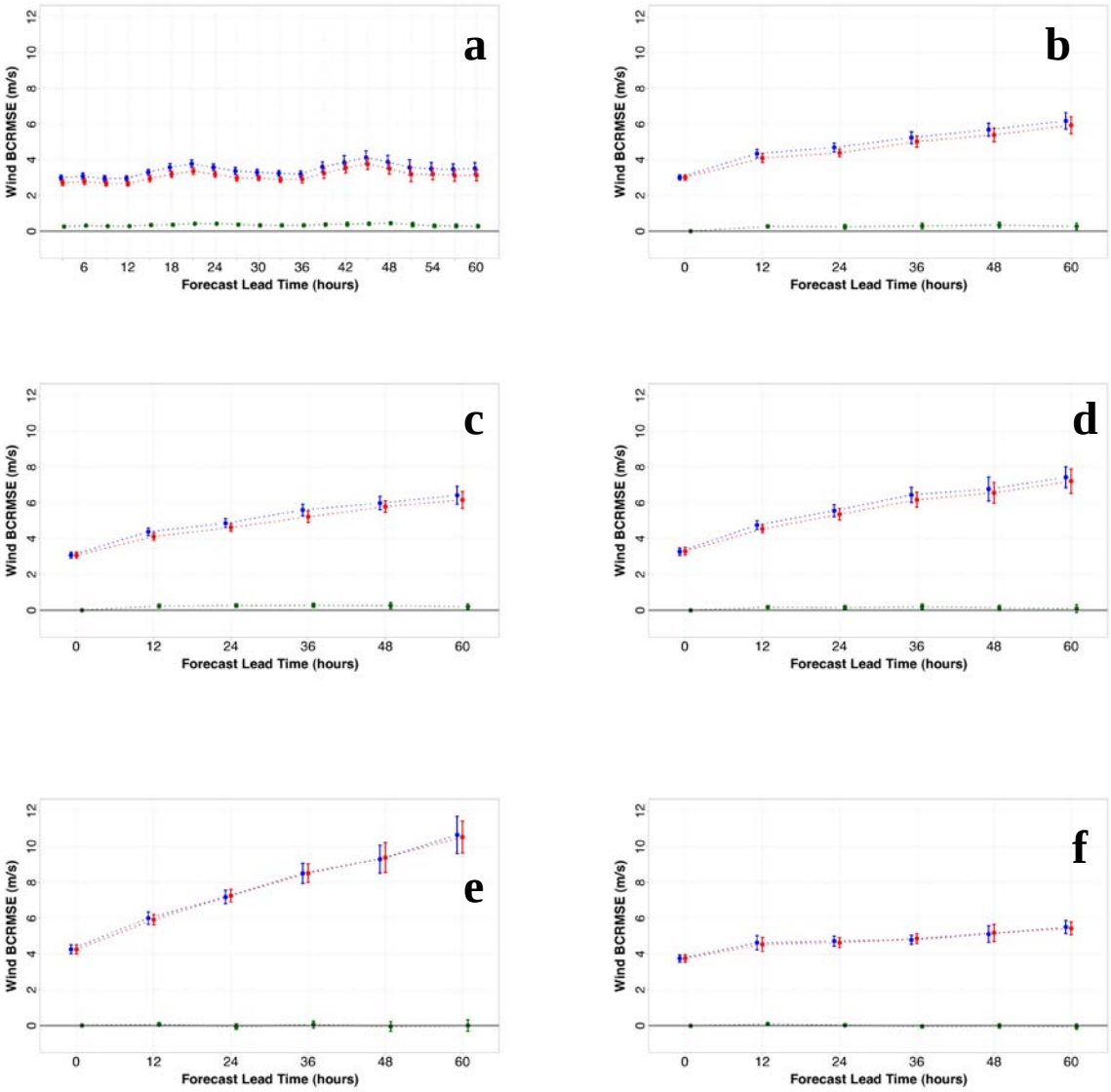

Figure 19. Same as Fig. 13 except for BCRMSE.

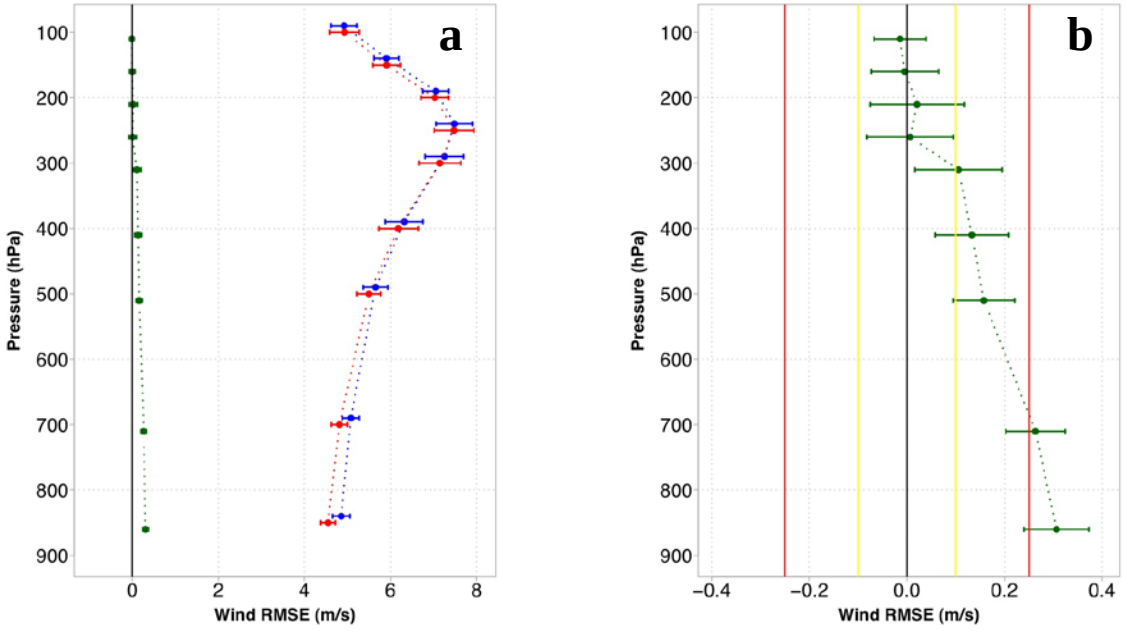

Figure 20. Vertical profile of annual mean wind RMSE (ms^{-1}) at the 24-h leadtime for a) ARW, NMM and ARW-NMM results, b) detail of ARW-NMM results, indicating the values of concern (between yellow and red lines) and serious concern (magnitude greater than red line). ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Horizontal bars represent the 99% CIs.

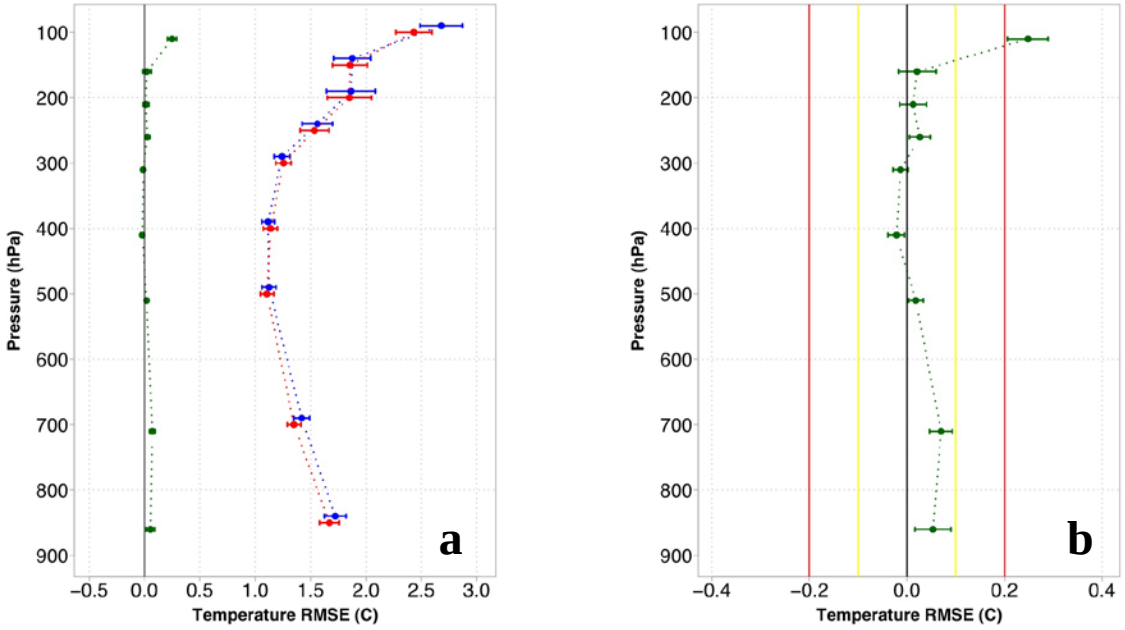

Figure 20. Vertical profile of annual mean wind RMSE (ms^{-1}) at the 24-h leadtime for a) ARW, NMM and ARW-NMM results, b) detail of ARW-NMM results, indicating the values of concern (between yellow and red lines) and serious concern (magnitude greater than red line). ARW is shown in blue, NMM in red and the ARW-NMM difference, in green. Horizontal bars represent the 99% CIs.

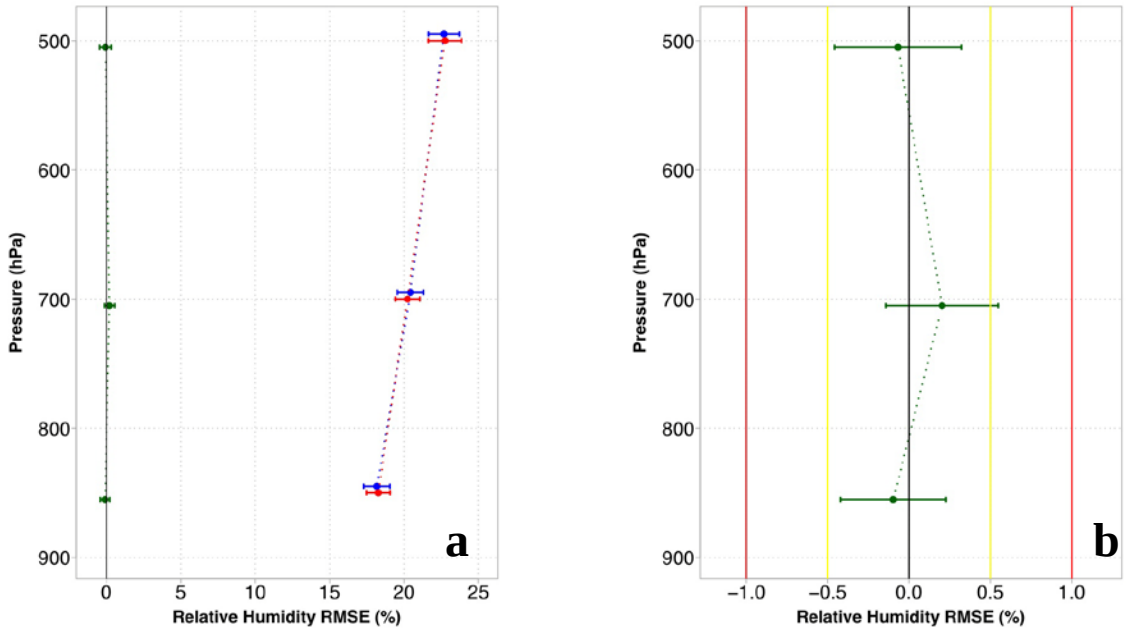

Figure 22. Same as Fig. 20 except RH RMSE (%).

The Developmental Testbed Center

2007 13-km Dynamic Core Test Report Addendum

Jamie Wolff

June 30, 2010

1. Introduction

Recently, the DTC has worked to streamline the end-to-end testing and evaluation system and in the process the scripts used to perform the precipitation analysis and plotting were rewritten. Due to this rewrite, the confidence intervals for the 24-h precipitation accumulation statistics were slightly altered. The analysis has been reexamined and detailed in this addendum. The overall result for the individual configurations and their tendencies to overpredict at lower thresholds and underpredict at higher thresholds did not change significantly. However, there are several SS pair-wise differences that are now evident for bias. The favored configuration was dependent on the threshold and season, with no consistently better performer.

2. Results addendum

Section 4a: Precipitation accumulated in 24 h

Since the RFC precipitation analysis used for verification is only published at 12 UTC, for the forecast cycles initialized at 12 UTC, verification is available for the 24- and 48-h lead times, while for the forecast cycles initialized at 00 UTC, verification is available for the 36- and 60-h lead times. The 24-h and 60-h results will be presented in this report.

The biases for the ARW and NMM for the 24-h lead time are shown in Fig. 4a for several thresholds. For the annual aggregation, overprediction is noted for both cores at the 0.01- and 0.10-in thresholds and underprediction occurs for the NMM core at the 0.50- and 0.75-in thresholds. For the other thresholds, the results are not conclusive regarding over or under-prediction. The CIs are noticeably larger for the higher thresholds, reflecting large variability and a smaller sample size.

When the seasonal distribution is examined, statistically significant (SS) bias results are present for additional thresholds (Table 3). For both cores, summer and spring have SS overprediction at the lower thresholds, while summer also displays a marked underprediction at the intermediate thresholds. Winter also has SS overprediction at the lower thresholds, but for the ARW core only, while fall has SS underprediction at high thresholds.

Table 3. *Forecasts of 24-h accumulated precipitation for the 24-h lead time for both dynamical cores classified as under or overprediction at all time periods and thresholds. Only SS results are presented.*

Threshold (in)	Time period	Prediction	Core
0.01	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
	Winter	Over	ARW

0.10	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
	Winter	Over	ARW
0.50	Annual	Under	NMM
	Summer	Under	Both
0.75	Annual	Under	NMM
	Summer	Under	Both
1.00	Summer	Under	Both
1.50	Fall	Under	NMM
2.0	Fall	Under	Both

At the 60-h lead time (Fig. 4b), the annual aggregation shows overprediction at the 0.01- and 0.10-in thresholds, and the results are inconclusive for higher thresholds. There is not a pronounced change in bias magnitude between the 24- and 60-h forecasts, however, the CIs are wider than at 24 h, indicating more variability in the sample. The seasonal decomposition is presented in Table 4. The differences among the seasons are more pronounced at the 60-h lead time, with fall generally displaying the lowest values of mean biases at all thresholds and the highest number of SS underprediction at thresholds above 0.50-in, especially for the NMM core. Winter has the highest mean biases at the mid thresholds with the ARW core exhibiting SS overprediction, while summer and spring have the highest mean biases and SS overprediction at the lower thresholds for both cores.

Table 4. *Same as Table 3, but for the 60-h lead time.*

Threshold (in)	Time period	Prediction	Core
0.01	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
	Winter	Over	Both
0.10	Annual	Over	Both
	Summer	Over	Both
	Spring	Over	Both
	Winter	Over	Both
0.25	Spring	Over	Both
	Winter	Over	ARW
0.50	Summer	Under	ARW
	Winter	Over	ARW
0.75	Fall	Under	NMM
1.00	Fall	Under	Both
2.0	Fall	Under	NMM

For both lead times (Fig 5), at the lowest thresholds for the annual and winter aggregations, the NMM core is favored when there are SS pair-wise differences, while most SS pair-wise differences during the summer favor the ARW core. For the 24-h lead time in the 0.5- to 1.5-in threshold range the SS pair-wise differences generally favor the ARW core. Overall, there are less SS pair-wise bias differences for the longer lead time.

Table 5. *SS pair-wise differences of bias for the 24-h lead time, where the favored core is highlighted.*

Threshold (in)	Time period	Core
0.01	Annual	NMM
	Summer	NMM
	Fall	ARW
	Winter	NMM
0.10	Summer	ARW
	Winter	NMM
0.25	Summer	ARW
	Fall	NMM
	Winter	NMM
0.50	Annual	ARW
	Winter	ARW
0.75	Annual	ARW
	Summer	ARW
1.0	Annual	ARW
	Summer	ARW
	Winter	ARW
	Spring	NMM
1.5	Annual	ARW
	Summer	ARW

Table 6. *Same as Table 5, but for the 60-h lead time.*

Threshold (in)	Time period	Core
0.01	Annual	NMM
	Fall	NMM
	Winter	NMM
0.10	Summer	ARW
	Winter	NMM
0.25	Summer	ARW
	Winter	NMM

The ETS for the 24–h lead time (Fig. 6a) displays its maximum values for the lowest thresholds and is smaller for higher thresholds. The seasonal breakdown indicates that summer has the lowest scores for all thresholds. The 60-h results (Fig. 6b) indicate a loss of skill with lead time, especially for the summer season. Fall and winter have the highest aggregated forecast skill at all thresholds at both lead times.

The ETS differences between ARW and NMM are shown in Fig. 7. The differences are very small and are not SS for any threshold at the 24-h lead time. Two SS pair-wise differences are noted at the 60-h lead time, both favoring the ARW core during fall at the 0.75- and 1.0-in thresholds.

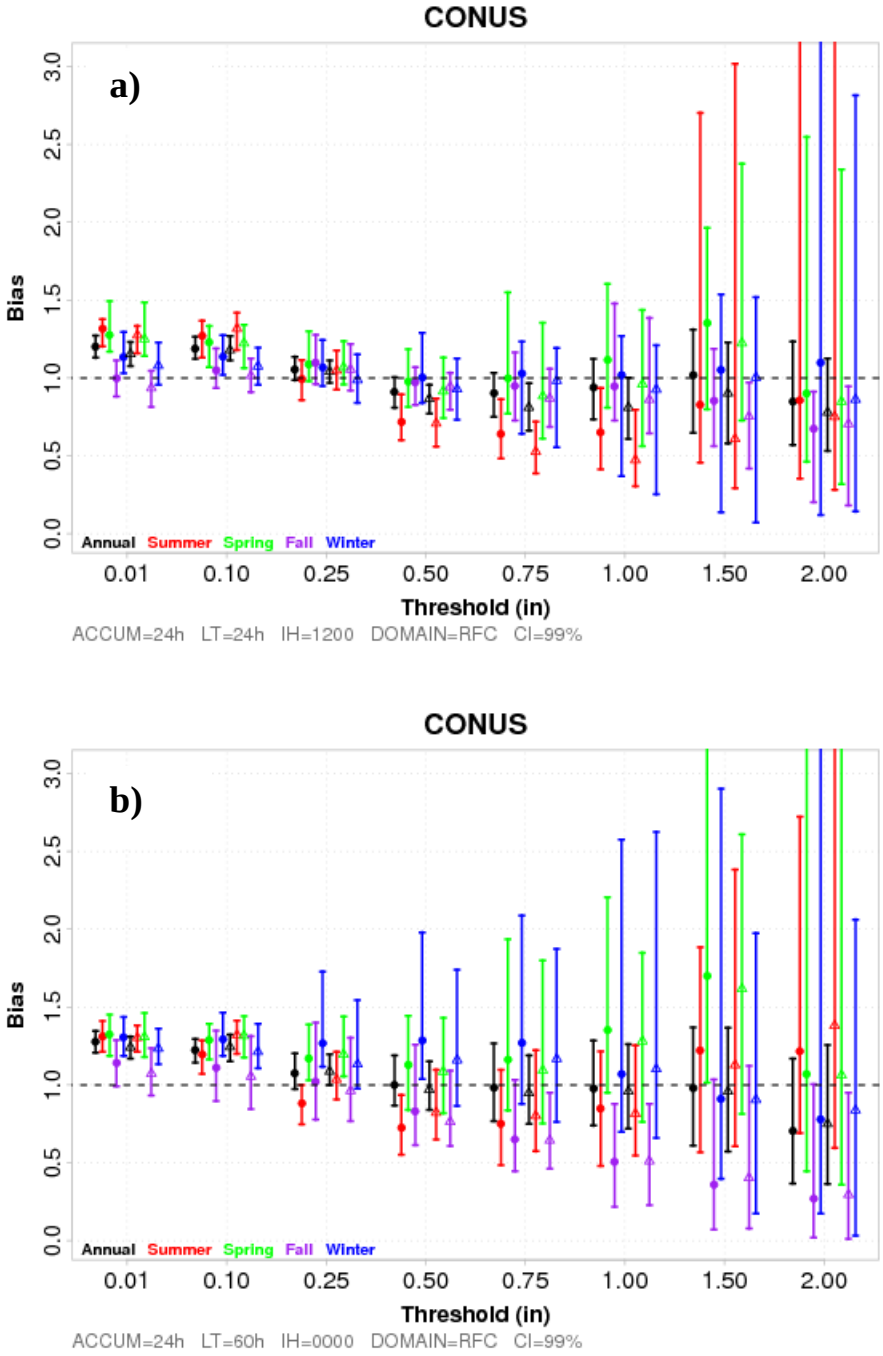

Figure 4. Bias for 24-h accumulated precipitation at the a) 24-h and b) 60-h lead times. ARW/NMM are circles/triangles. Annual mean in black, summer in read, spring in green, fall in purple and winter in blue. Vertical bars represent the 99% CIs.

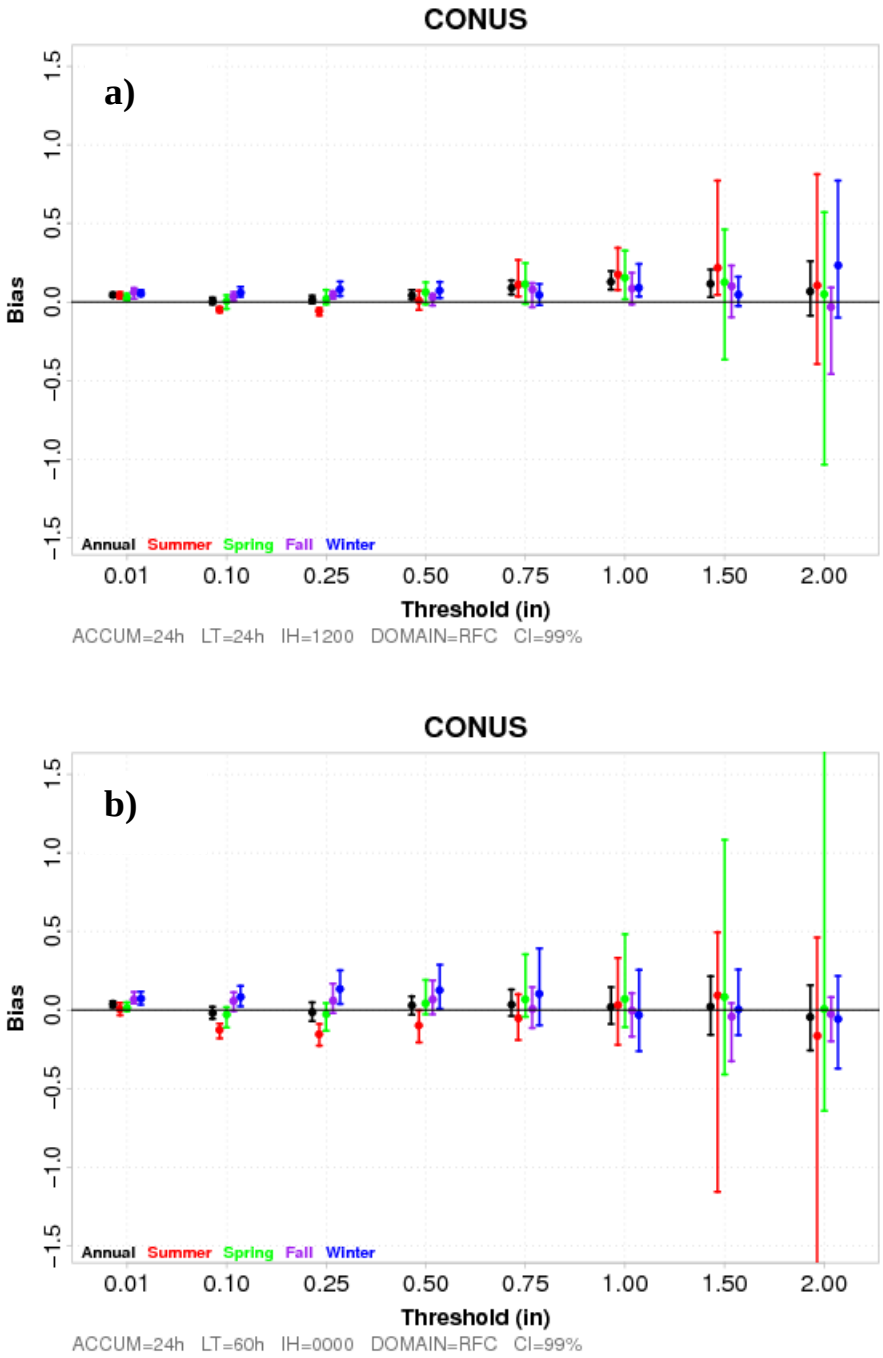

Figure 5. ARW-NMM difference in bias for 24-h accumulated precipitation at the a) 24-h and b) 60-h lead times. Annual mean in black, summer in read, spring in green, fall in purple and winter in blue. Vertical bars represent the 99% CIs.

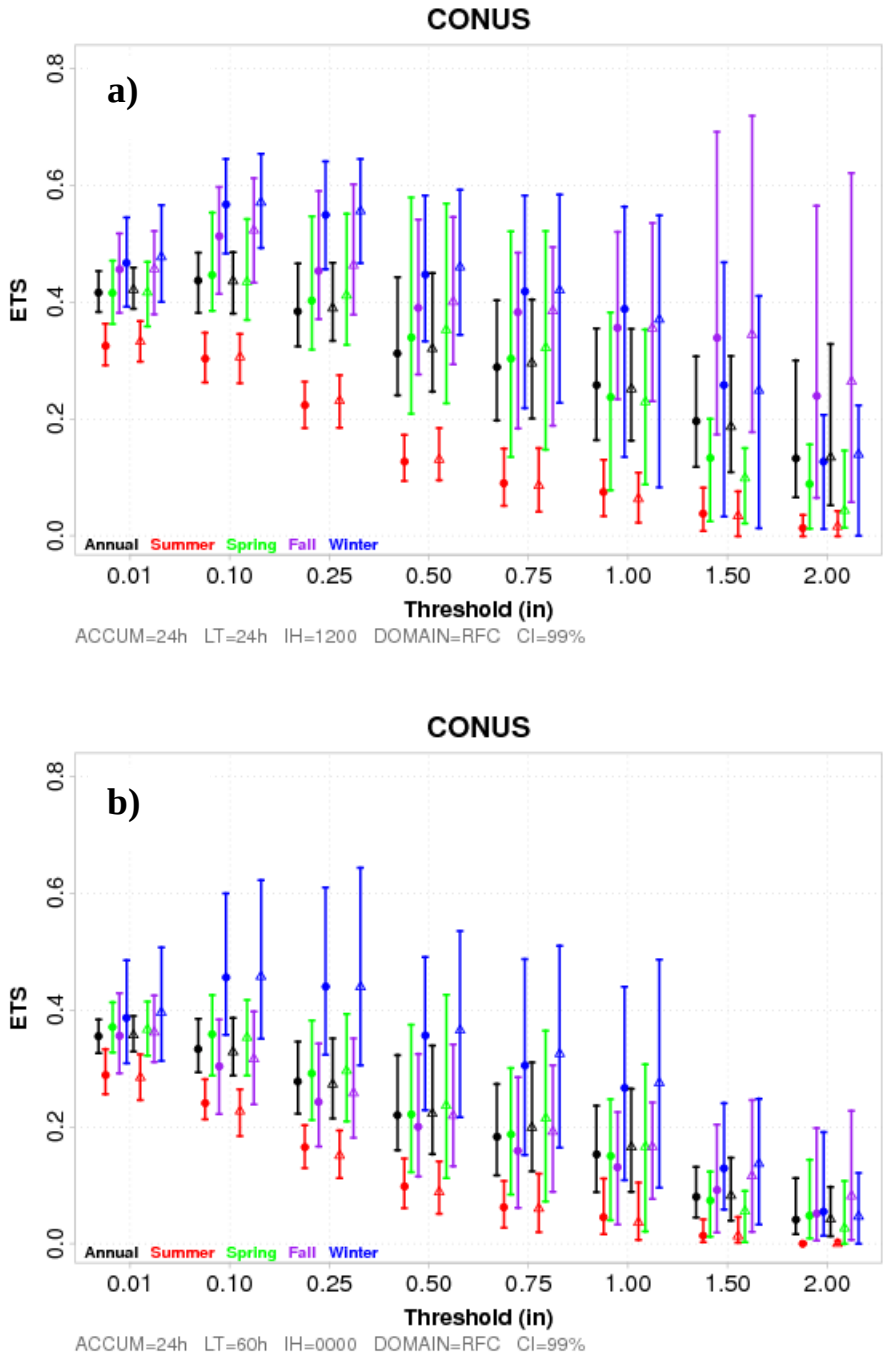

Figure 6. Same as Fig. 4, except for the ETS.

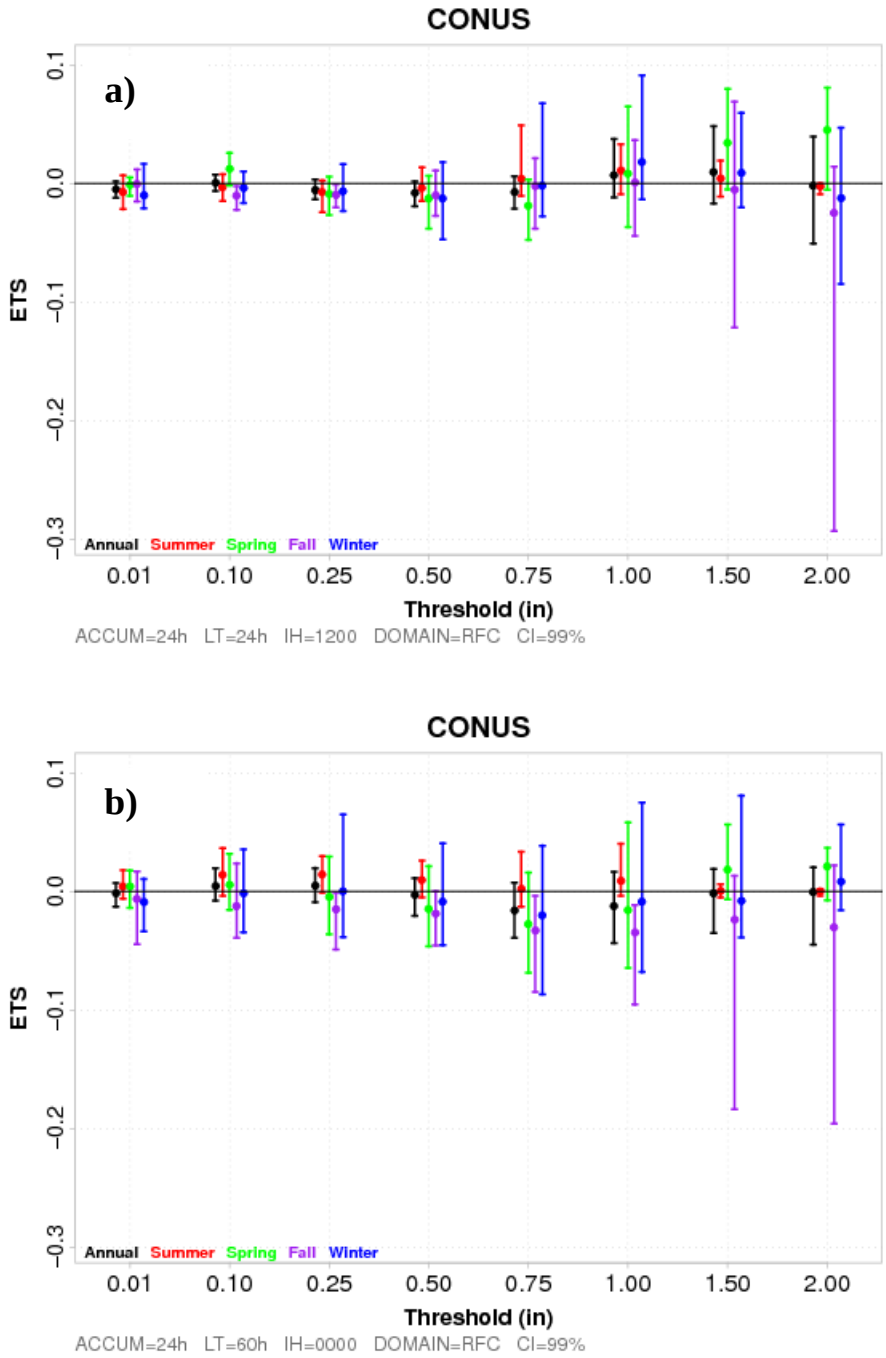

Figure 7. Same as Fig. 5, except for the ETS.

The Developmental Testbed Center

2007 13-km Dynamic Core Test Report ERRATA

Ligia Bernardet

October 30, 2008

1. Page 1. Andrew Loughe should be added as one of the contributors from NOAA ESRL.
2. Section 3d, last paragraph. The positive-definite advection scheme was not employed.
3. Section 4b, last paragraph. Instead of “errors increasing in the early morning and early afternoon.” it should read “with the largest errors at 3 AM and 3 PM MST.”
4. Fig. 16. The units of wind speed bias are m s^{-1} and not C.
5. Fig. 18, caption should read “Same as Fig. 16, except for BCRMSE”.
6. Fig. 19, caption should read “Same as Fig. 17, except for BCRMSE”.
7. The correct URL for Nance (2006) is http://ruc.fsl.noaa.gov/coretest2/DTC_report.pdf