The Developmental Testbed Center's Final Report on the Impact of Initializing with Differing Land Information System Output Data

30 January 2013

Executive Summary

The Weather Research and Forecasting (WRF) model is a mesoscale numerical weather prediction system utilized in both research and operational forecasting applications. The model is configurable to the users' requirements and suitable for a broad spectrum of weather regimes. Included in the flexibility provided by WRF is the users' ability to generate and/or select input files to initialize the model. While it is imperative to evaluate the performance of the model itself, it is also necessary to assess how forecast performance is affected by the data used in the initialization of WRF. At the request of the sponsor, the Air Force Weather Agency (AFWA), the Developmental Testbed Center (DTC) performed a rigorous test and evaluation which assessed forecast performance when initializing WRF with lower boundary conditions from differing versions of the Noah Land Surface Model (LSM) within AFWA's Land Information System (LIS). The test was conducted in a functionally similar operational environment to AFWA; each configuration was initialized with a 6-hour "warm-start" spin up, including the WRF Data Assimilation (WRFDA) component. Version 3.4 of the WRF model with the Advanced Research WRF (ARW) dynamic core was used. Two configurations were extensively tested, where the only difference between them was the version of the Noah LSM within the LIS; this allows for a direct assessment of how differing LIS output affects the performance of the model. For this testing, the two versions of the Noah LSM within the LIS were v2.7.1 and v3.3; the configuration using Noahv2.7.1 in the LIS was used as a baseline. The configurations will be referred to as LIS2 (WRFv3.4 initialized with LIS output using Noahv2.7.1) and LIS3 (WRFv3.4 initialized with LIS output using Noahv3.3). This report focuses on the pair-wise differences between the standard verification metrics for the two configurations, including an assessment of the statistical significance (SS) and practical significance (PS). Bias-corrected root-mean-square-error (BCRMSE) and bias were evaluated for surface and upper air temperature, dew point temperature, and wind speed; Gilbert Skill Score (GSS) and frequency bias were evaluated for 3-hourly and daily quantitative precipitation forecasts (QPF). The following points summarize the SS and PS differences seen in the verification results between LIS2 and LIS3.

- *Upper air temperature*
  - **BCRMSE**: LIS2 SS favored for all temporal aggregations and most lead times at and below 700 hPa; no PS differences were noted
  - **Bias**: Majority of SS/PS pair-wise differences show LIS2 as a better performer; the only PS pair-wise differences are at the 24 – 48 hour forecast lead times during the summer and spring aggregations, favoring LIS2
- *Upper air dew point temperature*
  - **BCRMSE**: All SS/PS pair-wise differences favor LIS2 and are seen predominantly during the annual, summer, and spring aggregations; PS pair-wise differences are seen in the summer and spring aggregations at and below 700 hPa
  - **Bias**: Most SS/PS pair-wise differences favor LIS3, with the most differences seen in the annual, winter, and spring aggregations; with exception to one PS pair-wise difference favoring LIS2 in the winter aggregation, all other PS pair-wise difference indicate LIS3 is the optimal performer
- Upper *air wind speed*
  - **BCRMSE**: Few SS pair-wise differences are observed, with all occurring at or below 500 hPa and with all but one favoring LIS2; none are PS
  - **Bias**: Few SS pair-wise differences are observed, with performance dependent on temporal aggregation, vertical level, and forecast lead hour; none are PS
- *Surface temperature*

        - **BCRMSE**: A diurnal trend is noted in SS/PS pair-wise differences, with the largest number of SS pair-wise differences favoring LIS2 and occurring during the daytime hours and a few SS differences favoring LIS3 during the overnight hours; only one difference, however, is PS
        - **Bias**: All forecast lead times for all temporal aggregations and both initializations, have not only SS but PS pair-wise differences indicating LIS2 is the better performer
    - *Surface dew point temperature*
        - **BCRMSE**: A diurnal trend in which configuration is favored is noted, with LIS3 generally favored with SS pair-wise differences during times valid overnight/early morning and LIS2 typically performing better during times valid in the afternoon/early evening hour; overall, more SS pair-wise differences are seen favoring LIS3; however, all PS pair-wise differences favor LIS2 and are seen in the spring aggregation at hours valid between 21 – 00 UTC
        - **Bias**: When differences are noted, almost all as PS; LIS3 is generally favored during times valid overnight/early morning and LIS2 typically performing better at all other times; very few differences are seen in the fall and winter temporal aggregations
    - *Surface wind speed*
        - **BCRMSE**: SS pair-wise differences are generally seen at times valid between 15 – 00 UTC, with more seen during the 12 UTC initializations; all but two SS pair-wise differences favor LIS2, and no differences are PS
        - **Bias**: A number of SS pair-wise differences are noted between 15 – 00 UTC, regardless of initialization or temporal aggregation, and predominantly indicating LIS3 is a better performer; no PS pair-wise differences are observed
    - *Three-hourly QPF*
        - **GSS**: SS pair-wise differences are noted but are dependent on initialization, forecast lead time, and precipitation threshold, with most differences favoring LIS2 and seen during the 12 UTC initializations; the only SS pair-wise differences for the 00 UTC initializations are noted at the 1.00" threshold
        - **Frequency Bias**: No SS differences were noted
    - *Daily QPF*
        - **GSS**: Several, scattered SS pair-wise differences observed, all favoring LIS2; a majority of the differences are seen at the 0.01" threshold
        - **Frequency Bias**: No SS differences were noted
    - Regardless of initialization or temporal aggregation, the GO Index indicates LIS2 configuration is more skillful than LIS3

1. Introduction

The Weather Research and Forecasting (WRF) model is a mesoscale numerical weather prediction system utilized in both research and operational forecasting applications. The model is configurable to the users' requirements and suitable for a broad spectrum of weather regimes. Included in the flexibility provided by WRF is the users' ability to generate and/or select input files to initialize the model. While it is imperative to evaluate the performance of the model itself, it is also necessary to assess how forecast performance is affected by the data used in the initialization of WRF. The Air Force Weather Agency (AFWA) is interested in determining whether consistent LSMs in the model and LIS produce better model forecasts. To address this inquiry, the Developmental Testbed Center (DTC) performed a rigorous test and evaluation which assessed forecast performance when initializing the WRF model with data from two, differing versions of the Noah LSM. The test was conducted in a functionally similar operational environment to AFWA, with each configuration initialized with a 6-hour "warm-start" spin up, including data assimilation; the Advanced Research WRF (ARW) dynamic core in WRF (Skamarock et al. 2008) was used. The only difference in the extensive testing was the version of the Noah LSM within the LIS; this allows for a direct assessment of how differing LIS initialization data affects the performance of the model. For this testing, the two versions of the Noah LSM within the LIS were v2.7.1 and v3.3; the configuration using Noahv2.7.1 in the LIS was used as a baseline. The configurations will be referred to as LIS2 (WRFv3.4 initialized with output from the LIS using Noahv2.7.1) and LIS3 (WRFv3.4 initialized

with output from the LIS using Noahv3.3). In addition to documenting the performance of the two configurations against each other, LIS2 will be designated as DTC Reference Configuration (RC) and the results made available to the WRF community.

2. Experiment Design

For this test, the end-to-end forecast system consisted of the WRF Preprocessing System (WPS), WRF Data Assimilation (WRFDA) system, WRF, Unified Postprocessor (UPP) and the NCAR Command Language (NCL) for graphics generation. Post-processed forecasts were verified using the Model Evaluation Tools (MET). In addition, the full data set was archived and made available for dissemination to the user community. The codes utilized were based on the official released versions of WPS (v3.4), WRFDA (v3.4), WRF (v3.4), UPP (v1.1), and MET (v4.0). MET included relevant bug fixes that were checked into the code repository prior to testing.

2.1 Forecast Periods

Forecasts were initialized every 36 hours from 1 July 2011 through 29 June 2012, consequently creating a default of initialization times including both 00 and 12 UTC, for a total of 244 cases (see Appendix A for a list of the cases). The forecasts were run out to 48 hours with output files generated every 3 hours.

The tables below list the forecast initializations that failed to complete the end-to-end process; the missing data and reason for failure is described in the table. All missing forecasts were due to missing or bad input data sets, not model crashes. A total of 239 cases ran to completion and were used in the verification results.

Missing forecasts:

Affected Cycle	Missing data	Reason
2011080112	WRF output	Bad SST input data
2011082400	WRF output	Missing SST input data
2012050312	WRF output	Missing GFS input data
2012050612	WRF output	Bad obs_gts input data
2012060400	WRF output	Bad SST input data

Missing verification:

Affected Cycle	Missing data	Reason
2011072500	Missing 3-h QPF verification for 18 – 21-h Missing 24-h QPF verification for 36-h	Missing ST2 analysis

2.2 Initial and Boundary Conditions

Initial conditions (ICs) and lateral boundary conditions (LBCs) were derived from the 0.5° x 0.5° Global Forecast System (GFS). Output from AFWA's LIS running with version 2.7.1 and version 3.3 of the Noah LSM were used to initialize the lower boundary conditions (LoBCs); the two sets of files used for initializing the LoBCs were generated by AFWA and then provided to the DTC for the testing period. In addition, a daily, real-time sea surface temperature product from Fleet Numerical Meteorology and Oceanography Center (FNMOC) was used to initialize the sea surface temperature (SST) field for the forecasts.

The time-invariant components of the LoBCs (topography, soil, vegetation type, etc.) were derived from United States Geological Survey (USGS) input data and were generated through the *geogrid* program of WPS. The *avg_tsfc* program of WPS was also used to compute the mean surface air temperature in order to provide improved water temperature initialization for lakes and smaller bodies of water in the domain that are further away from an ocean.

A 6-hour "warm start" spin-up procedure (Fig. 1) preceded each forecast. Data assimilation using WRFDA was conducted at the beginning and end of the 6-hour window using observation data files provided by AFWA. At the beginning of the data assimilation window, the GFS derived initial conditions were used as the model background, and at the end of the window, the 6-hour WRF forecast initialized by the WRFDA analysis was used. After each WRFDA run, the LBCs initially derived from GFS were updated and used in the subsequent forecasts.

Seasonal, domain-specific model background error statistics (BE) files were created and used in WRFDA. To create the appropriate BE files, cold-start WRF forecasts were conducted on the 15 km grid twice daily for 15 days each season. Essentially, this was 30 forecasts per season, or 120 total forecasts (24-h forecasts, in 12-h increments). The *gen_be* utility in WRFDA was then used to generate BE files from those model runs.

2.3 Model Configuration Specifics

2.3.1 Domain Configuration

A 15-km contiguous U.S. (CONUS) grid was employed for this test. The domain (Fig. 2) was selected such that it covers complex terrain, plains, and coastal regions spanning from the Gulf of Mexico, north, to Central Canada in order to capture diverse regional effects for worldwide comparability. The domain was 403 x 302 gridpoints, for a total of 121,706 gridpoints. The Lambert-Conformal map projection was used and the model was configured to have 56 vertical levels (57 sigma entries) with the model top at 10 hPa.

2.3.2 Other Aspects of Model Configuration

The table below lists AFWA's current OC that was used in this testing. The model configuration based on version 3.4 of the WRF system.

	Current AFWA OC (AFWA)
Microphysics	WRF Single-Moment 5 scheme
Radiation LW and SW	RRTM/Dudhia schemes
Surface Layer	Monin-Obukhov similarity theory
Land-Surface Model	Noah
Planetary Boundary Layer	Yonsei University scheme
Convection	Kain-Fritsch scheme

Both configurations were run with a long timestep of 90 s, and an acoustic step of 4 was used. Calls to the boundary layer, and microphysics were performed every time step, whereas the cumulus parameterization was called every 5 minutes and every 30 minutes for the radiation.

The ARW solver offers a number of run-time options for the numerics, as well as various filter and damping options (Skamarock et al. 2008). The ARW was configured to use the following numeric options: 3rd-order Runge-Kutta time integration, 5th-order horizontal momentum and scalar advection, and 3rd-order vertical momentum and scalar advection. In addition, the following filter/damping options were utilized: three-dimensional divergence damping (coefficient 0.1), external mode filter (coefficient 0.01), off-center integration of vertical momentum and geopotential equations (coefficient 0.1), vertical-velocity damping, and a 5-km-deep diffusive damping layer at the top of the domain (coefficient 0.02). Positive-definite moisture advection was also turned on.

In the extensive testing, it was discovered that two namelist options in WRFv3.4 are not compatible [*hypsoemtric_opt* set to 2 (default) and *adjust_heights* set to true (default is false)]. The *hypsoemtric_opt = 2* option is new to WRF as of version 3.4 and was incorporated as the default. The AFWA OC that the DTC has been testing over several years (i.e., versions prior to v3.4) has the *adjust_heights* namelist option set to true. The original set up for this test caused several model crashes in the summer months of

the year-long test due to the incompatibility of the namelist options. A new check to assure *adjust_heights* set to true is not used with *hypsoemtric_opt* set to 2 has been added to the WRF code repository due to discovering this issue and will be released with the next official code distribution. Users will want to examine their v3.4 namelists to ensure they do not use these options concurrently.

Appendix B provides relevant portions of the *namelist.input* file.

2.4 Post-processing

The *unipost* program within UPP was used to destagger the forecasts, to generate derived meteorological variables, and to vertically interpolate fields to isobaric levels. The post-processed files included two- and three-dimensional fields on constant pressure levels, both of which were required by the plotting and verification programs. Three-dimensional post-processed fields on model native vertical coordinates were also output and used to generate graphical forecast sounding plots.

3. Model Verification

The MET package was used to generate objective model verification. MET is comprised of grid-to-point verification, which was utilized to compare gridded surface and upper-air model data to point observations, as well as grid-to-grid verification, which was utilized to verify QPF. Verification statistics generated by MET for each retrospective case were loaded into a MySQL database. Data was then retrieved from this database to compute and plot specified aggregated statistics using routines developed by the DTC in the statistical programming language, R.

Several domains were verified for the surface and upper air, as well as precipitation variables. Area-average results were computed for the CONUS domain, East and West domains, and 14 sub-domains (Fig. 3). Only the CONUS domain is described in detail for this report, with occasional mention to the East and West domains; however, all East, West, and sub-domain results are available on the DTC website (http://www.dtcenter.org/eval/afwa_test/wrf_v3.4/). In addition to the regional stratification, the verification statistics were also stratified by vertical level and lead time for the 00 UTC and 12 UTC initialization hours combined, and by forecast lead time and precipitation threshold for 00 UTC and 12 UTC initialized forecasts individually for surface fields in order to preserve the diurnal signal.

Each type of verification metric is accompanied by confidence intervals (CIs), at the 99% level, computed using the appropriate statistical method. Both configurations were run for the same cases allowing for a pair-wise difference methodology to be applied, as appropriate. The CIs on the pair-wise differences between statistics for the two configurations objectively determines whether the differences are statistically significant (SS); if the CIs on the median pair-wise difference statistics include zero, the differences are not statistically significant. Due to the nonlinear attributes of frequency bias, it is not amenable to a pair-wise difference comparison. Therefore, the more powerful method to establish SS could not be used and, thus, a more conservative estimate of SS was employed based solely on whether the aggregate statistics, with the accompanying CIs, overlapped between the two configurations. If no overlap was noted for a particular threshold, the differences between the two configurations were considered SS.

Due to the large number of cases used in this test, many SS pair-wise differences were anticipated. In many cases, the magnitude of the SS differences was quite small and did not yield practically meaningful results. Therefore, in addition to determining SS, the concept of establishing practical significance (PS) was also utilized this test. PS was determined by filtering results to highlight pair-wise differences greater than the operational measurement uncertainty requirements and instrument performance as specified by the World Meteorological Organization (WMO; http://www.wmo.int/pages/prog/www/IMOP/publications/CIMO-Guide/1st-Suppl-to-7th_draft/pdf/Annex_I_1B.pdf). To establish PS between the two configurations, the following criteria was applied: temperature and dew point temperature differences greater than 0.1 K and wind speed

differences greater than 0.5 m s^{-1}. PS was not considered for metrics used in precipitation verification [i.e., Gilbert Skill Score (GSS) or frequency bias] because those metrics are calculated via a contingency table, which is based on counts of yes and no forecasts.

3.1 Temperature, Dew Point Temperature, and Winds

Forecasts of surface and upper air temperature, dew point temperature, and wind were bilinearly interpolated to the location of the observations (METARs and RAOBS) within the National Centers for Environmental Prediction (NCEP) North American Data Assimilation System (NDAS) prepbufr files. Objective model verification statistics were then generated for surface (using METAR) and upper air (using RAOBS) temperature, dew point temperature, and wind. Because shelter-level variables are not available from the model at the initial time, surface verification results start at the 3-hour lead time and go out 48 hours by 3-hour increments. For upper air, verification statistics were computed at the mandatory levels using radiosonde observations and computed at 12-hour intervals out to 48 hours. Because of known errors associated with radiosonde moisture measurements at high altitudes, the analysis of the upper air dew point temperature verification focuses on levels at and below 500 hPa. Bias and bias-corrected root-mean-square-error (BCRMSE) were computed separately for surface and upper air observations. The CIs were computed from the standard error estimates about the median value of the stratified results using a parametric method and a correction for first-order autocorrelation.

3.2 Precipitation

For the QPF verification, a grid-to-grid comparison was made by first bilinearly interpolating the precipitation analyses to the 15-km model integration domain. This regridded analysis was then used to evaluate the forecast. Accumulation periods of 3 and 24 hours were examined. NCEP Stage II analysis was used as the observational dataset, and the data is available in hourly, 6-hourly, and 24-hourly accumulations. For this test, hourly data was summed for the 3-hour QPF verification, and daily QPF verification utilized the 24-hour accumulation files. The 24-hour accumulation observations are valid at 12 UTC; therefore, the daily QPF was examined for the 24- and 48-hour lead times for the 12 UTC initializations and 36-hour lead time for the 00 UTC initializations. Traditional verification metrics computed included the GSS and frequency bias. For the precipitation statistics, a bootstrapping CI method was applied.

3.3 GO Index

Skill scores (S) were computed for wind speed (at 250 hPa, 400 hPa, 850 hPa and surface), dew point temperature (at 400 hPa, 700 hPa, 850 hPa and surface), temperature (at 400 hPa and surface), height (at 400 hPa), and mean sea level pressure, using root-mean-square-error (RMSE) for both the LIS2 and LIS3 configurations using the formula:

$$S = 1 - \frac{(RMSE_{LIS3})^2}{(RMSE_{LIS2})^2}$$

For each variable, level, and forecast hour, predefined weights (w_i), shown in the table below, were then applied and a weighted sum, S_W, was computed

Variable	Level	Weights by lead time			
		12 h	24 h	36 h	48 h
Wind Speed	250 hPa	4	3	2	1
	400 hPa	4	3	2	1
	850 hPa	4	3	2	1
	Surface	8	6	4	2
Dew Point Temperature	400 hPa	8	6	4	2
	700 hPa	8	6	4	2
	850 hPa	8	6	4	2
	Surface	8	6	4	2
Temperature	400 hPa	4	3	2	1
	Surface	8	6	4	2
Height	400 hPa	4	3	2	1
Pressure	Mean sea level	8	6	4	2

where,

$$S_w = \frac{1}{\sum_i w_i}\left(\sum_i (w_i S_i)\right)$$

Once the weighted sum of the skill scores, S_w, was computed, the Index value (N) is defined as:

$$N = \sqrt{\frac{1}{1 - S_w}}$$

Given this definition, which is based on the General Operations (GO) Index, values (N) less than one indicate the LIS2 configuration has higher skill and values greater than one indicate the LIS3 configuration has higher skill.

4. Verification Results

Differences between two versions of the Noah LSM within AFWA's LIS are computed by subtracting LIS2 from LIS3. BCRMSE is always a positive quantity, and a perfect score is zero. This results in positive (negative) differences indicating the LIS2 (LIS3) configuration has a lower BCRMSE and is favored. Bias also has a perfect score of zero but can have positive or negative values; therefore, when examining pair-wise differences, it is important to note the magnitude of the bias in relation to the perfect score for each individual configuration to know which configuration has a smaller bias and is, thus, favored. For GSS, the perfect score is one, and the no-skill forecast is zero. Thus, if the pair-wise difference is negative (positive), the LIS2 (LIS3) configuration has a higher GSS and is favored. Frequency bias has a perfect score of one, but as described earlier, SS is determined by the overlap of CIs attached to the median value. A breakdown of the configuration with SS and PS better performance by variable, season, statistic, initialization hour, forecast lead time, and level is summarized in Tables 1-8, where the favored configuration is highlighted. The discussion below is focused on the CONUS domain; however, the East and West sub-domains were also investigated. If dissimilarities were noted when considering PS pair-wise differences in comparing the East and West sub-domains to the CONUS domain, the differences are briefly discussed. All verification plots generated (by plot type, metric, lead time, threshold, season, etc.) can be viewed on the DTC webpage.

4.1 Upper Air

4.1.1 Temperature BCRMSE and bias

For both configurations with the differing LSM versions, regardless of temporal aggregation or forecast lead time, a minimum in BCRMSE values is between 500 and 300 hPa, with the largest error occurring at the lower and upper-levels (Fig. 4). In general, the largest differences between the two configurations are seen at 850 and 700 hPa. This is reflected in the SS pair-wise differences, with a majority of the differences occurring at and below 700 hPa and favoring LIS2. For the lowest-levels, SS pair-wise differences are observed for all temporal aggregations and most forecast lead times, with the winter aggregation having the smallest number of differences (see Table 1). In the middle-to-upper levels, only a few differences are noted, and the configuration depends on forecast lead time and temporal aggregation. None of the SS pair-wise differences are PS.

For all temporal aggregations and forecast lead times, both configurations have a cool bias at 850 hPa, which transitions to a warm bias with height (Fig. 5). The annual, summer, and spring aggregations generally see the transition between 700 and 500 hPa, with fall and winter aggregations transitioning between 500 and 400 hPa. At 150 hPa, the bias value is either cold or have CIs encompassing zero for all forecast lead times and temporal aggregations except summer, where a warm bias is observed. For several forecast lead times, SS pair-wise differences are seen in all temporal aggregations and forecast lead times, with most differences favoring LIS2 (see Table 1). Several SS pair-wise differences showing LIS3 as the better performer are noted at 150 hPa in all but the summer aggregation. PS pair-wise differences, favoring LIS2, are seen at 850 hPa during the summer and spring aggregations at the 24-48 h forecasts.

4.1.2 Dew Point Temperature BCRMSE and bias

Both configurations display an increase in dew point BCRMSE as forecast lead time increases and pressure decreases in all temporal aggregations, except the winter aggregation, where at most forecast lead times, a slight decrease in median BCRMSE from 700 hPa to 500 hPa is observed (Fig. 6). All SS pair-wise differences noted at or below 700 hPa and show LIS2 as a better performer; all PS pair-wise differences noted are at 850 hPa in the summer and spring aggregations (see Table 2). In the East domain, similar results were seen as compared to the CONUS domain; however, no PS pair-wise differences were observed in the West domain (not shown).

For all temporal aggregations except summer, 850 hPa displays a wet bias at the 12-h forecast lead time, with dew point temperature bias nearing zero and becoming unbiased as forecast lead time increases; at 700 and 500 hPa, for all forecast lead times, a wet bias is present (Fig. 7). During the summer aggregation, at 850hPa, the 12 – 48-h forecast lead times have CIs that encompass zero, indicating an unbiased forecast, and the 500 hPa level has a wet bias. At 700 hPa, the 12 – 24-h forecast lead times are unbiased but transition to a cool bias at longer forecast lead times. Most SS/PS pair-wise differences favor LIS3; the only differences favoring LIS2 occur at the 48-h forecast lead time at 850 hPa during the annual, fall and winter aggregations (see Table 2). When examining the PS pair-wise differences in the East and West domains, as compared to the CONUS domain, several differences are observed (not shown). In the East domain, a number of PS pair-wise differences are observed during the annual and summer temporal aggregations with most favoring LIS2. The West domain also has a number of PS pair-wise differences, many of which favor LIS2, especially in the spring temporal aggregation. In the summer aggregation, all forecast lead times at 850 hPa are PS, favoring LIS2; whereas, no differences are noted in the CONUS domain.

4.1.3 Wind Speed BCRMSE and bias

For both configurations, BCRMSE generally increases from a minimum at 850 hPa to a maximum around 300 – 200 hPa, with decreasing error further aloft (Fig. 8). As forecast lead time increases, the errors

also show a tendency to increase. Overall, both configurations have very similar distributions of BCRMSE; this is reflected when examining the statistical significance, where only a few scattered SS pair-wise differences are noted (see Table 3). All differences are at and below 500 hPa, with no differences in the spring aggregation; all but one difference favors LIS2. No differences were considered PS.

At 850 hPa, both configurations, for all temporal aggregations, generally have a negative bias (i.e., winds are too light) that transitions towards an unbiased forecast with forecast lead time (Fig. 9); the unbiased forecasts in the seasonal aggregations are often a product of larger CIs encompassing zero. A negative bias is noted at nearly all forecast lead times and seasonal aggregations at 200 hPa; at levels above 200 hPa, forecasts become either unbiased or display a positive bias. A few SS pair-wise differences are seen, depending on forecast lead time and vertical level. Most occur in the annual aggregation and none in the spring aggregation, with only a few exceptions, LIS3 is the better performer (see Table 3).

4.2 Surface

4.2.1 Temperature BCRMSE and bias

The surface temperature BCRMSE generally displays an increase with forecast lead time, evidenced in both the 00 and 12 UTC initializations (Fig. 10). A diurnal trend is also noted with lower errors seen at times valid 06 – 09 UTC, with exception to the fall aggregation where the lowest errors are near 03 UTC. Maximum BCRMSE values typically occur at lead times valid at 15 UTC with a secondary peak near 00 UTC. Several SS pair-wise differences are observed, favoring LIS2 during times valid 15 – 00 UTC (see Table 4). A few SS pair-wise differences favoring LIS3 are seen in the annual and summer aggregations at times valid between 06 – 09 UTC. One PS difference is noted, favoring LIS2.

In general, both configurations have a cold bias, regardless of initialization time, forecast lead time, or temporal aggregations, with only a few exceptions noted in the temporal aggregations (Fig. 11). A strong diurnal signal is present for both configurations, regardless of initialization hour or temporal aggregation. The largest cold bias is seen in hours valid between 18 – 00 UTC; a less negative and occasionally unbiased forecast is evident at hours valid between 06 – 09 UTC. In general, median bias values for LIS2 are higher (i.e., less negative) than LIS3. This is reflected in the pair-wise differences; regardless of initialization, temporal aggregation, or forecast lead time, PS pair-wise differences favor LIS2 (see Table 4). This same behavior is also seen within the East verification domain; however, in the West verification domain, several PS pair-wise differences are observed in the fall and winter temporal aggregations, near the 12 UTC valid time, showing LIS3 as the better performer (not shown).

4.2.2 Dew Point Temperature BCRMSE and bias

Regardless of initialization and temporal aggregation, dew point temperature BCRMSE increases as lead time increases (Fig. 12). In addition, a diurnal trend in noted for all temporal aggregations, with the fall aggregation displaying only a weak signal. Depending on the exact initialization and temporal aggregation, the smallest errors are observed at times valid between 03 – 15 UTC, with the largest errors during times valid between 18 – 00 UTC. In general, the two configurations display similar trends in the distribution; however, a large number of SS pair-wise differences exist. For both the 00 and 12 UTC initialization, a number of the SS pair-wise differences show LIS3 as the better performer, typically at times valid between 03 – 12 UTC, with the least amount of differences occurring during the winter aggregation see (Table 5). However, all PS pair-wise differences noted favor LIS2 and occur at valid times of 21 or 00 UTC during the Spring temporal aggregation.

Both configurations display a diurnal signal regardless of initialization or season; however, the trend is strongest in the annual, spring, and summer aggregations (Fig. 13). In general, a SS wet bias is seen at hours valid from 18 – 00 UTC for both configurations, with the fewest differences in the fall and winter aggregations, where CIs are large and encompass zero. An unbiased or dry forecast is present for times valid between 06 – 12 UTC; this is especially apparent at the longer forecast lead times. The annual,

summer, and spring aggregations generally indicate the median values of the LIS3 configuration are consistently higher than LIS2, which translates to better performance when the bias is dry but worse performance when there is a wet bias. The fall and winter aggregations show more uncertainty with larger CIs, lending to a smaller number of SS/PS pair-wise differences see (Table 5). A number of SS/PS pair-wise differences are noted, with most being PS and occurring in the annual, summer, and spring aggregations. In the overnight hours, LIS3 is a better performer, while LIS2 is favored in the mid-morning to evening hours. The East verification region closely follows the CONUS domain; however, in the West, no PS pair-wise differences are seen in the annual temporal aggregation and very few forecast lead times are PS in the spring aggregation (not shown).

4.2.3 Wind BCRMSE and bias

Wind speed BCRMSE distribution displays a diurnal signal regardless of initialization in the annual, spring, and summer aggregations; lowest errors are seen at times valid at or near 12 UTC, with the largest errors seen at or near times valid at 00 UTC (Fig. 14). Increasing error growth with forecast lead time is also noted for the annual, spring, and summer aggregations. The fall and winter aggregations have a very weak diurnal signal and show relatively steady error with forecast lead time. SS pair-wise differences are generally seen at times valid between 15 – 00 UTC, with more occurring for the 12 UTC initializations (see Table 6). All but two SS pair-wise differences favor LIS2; no differences are PS.

With exception to a few lead times during the spring and summer aggregations, a high wind speed bias is observed regardless of initialization (Fig. 15). A prominent diurnal trend is also noted, with highest errors seen at times valid between 03 – 12 UTC and with lowest errors at times valid between 15 – 21 UTC. A number of SS pair-wise differences are present, with all but two differences favoring LIS3 (see Table 6). Differences are generally noted between 15 – 00 UTC, regardless of initialization or temporal aggregation. No PS pair-wise differences are observed.

4.2.4 3-hourly QPF GSS and bias

Regardless of configuration, initialization, or forecast lead time, median GSS values decrease as the threshold increases from 0.01" to 1.00" (Fig. 16). This behavior is also exhibited in the base rate, which is the measurement of observed grid box events to the total number of grid boxes in the domain. Higher base rates are often observed at lower thresholds and during the summer, regardless of threshold, due to an overall higher number of observed events. Lower base rates are often associated with higher thresholds due to the infrequency of high-precipitation and spatially expansive events. Several SS pair-wise differences are noted but are dependent on initialization, forecast lead time, and precipitation threshold, with most favoring LIS2 for the 12 UTC initializations; the only SS pair-wise differences for the 00 UTC initializations are noted at the 1.00" threshold (see Table 7).

In general, a high bias, regardless of initialization hour or forecast lead time, is present in all but the highest thresholds in the annual, fall, and winter aggregations (Fig. 17). The spring and summer aggregations show similar trends for times valid at 00 UTC for both initialization hours; at times valid at 12 UTC, unbiased forecasts are seen at most thresholds. Due to a small base rate, large CIs are noted at the highest thresholds, which leads to unbiased forecasts. No SS pair-wise differences are noted, regardless of initialization hour, forecast lead time, or threshold (not shown).

4.2.5 Daily Precipitation GSS and bias

In general, for both configurations, initializations, and forecast lead times, GSS decreases as threshold increases (Fig. 18). While the width of the CIs is large for most forecast lead times and thresholds, the fall aggregation displays an increase in median GSS from 0.01" to 0.25" for both initializations and all forecast lead times; for thresholds higher than 0.25", a decrease in median GSS is noted as threshold increases. In the winter season, a less-exponential decrease in GSS in noted, with larger CIs bounding the median GSS. The scattered pair-wise differences that are observed all favor LIS2, with the most differences seen at the 0.01" threshold (see Table 8).

Generally, regardless of configuration, initialization hour, temporal aggregation, or forecast lead hour, a high bias is present at most thresholds, with exception to the lowest and highest thresholds (Fig. 19). The highest threshold is occasionally unbiased, often due to the large width of the CIs; the lowest threshold has a high or unbiased forecast in all aggregations with exception to the summer, where the forecast is either unbiased or under-forecast. Similar to the 3-hour QPF, the width of the CIs increase with increasing threshold, likely due to the decreasing base rate, which indicates a lower level of confidence. No differences are noted for either initialization for any forecast hours or thresholds (not shown).

4.3 GO Index

Regardless of temporal aggregation or initialization hour, the median GO Index values and associated CIs, indicated by the width of the notches about the median on the boxplot, are less than one, indicating the LIS2 configuration is more skillful than the LIS3 configuration (Fig. 20). This finding is particularly enhanced at the 12 UTC initializations, where the median GO Index values are less than the 00 UTC initializations.

5. Summary

An end-to-end sensitivity test was performed to test and evaluate the performance of AFWA's operational physics suite when initialized with LIS output generated with differing versions of the Noah LSM. One configuration utilized version 2.7.1 of the Noah LSM, and the second configuration used version 3.3 of the Noah LSM; both configurations were run with version 3.4 of WRF-ARW. Each configuration included a 6-hr warm-start data assimilation procedure and was run over the same set of cases, spanning one year. The goal of this testing effort was to assess the potential impacts of upgrading the LSM within an LIS.

Testing methodology allowed for pair-wise differences to be computed for several verification metrics, with an assessment of SS and PS pair-wise difference. Overall, a large number of SS and PS pair-wise differences were observed; however, a sensitivity in which configuration was favored is dependent on verification metric, temporal aggregation, initialization time, vertical level, lead time, and threshold. In general, more PS pair-wise differences were observed in the annual, spring and summer aggregations at the surface and at 850 hPa, perhaps indicating that changes to the lower boundary conditions have the largest effects in the lowest levels. PS pair-wise differences most often showed LIS2 was a better performer; a few exceptions were noted for dew point temperature, where LIS3 was favored at 850 hPa in the spring aggregation as well as at the surface in the overnight hours during the annual, spring, and summer aggregations. When considering the GO Index, regardless of temporal aggregation or initialization hour, LIS2 outperforms LIS3. On a regional scale, most PS pair-wise differences seen for surface variables in the CONUS domain were often seen in the East domain as well, while the West domain tended to display more dissimilarity from the CONUS domain.

6. References

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3, NCAR Tech Note, NCAR/TN–475+STR, 113 pp.

Table 1. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for upper air temperature BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the CONUS verification domain.

Upper Air Temperature		Annual				Summer				Fall				Winter				Spring			
		f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1
	700	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1
	500	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--
	400	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	300	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	200	--	--	LSM_v3.3	--	--	--	LSM_v3.3	--	--	--	--	--	--	--	--	--	--	--	--	--
	150	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--
	100	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	850	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
	700	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1
	500	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--	--	LSM_v3.3
	400	--	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--	--	--
	300	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--
	200	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1
	150	--	--	LSM_v3.3	LSM_v3.3	--	--	--	--	LSM_v3.3	--	LSM_v3.3	--	--	--	--	LSM_v3.3	--	--	--	LSM_v3.3
	100	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table 2. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for upper air dew point temperature BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the CONUS verification domain.

Upper Air Dew Point Temperature		Annual				Summer				Fall				Winter				Spring			
		f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	--	--	LSM_v2.7.1	--	--	--	--	--	LSM_v2.7.1 *	LSM_v2.7.1 *	--	LSM_v2.7.1 *
	700	LSM_v2.7.1	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	LSM_v2.7.1 *	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--
	500	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	850	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v2.7.1	--	--	--	--	--	--	--	LSM_v2.7.1	LSM_v3.3	LSM_v3.3	LSM_v3.3 *	LSM_v2.7.1 *	--	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *
	700	--	LSM_v3.3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	LSM_v3.3	LSM_v3.3	--	--
	500	--	--	LSM_v3.3	--	--	--	--	LSM_v3.3 *	--	--	--	--	--	--	--	--	--	--	LSM_v3.3	--

Table 3. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for upper air wind BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the CONUS verification domain.

Upper Air Wind Speed		Annual				Summer				Fall				Winter				Spring			
		f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--
	700	--	--	--	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--
	500	--	--	--	--	--	--	--	--	--	--	--	--	--	--	LSM_v3.3	--	--	--	--	--
	400	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	300	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	200	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	150	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	100	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	850	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v2.7.1	--	--	LSM_v3.3	LSM_v3.3	--	--	LSM_v3.3	--	--	--	--	--	--	--
	700	--	LSM_v3.3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	500	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	400	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	300	--	--	--	LSM_v3.3	LSM_v3.3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	200	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	--	--	--	--	--	--	LSM_v3.3	--	--	--	--
	150	--	LSM_v2.7.1	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	100	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table 4. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for surface temperature BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Surface Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	LSM_v3.3	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1
		Summer	--	--	--	--	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v3.3	LSM_v3.3	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1
		Fall	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1
		Winter	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	LSM_v2.7.1
		Spring	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1 *
	12 UTC Initializations	Annual	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	LSM_v3.3	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	LSM_v2.7.1
		Summer	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v3.3	LSM_v3.3	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v3.3	LSM_v3.3	--
		Fall	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--
		Winter	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1
		Spring	LSM_v2.7.1	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	LSM_v2.7.1
Bias	00 UTC Initializations	Annual	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Summer	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Fall	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Winter	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Spring	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
	12 UTC Initializations	Annual	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Summer	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Fall	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Winter	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Spring	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *

Table 5. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for surface dew point temperature BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Surface Dew Point Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--
		Summer	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	LSM_v3.3
		Fall	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--
		Winter	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	--	--	--	--	--	--	--
		Spring	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	LSM_v2.7.1 *	LSM_v2.7.1	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	LSM_v2.7.1	LSM_v2.7.1 *	LSM_v2.7.1
	12 UTC Initializations	Annual	--	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3
		Summer	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3
		Fall	--	--	LSM_v2.7.1	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	--	LSM_v3.3	LSM_v3.3
		Winter	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--	--	--	--
		Spring	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1 *	LSM_v2.7.1 *	--	LSM_v3.3	LSM_v3.3	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1 *	LSM_v2.7.1 *	--	--	--	--
Bias	00 UTC Initializations	Annual	LSM_v3.3 *	--	--	--	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Summer	LSM_v3.3 *	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
		Fall	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		Winter	--	--	LSM_v3.3 *	--	--	--	--	--	--	--	--	--	--	--	--	--
		Spring	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *
	12 UTC Initializations	Annual	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *
		Summer	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *
		Fall	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		Winter	--	--	--	--	--	--	LSM_v3.3 *	LSM_v2.7.1 *	--	--	--	--	--	--	--	--
		Spring	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v2.7.1 *	LSM_v3.3 *	LSM_v3.3 *	LSM_v3.3 *

Table 6. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for surface wind BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Surface Wind Speed			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	--	--	--	--	--	--	--
		Summer	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		Fall	--	LSM_v3.3	--	--	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--
		Winter	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--
		Spring	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--
	12 UTC Initializations	Annual	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--
		Summer	--	--	LSM_v3.3	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--
		Fall	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	LSM_v2.7.1	--	--	--	--	--	--
		Winter	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--	--	--	--	--	--	--
		Spring	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	--	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--
Bias	00 UTC Initializations	Annual	LSM_v2.7.1	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3
		Summer	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3
		Fall	LSM_v2.7.1	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3
		Winter	LSM_v2.7.1	--	LSM_v3.3	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--
		Spring	--	LSM_v3.3	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3
	12 UTC Initializations	Annual	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--
		Summer	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--
		Fall	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	LSM_v3.3	--	--	--
		Winter	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	LSM_v3.3	--	LSM_v3.3
		Spring	LSM_v3.3	LSM_v2.7.1	LSM_v3.3	LSM_v3.3	--	--	--	--	--	LSM_v3.3	LSM_v3.3	LSM_v3.3	LSM_v3.3	--	--	--

Table 7. SS differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for 3-hour QPF GSS by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

3-hour QPF			00 UTC Initializations									12 UTC Initializations								
			>0.01	>0.02	>0.05	>0.1	>0.15	>0.25	>0.35	>0.5	>1	>0.01	>0.02	>0.05	>0.1	>0.15	>0.25	>0.35	>0.5	>1
Gilbert Skill Score	Annual	f12	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	LSM_v2.7.1	--
		f36	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--	--	LSM_v2.7.1
		f36	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	LSM_v3.3	--	--	--	--	--	--	--	--	--
	Fall	f12	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	LSM_v3.3
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	LSM_v2.7.1	--
		f48	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	LSM_v2.7.1	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1
		f24	--	--	--	--	--	--	--	--	LSM_v3.3	--	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--
		f36	--	--	--	--	--	--	--	--	--	--	--	LSM_v3.3	--	--	LSM_v2.7.1	--	LSM_v2.7.1	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Spring	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	LSM_v2.7.1	LSM_v2.7.1	--	LSM_v2.7.1	--
		f36	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	LSM_v2.7.1

Table 8. SS differences for the AFWA configuration run with WRF v3.4 and LIS input using either Noah v2.7.1 or v3.3 (where the highlighted configuration is favored) for 24-hour QPF GSS by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Daily QPF				>0.01	>0.25	>0.5	>0.75	>1	>1.25	>1.5	>2	>3
Gilbert Skill Score	00 UTC Initializations	Annual	f36	LSM_v2.7.1	--	--	--	--	--	--	--	--
		Summer	f36	LSM_v2.7.1	--	--	--	--	--	--	--	--
		Fall	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
		Spring	f36	LSM_v2.7.1	--	--	--	--	--	--	--	--
	12 UTC Initializations	Annual	f24	LSM_v2.7.1	--	--	LSM_v2.7.1	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Summer	f24	--	--	--	LSM_v2.7.1	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Fall	f24	--	--	--	--	--	--	LSM_v2.7.1	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	LSM_v2.7.1	--	--
		Spring	f24	LSM_v2.7.1	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--

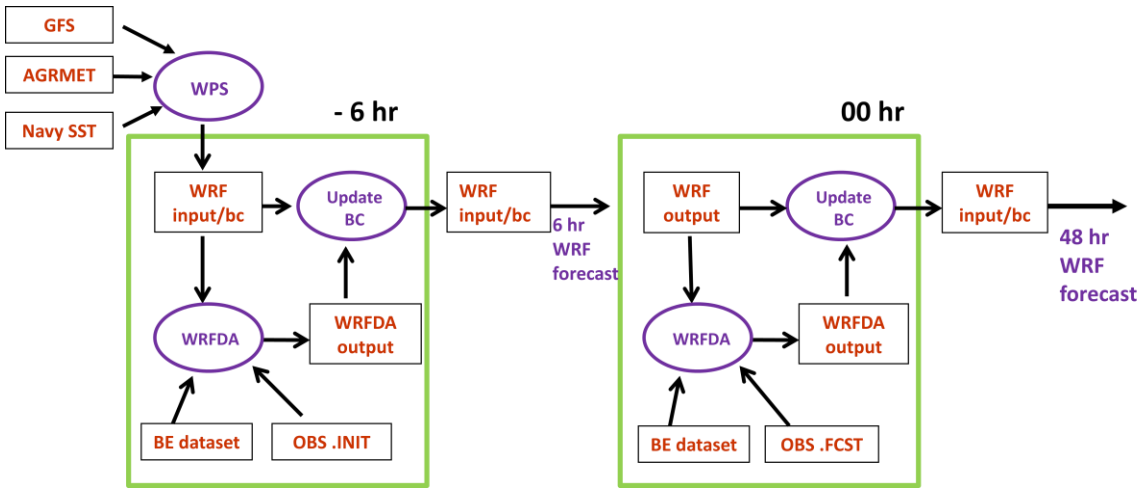


Figure 1. Overview of 6-hr “warm start” spin-up.

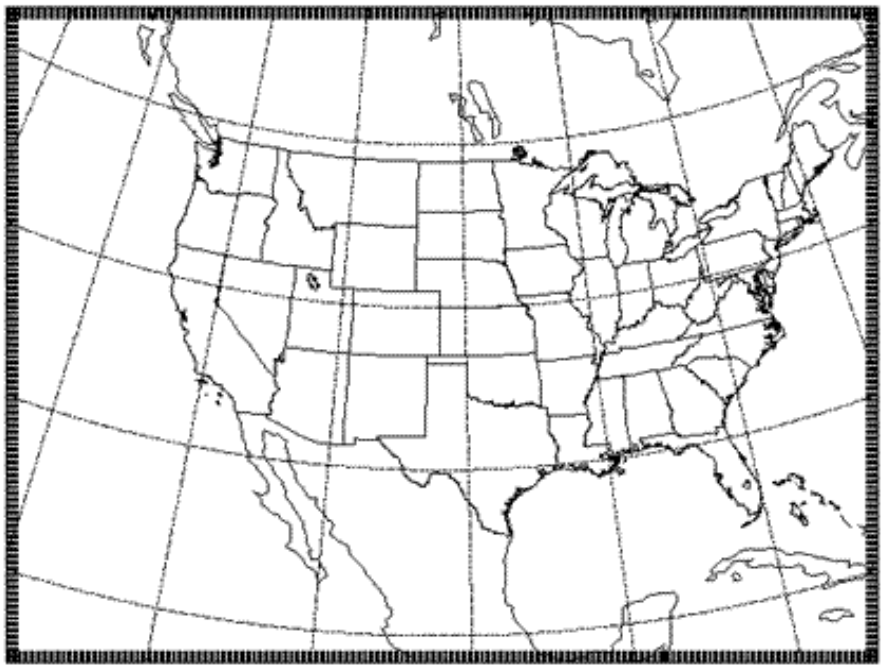


Figure 2. WRF-ARW computational domain.

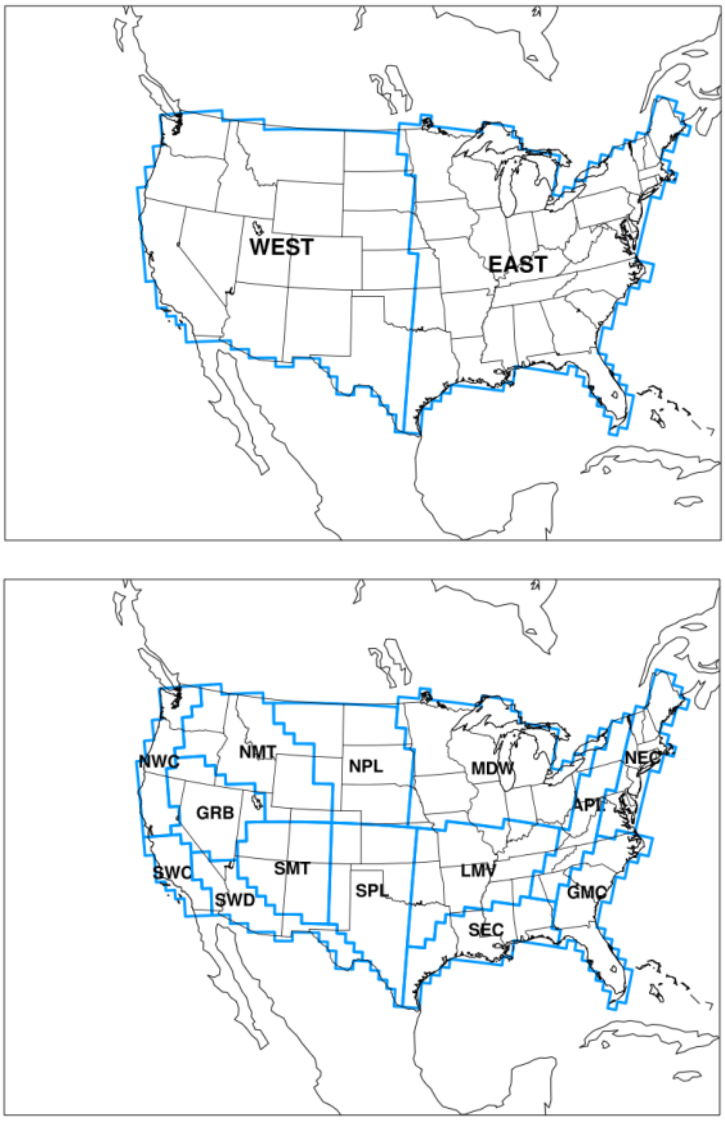
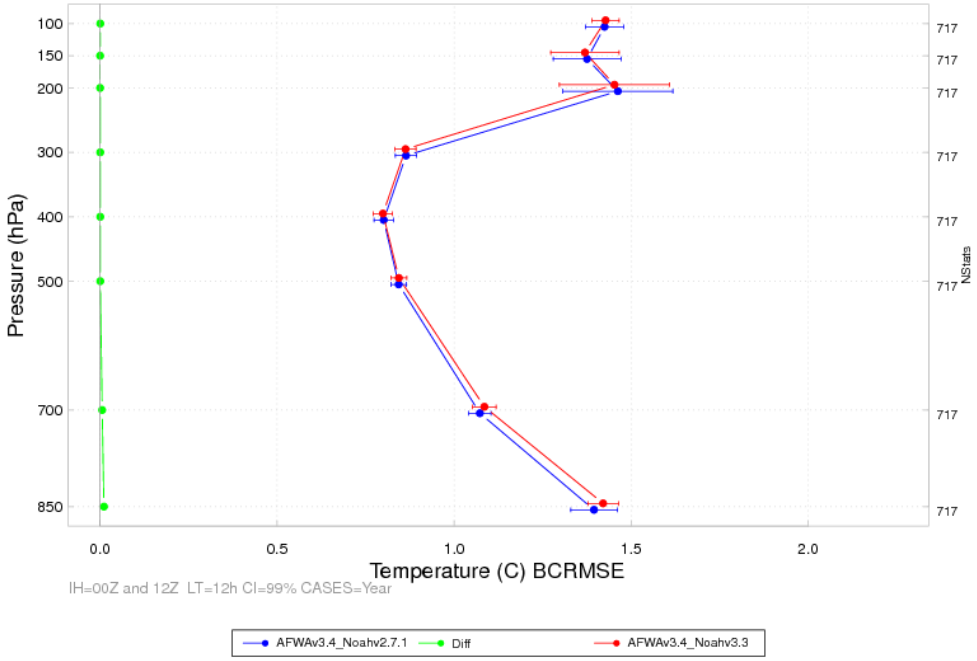


Figure 3. Map showing the locations of the CONUS-West, CONUS-East (top) and 14 regional verification domains (bottom). The outermost outline of the regional domains depict the CONUS verification domain.

(a) LT=12 h



(b) LT=48 h

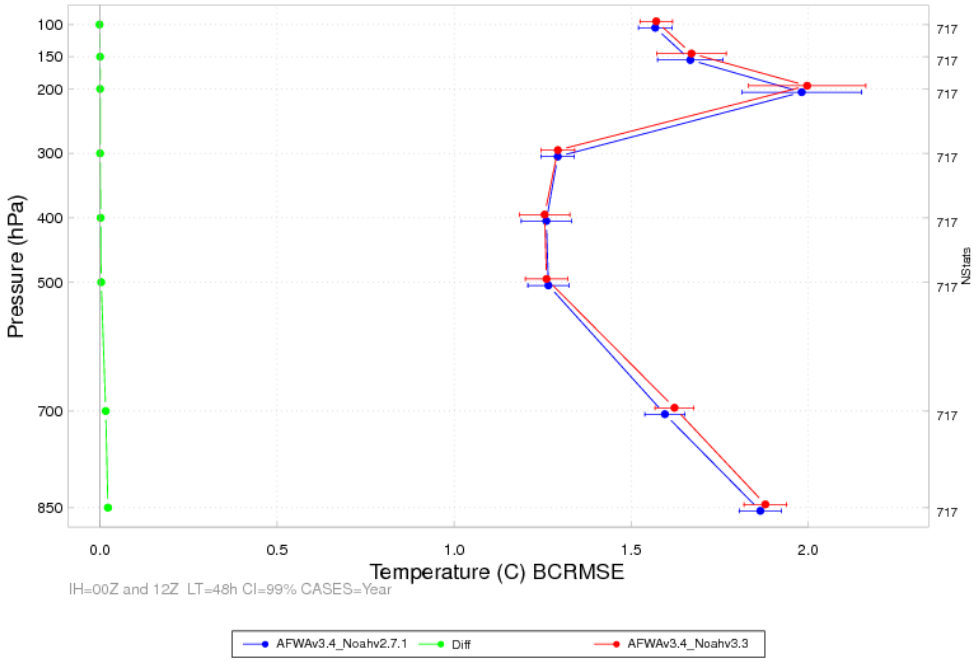
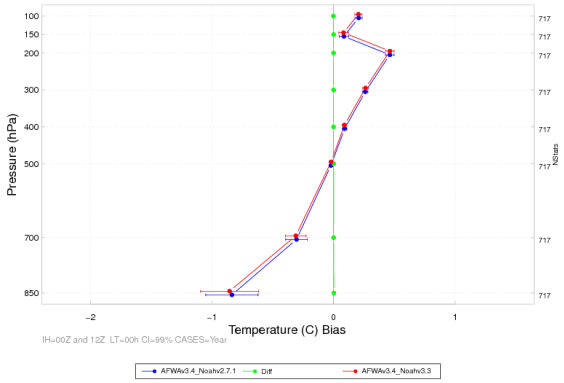
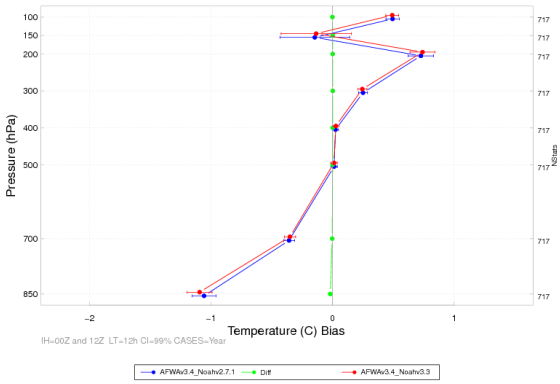


Figure 4. Vertical profile of the median BCRMSE for temperature (°C) for the full integration domain aggregated across the entire year of cases for the (a) 12- and (b) 48-h lead times. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The horizontal bars attached to the median represent the 99% CIs.

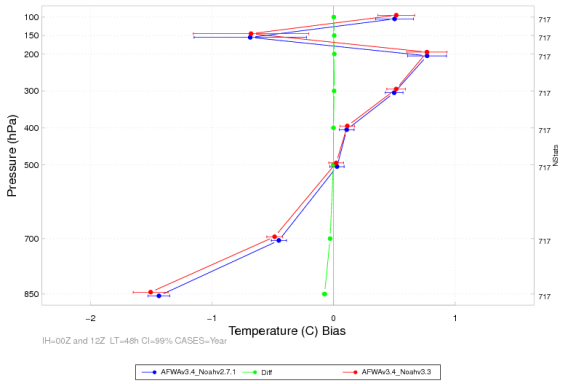
(a) Annual LT=00 h



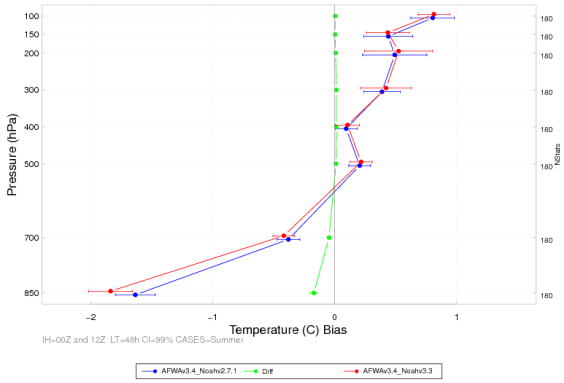
(b) Annual LT=12 h



(c) Annual LT=48 h



(d) Summer LT=48 h



(e) Winter LT=48 h

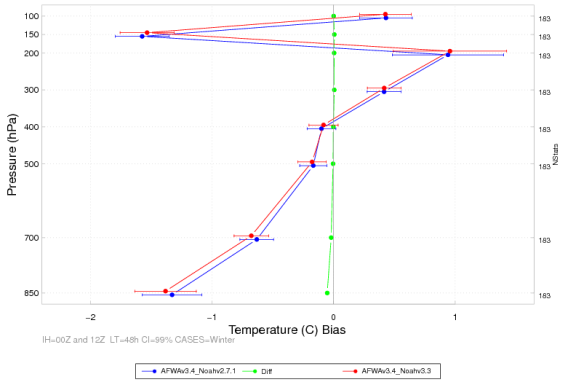
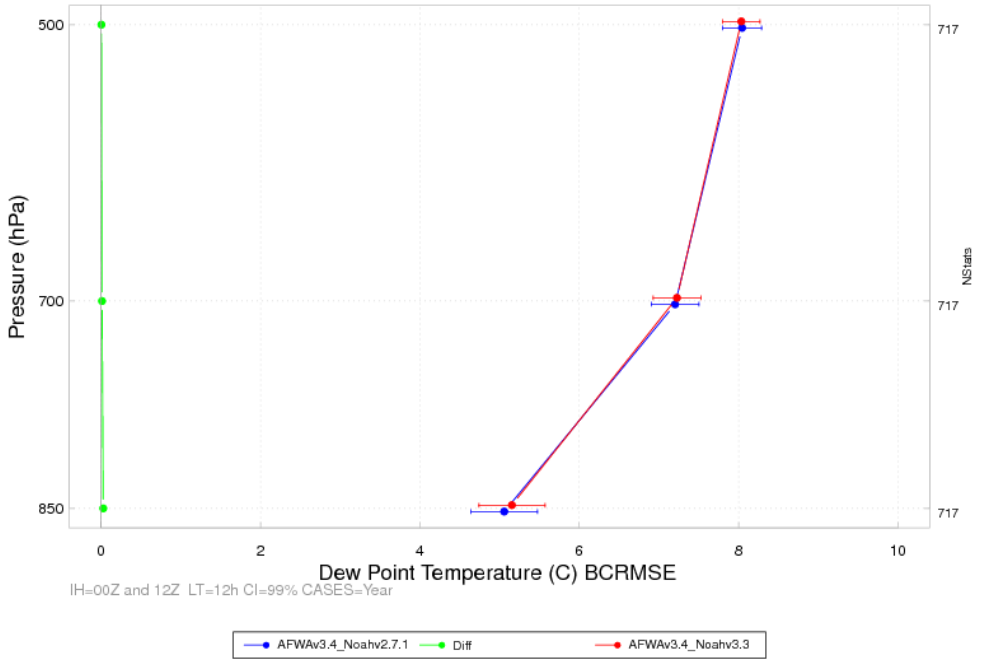


Figure 5. Vertical profile of the median bias for temperature (°C) for the full integration domain aggregated across the entire year of cases for the (a) initialization time and (b) 12- and (c) 48-h lead times and for 48-h lead time for the (d) summer aggregation and (e) winter aggregation. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) LT=12 h



(b) LT=48 h

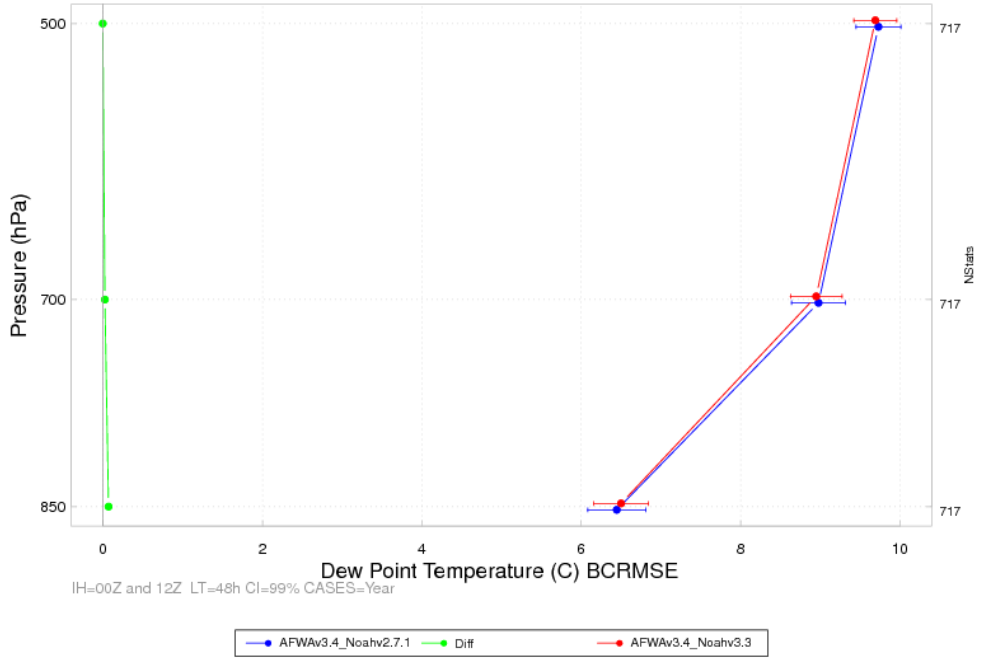
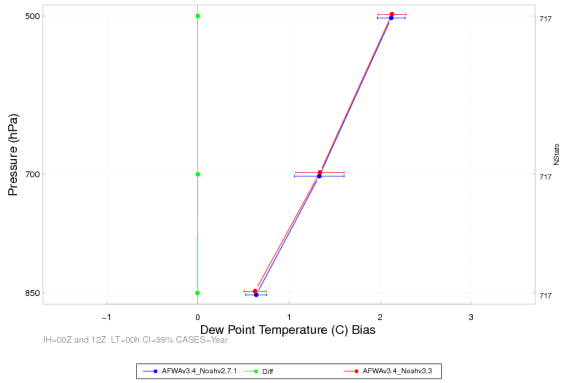
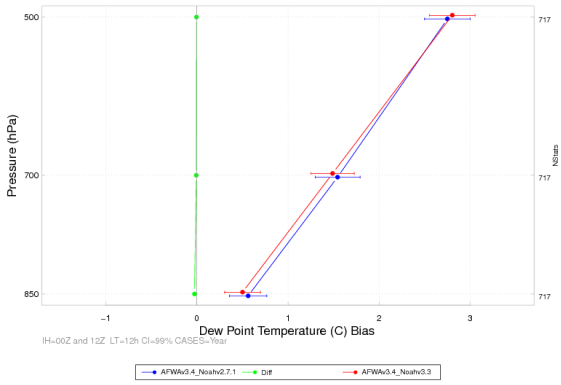


Figure 6. Vertical profile of the median BCRMSE for dew point temperature (°C) for the full integration domain aggregated across the entire year of cases for the (a) 12- and (b) 48-h lead times. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The horizontal bars attached to the median represent the 99% CIs.

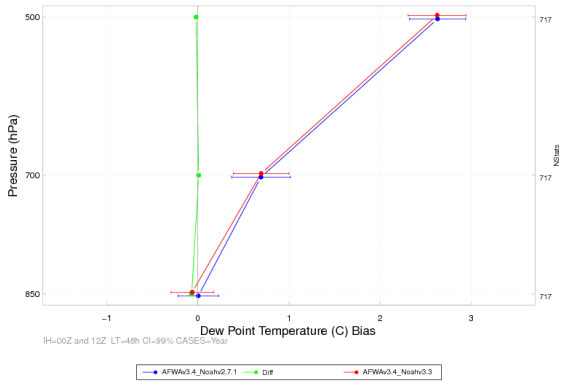
(a) Annual LT=00 h



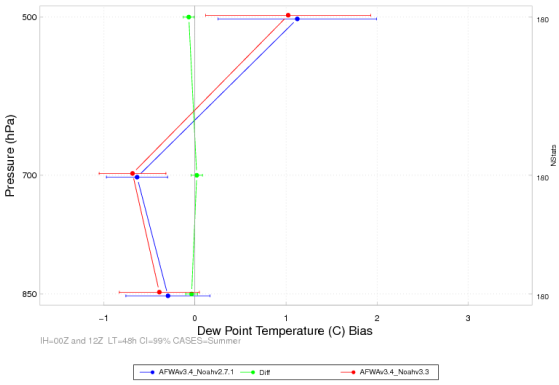
(b) Annual LT=12 h



(c) Annual LT=48 h



(d) Summer LT=48 h



(e) Winter LT=48 h

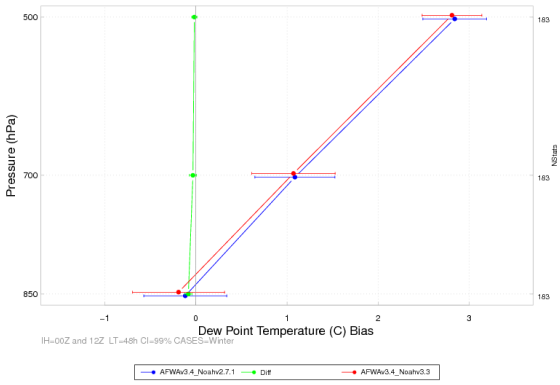
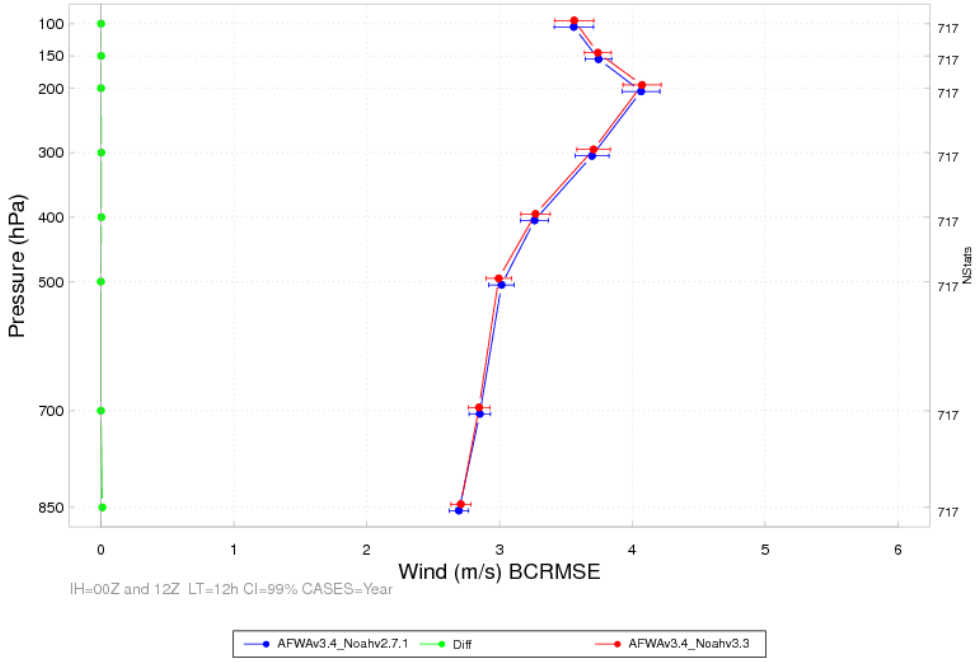


Figure 7. Vertical profile of the median bias for dew point temperature (°C) for the full integration domain aggregated across the entire year of cases for the (a) initialization time and (b) 12- and (c) 48-h lead times and for 48-h lead time for the (d) summer aggregation and (e) winter aggregation. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) LT=12 h



(b) LT=48 h

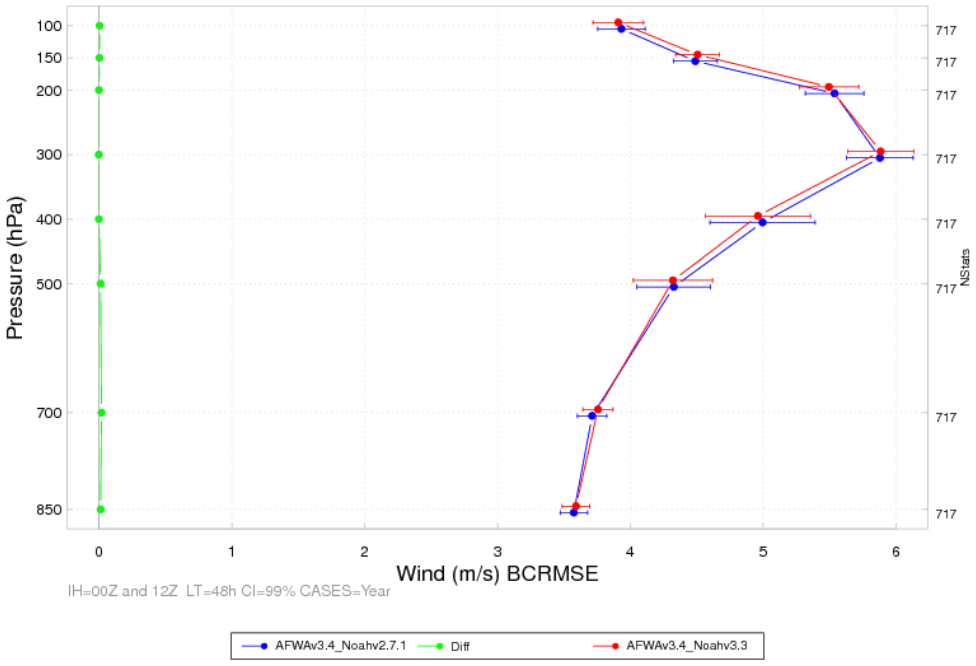


Figure 8. Vertical profile of the median BCRMSE for wind speed (m s^{-1}) for the full integration domain aggregated across the entire year of cases for the (a) 12- and (b) 48-h lead times. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The horizontal bars attached to the median represent the 99% CIs.

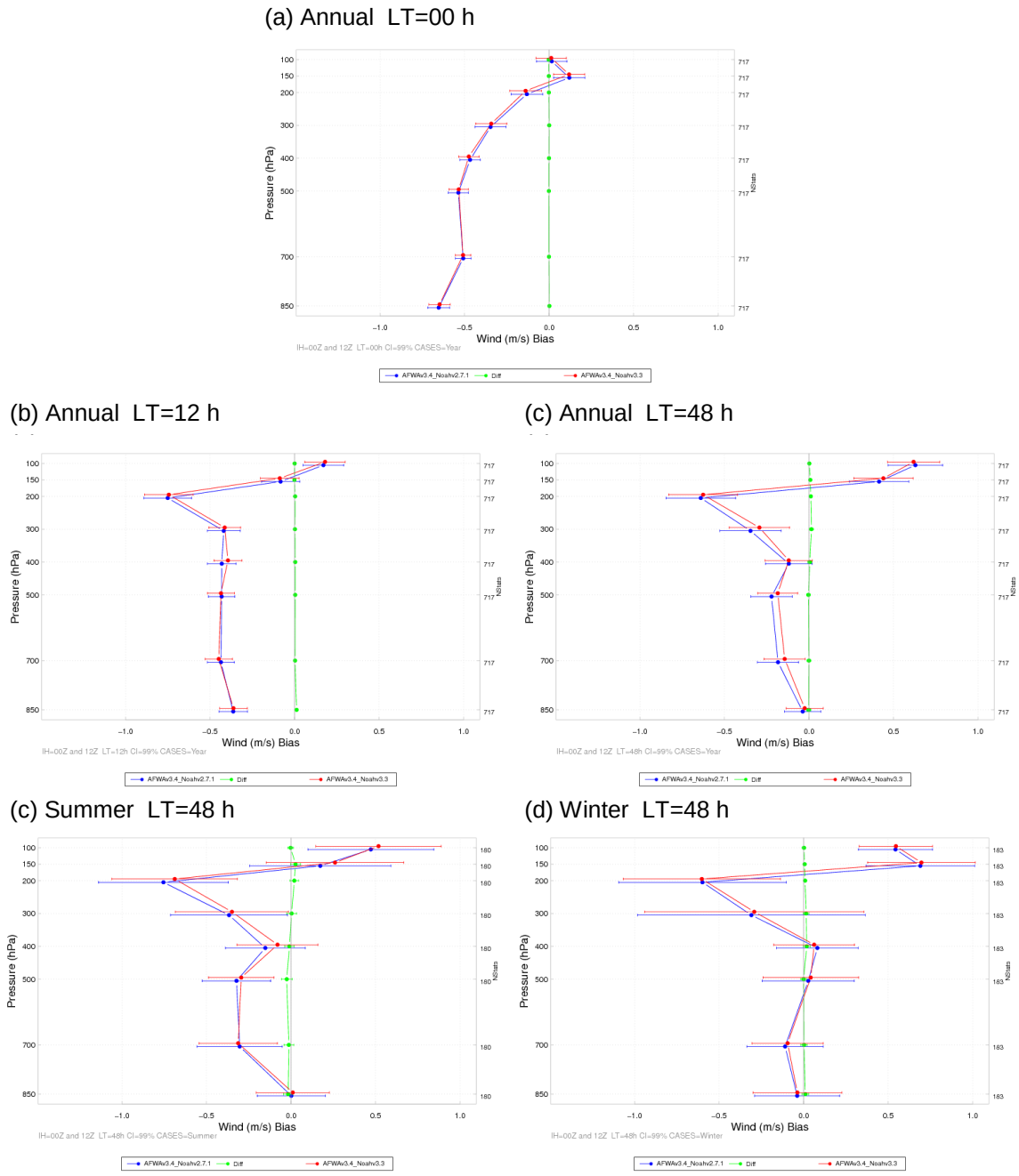
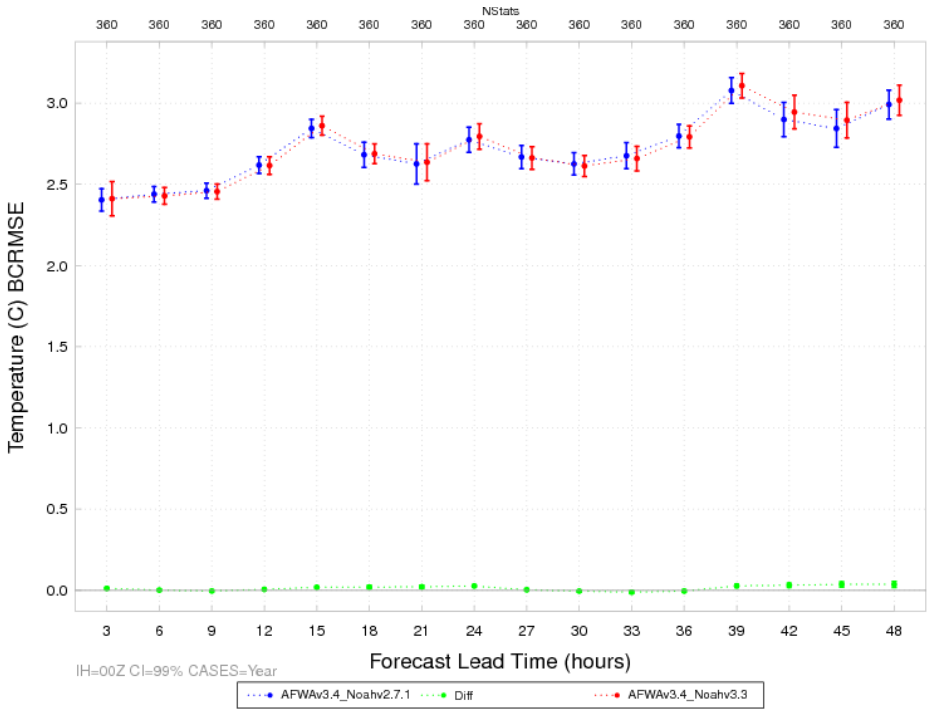


Figure 9. Vertical profile of the median bias for wind speed (m s^{-1}) for the full integration domain aggregated across the entire year of cases for the (a) initialization time and (b) 12- and (c) 48-h lead times and for 48-h lead time for the (d) summer aggregation and (e) winter aggregation. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) IH=00 UTC



(b) IH=12 UTC

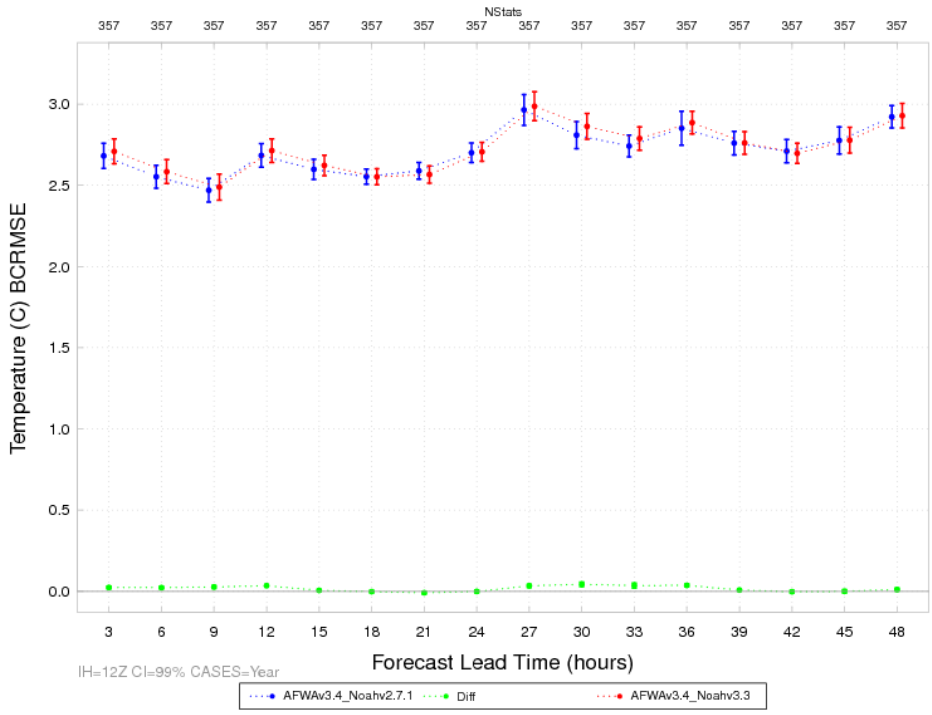
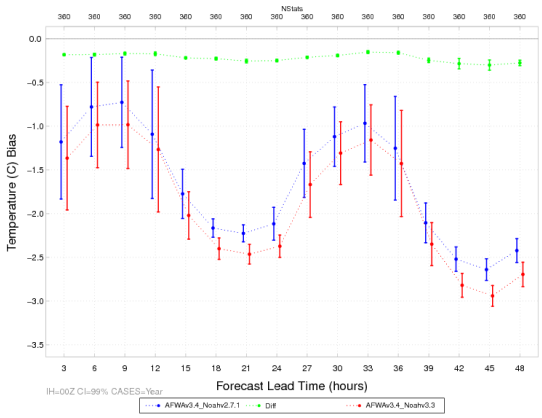
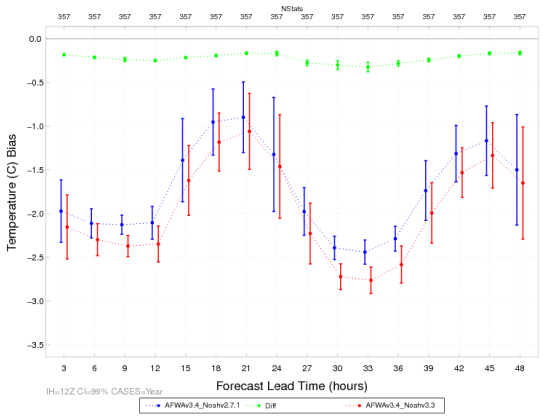


Figure 10. Time series plot of 2 m AGL temperature (°C) for median BCRMSE for the (a) 00 UTC initializations and (b) 12 UTC initializations aggregated across the entire year of cases. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

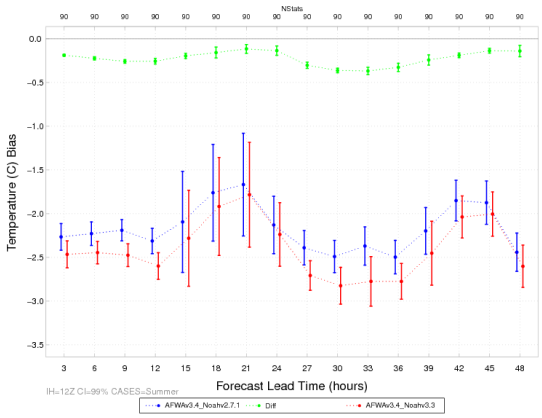
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=12 UTC



(d) Winter IH=12 UTC

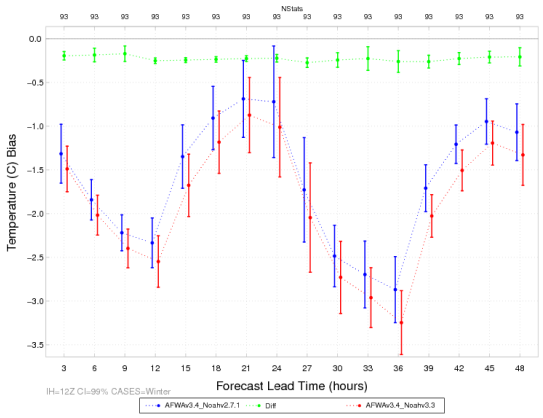
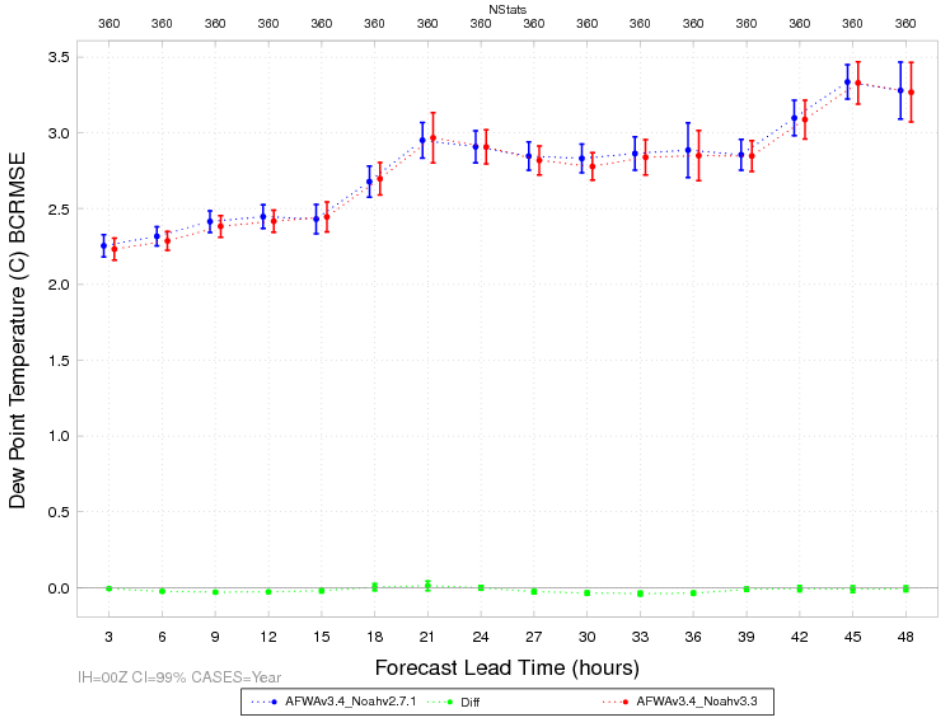


Figure 11. Time series plot of 2 m AGL temperature (°C) for median bias for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 12 UTC initializations for the (c) summer aggregation and (d) winter aggregation. LIS2 is in blue, LIS3 in red, and the differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

(a) IH=00 UTC



(b) IH=12 UTC

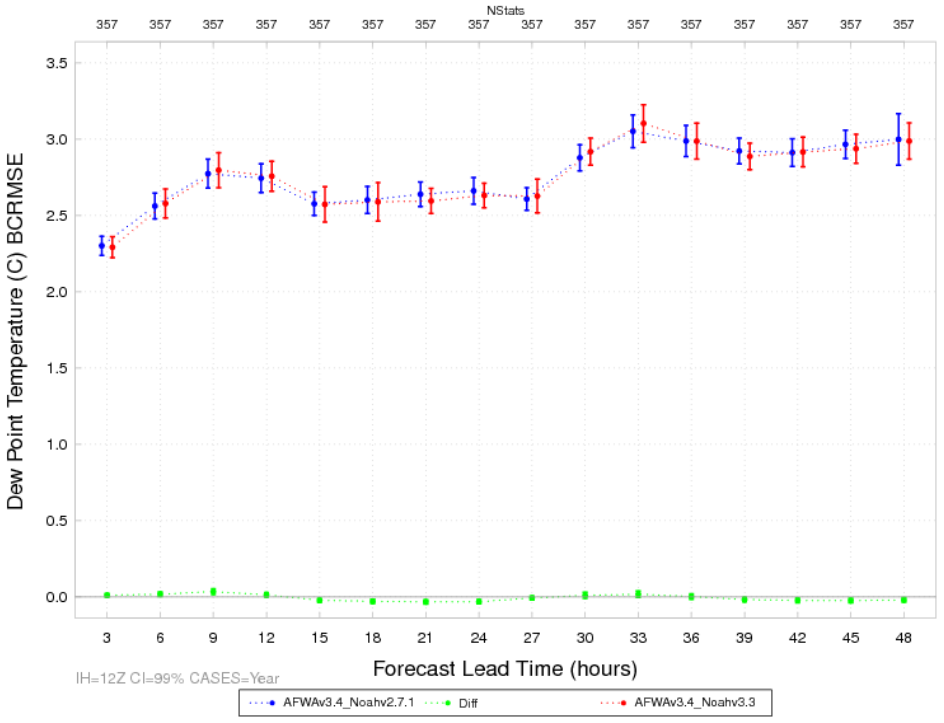
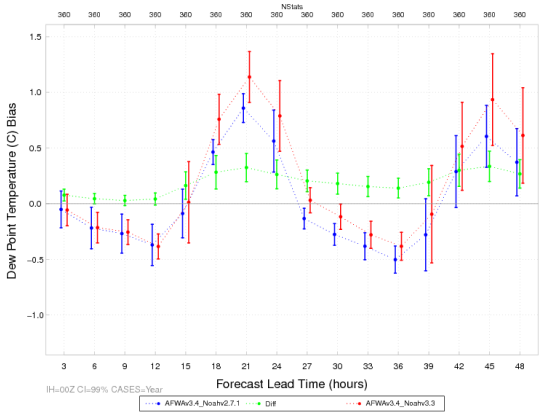
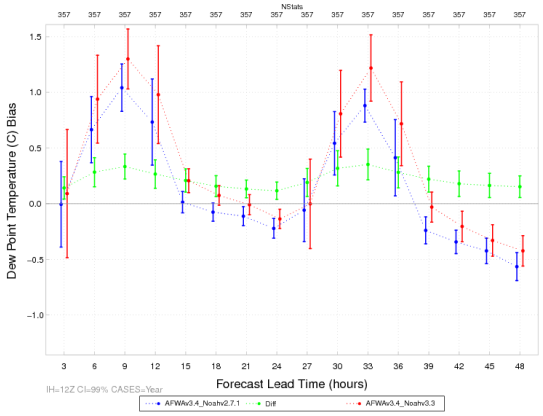


Figure 12. Time series plot of 2 m AGL dew point temperature (°C) for median BCRMSE for the (a) 00 UTC initializations and (b) 12 UTC initializations aggregated across the entire year of cases. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

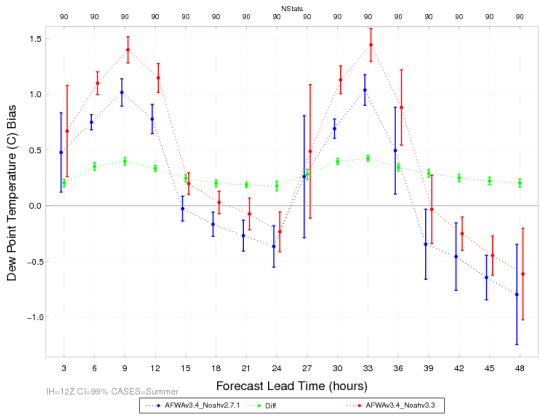
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=12 UTC



(d) Winter IH=12 UTC

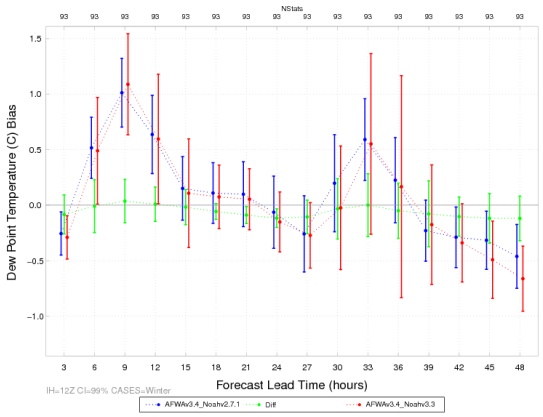
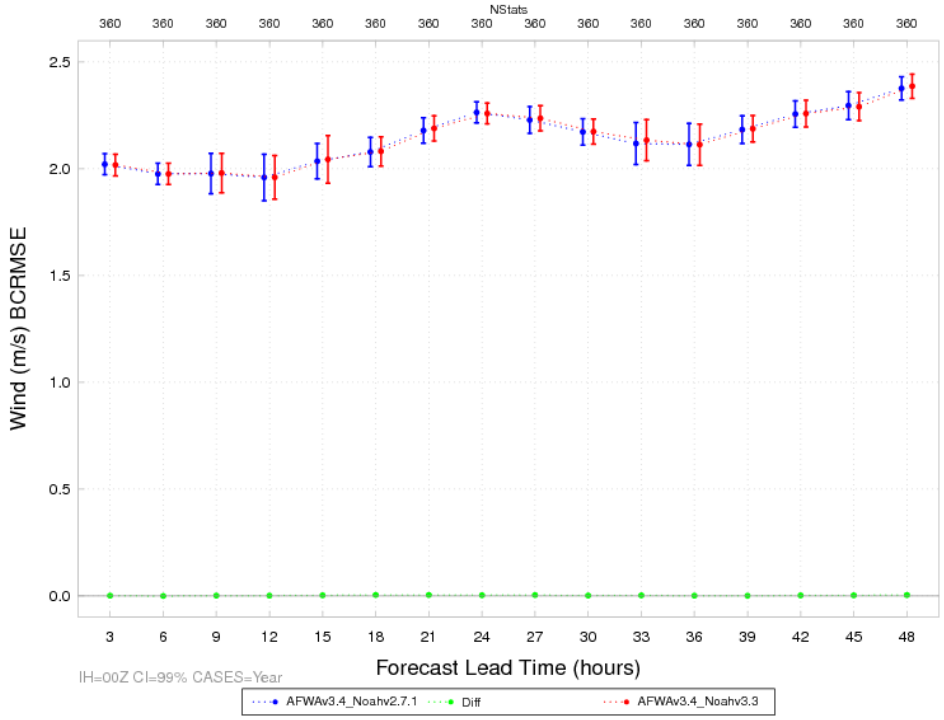


Figure 13. Time series plot of 2 m AGL dew point temperature (°C) for median bias for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 12 UTC initializations for the (c) summer aggregation and (d) winter aggregation. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

(a) IH=00 UTC



(b) IH=12 UTC

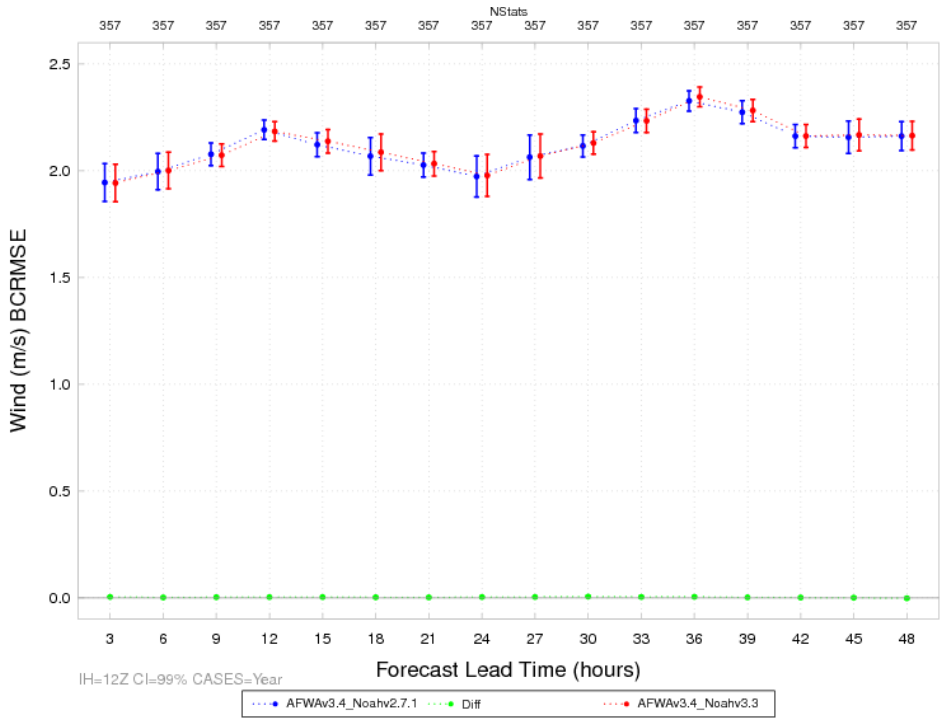
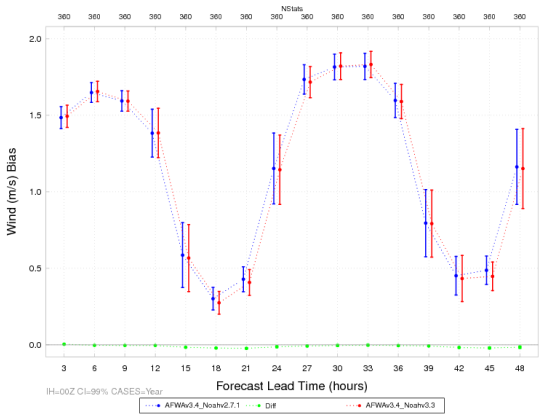
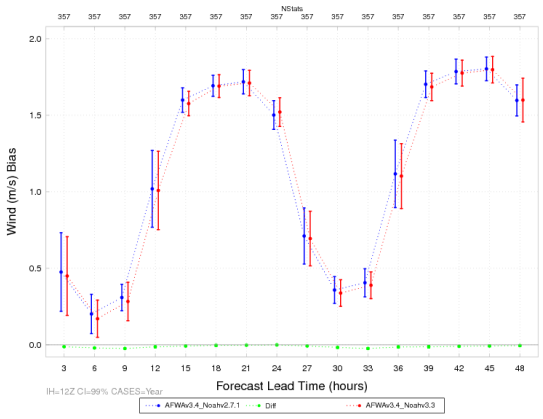


Figure 14. Time series plot of 10 m AGL wind speed (m s^{-1}) for median BCRMSE for the 12 UTC initializations (a) aggregated across the entire year of cases and (b) aggregated across the summer season. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

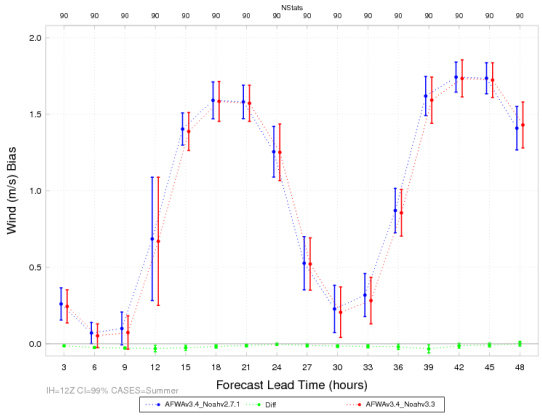
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=00 UTC



(d) Winter IH=00 UTC

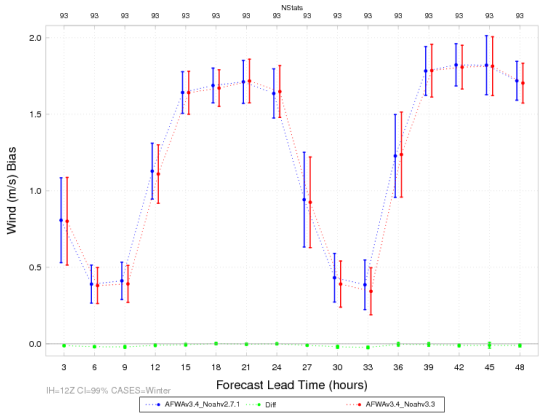
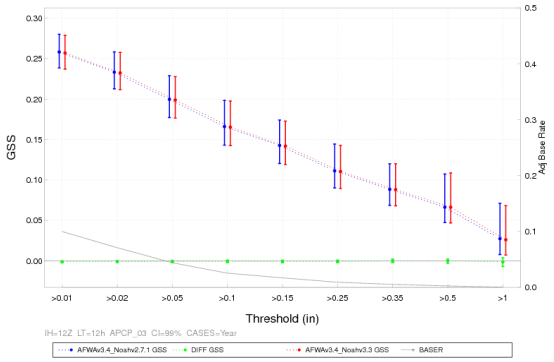
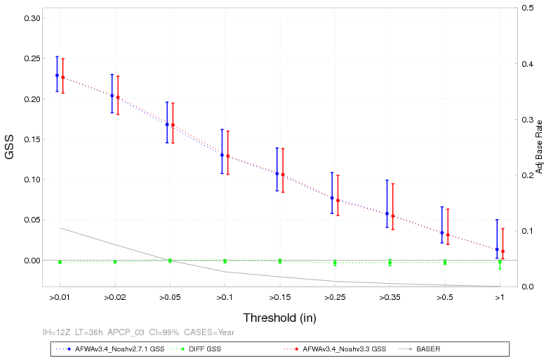


Figure 15. Time series plot of 2 m AGL wind speed (m s^{-1}) for median bias for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. LIS2 is in blue, LIS3 in red, and the differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

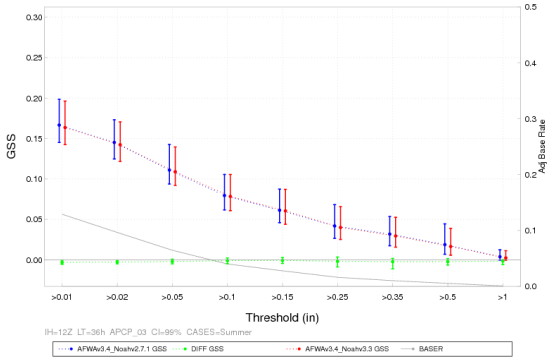
(a) Annual LT=12 h



(b) Annual LT=36 h



(c) Summer LT=12 h



(d) Winter LT=12 h

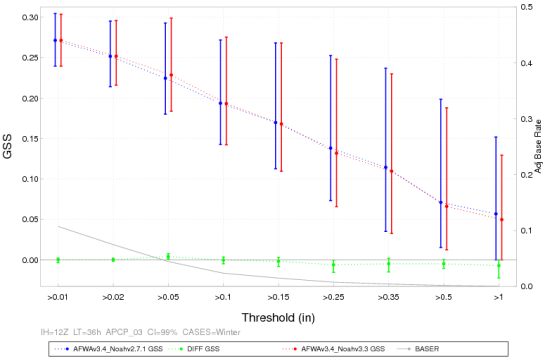
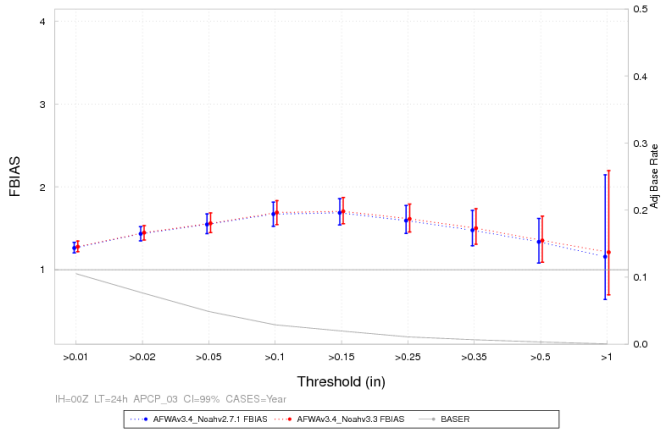
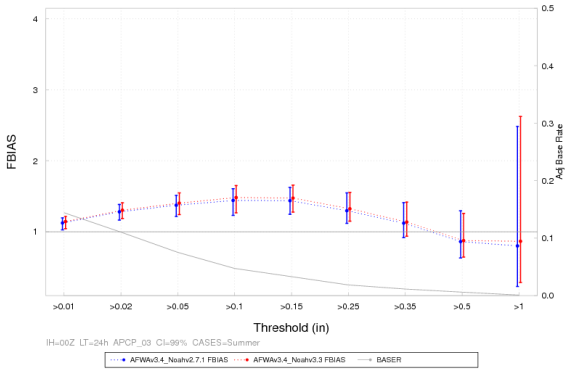


Figure 16. Threshold series plot of 3-h accumulated precipitation (in) for median GSS for the 12 UTC initializations aggregated across the entire year of cases for the (a) 12-h lead time and the (b) 36-h lead time and for the 12 UTC initializations for the 36-h lead time for the (c) summer aggregation and (d) winter aggregation. LIS2 is in blue, LIS3 in red, and the pair-wise differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

(a) Annual



(b) Summer



(c) Winter

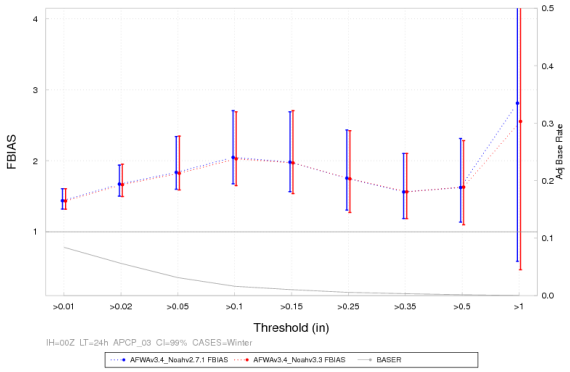
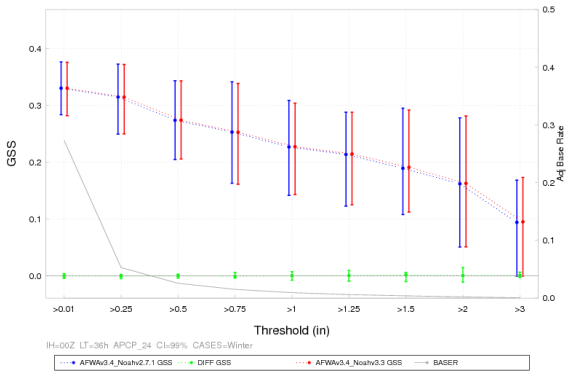
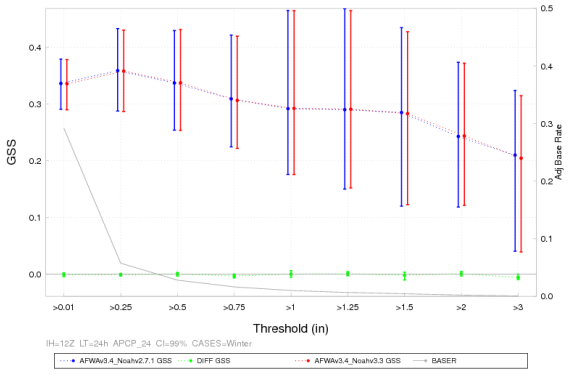


Figure 17. Threshold series plot of 3-h accumulated precipitation (in) for median frequency bias for the 00 UTC initialization for the 24-h lead time aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. LIS2 is in blue, LIS3 in red, and the differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

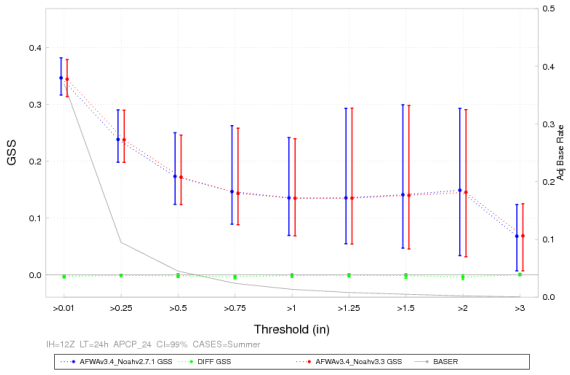
(a) Annual IH=00 UTC LT=36 h



(b) Annual IH=12 UTC LT=24 h



(c) Summer IH=12 UTC LT=24 h



(d) Winter IH=12 UTC LT=24 h

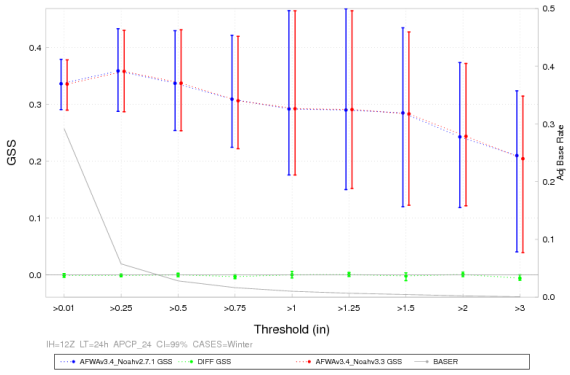
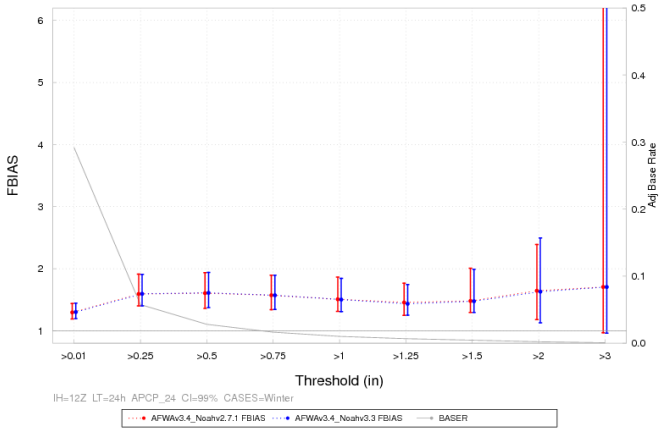
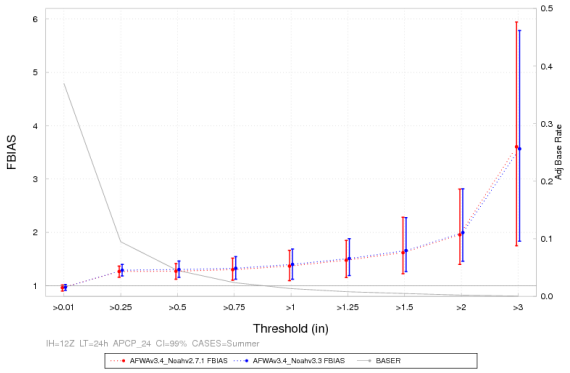


Figure 18. Threshold series plot of 24-h accumulated precipitation (in) for median GSS for the (a) 00 UTC for the 36-h lead time aggregated across the entire year of cases, the 12 UTC initialization for the 24-h lead time aggregated across the (b) entire year of cases, (c) summer aggregation, and (d) winter aggregation. LIS2 is in blue, LIS3 in red, and the differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

(a) Annual



(b) Summer



(c) Winter

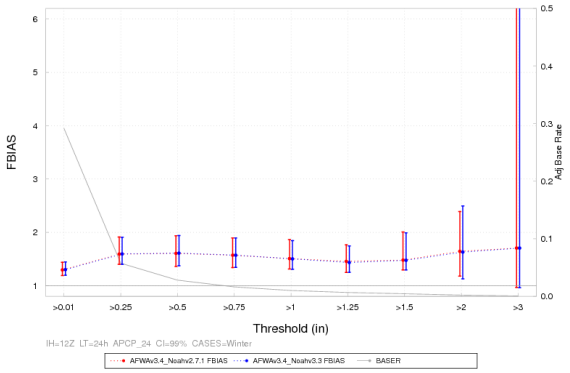


Figure 19. Threshold series plot of 24-h accumulated precipitation (in) for median frequency bias for the 12 UTC initializations for the 24-h lead time aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. LIS2 is in blue, LIS3 in red, and the differences (LIS3-LIS2) in green. The vertical bars attached to the median represent the 99% CIs.

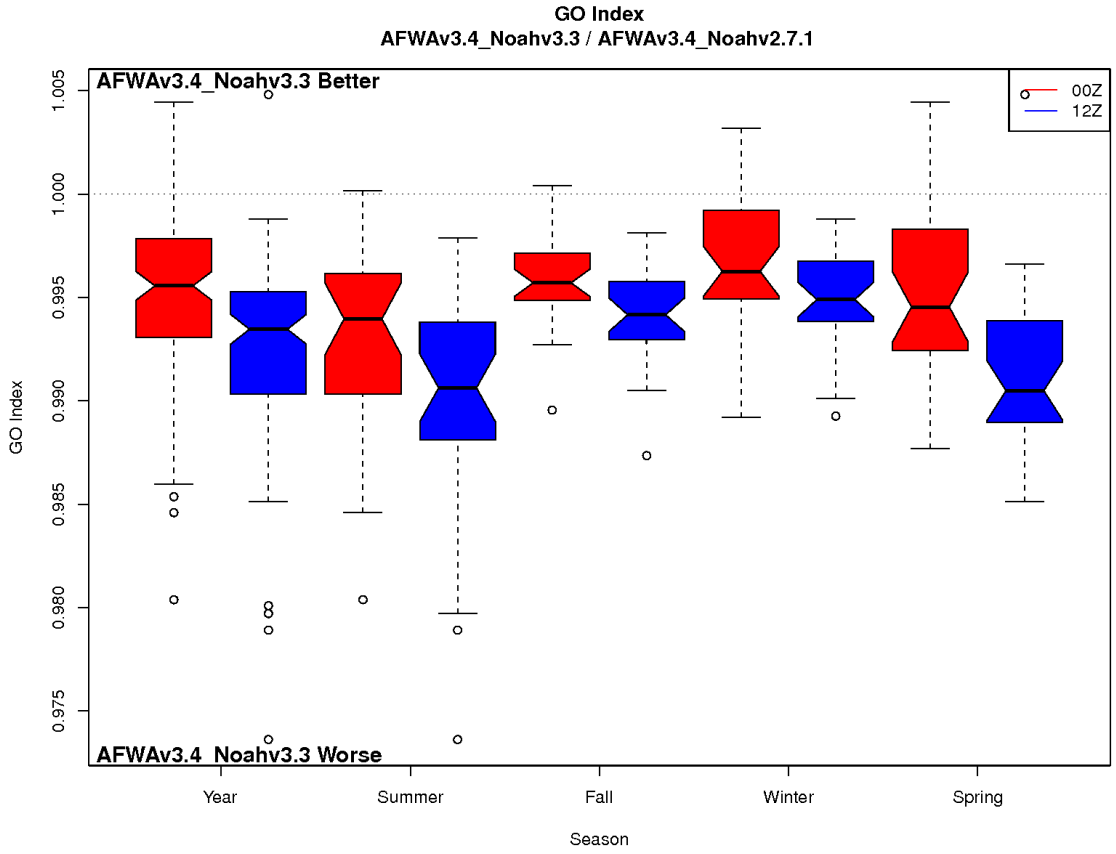


Figure 20. Boxplot of GO Index values aggregated across the entire year of cases and for all seasons, stratified by initialization time where 00 UTC is in red and 12 UTC is in blue. The median value is the thick black line located at the vertex of the notches, the notches around the median is an approximation of the 95% confidence about the median, the whiskers, denoted by the black, dashed lines, denote the largest values that are not outliers, and the circles represent the outliers.

Appendix A: Case list. Dates in bold were not included in the verification due to bad or missing input data.

00 UTC Initialization	12 UTC Initialization
July 2011: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	July 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
August 2011: 3, 6, 9, 12, 15, 18, 21, **24**, 27, 30	August 2011: **1**, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
September 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	September 2011: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
October 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	October 2011: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
November 2011: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28	November 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
December 2011: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	December 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
January 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	January 2012: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
February 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	February 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27
March 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	March 2012: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
April 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	April 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
May 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	May 2012: **3**, **6**, 9, 12, 15, 18, 21, 24, 27, 30
June 2012: 1, **4**, 7, 10, 13, 16, 19, 22, 25, 28	June 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29

Appendix B: Subset a WRF ***namelist.input*** used in this test

```
&wrfva4
thin_conv       = .TRUE.,
use_synopobs    = .TRUE.,
use_shipsobs    = .TRUE.,
use_metarobs    = .TRUE.,
use_soundobs    = .TRUE.,
use_mtgirsobs   = .TRUE.,
use_tamdarobs   = .TRUE.,
use_pilotobs    = .TRUE.,
use_airepobs    = .TRUE.,
use_geoamvobs   = .TRUE.,
use_polaramvobs  = .TRUE.,
use_buoyobs     = .TRUE.,
use_profilerobs   = .TRUE.,
use_satemobs    = .TRUE.,
use_gpspwobs    = .TRUE.,
use_gpsrefobs   = .TRUE.,
top_km_gpsro    = 30.0,
bot_km_gpsro    = 0.0,
use_ssmiretrievalobs = .TRUE.,
use_qscatobs    = .TRUE.,

&wrfvar6
max_ext_its     = 2,
ntmax           = 200,
nsave           = 4,
write_interval = 5,
eps             = 1.E-02,
```

```
precondition_cg = .FALSE.,
precondition_factor = 1.0,
use_lanczos    = .FALSE.,
orthonorm_gradient = .FALSE.,

&time_control
run_hours                           = 48,
interval_seconds                    = 10800,
history_interval                    = 180,
frames_per_outfile                  = 1,
restart                             = .false.,
io_form_history                     = 2,
input_outname                       = "wrfinput_d<domain>_<date>",
/

&domains
time_step                           = 90,
time_step_fract_num                 = 0,
time_step_fract_den                 = 1,
max_dom                             = 1,
e_we                                = 403,
e_sn                                = 302,
e_vert                              = 57,
num_metgrid_levels                  = 27,
num_metgrid_soil_levels             = 4,
dx                                  = 15000,
dy                                  = 15000,
p_top_requested                     = 1000,
interp_type                         = 1,
lowest_lvl_from_sfc                 = .false.,
lagrange_order                      = 1,
force_sfc_in_vinterp                = 6,
zap_close_levels                    = 500,
adjust_heights                      = .false.,
eta_levels                          = 1.000, 0.997, 0.992, 0.985, 0.978, 0.969, 0.960, 0.950,
                                      0.938, 0.925, 0.910, 0.894, 0.876, 0.857, 0.835, 0.812,
                                      0.787, 0.760, 0.731, 0.700, 0.668, 0.635, 0.600, 0.565,
                                      0.530, 0.494, 0.458, 0.423, 0.388, 0.355, 0.323, 0.293,
                                      0.264, 0.237, 0.212, 0.188, 0.167, 0.147, 0.130, 0.114,
                                      0.099, 0.086, 0.074, 0.064, 0.054, 0.046, 0.039, 0.032,
                                      0.027, 0.022, 0.017, 0.013, 0.010, 0.007, 0.004, 0.002,
                                      0.000,
/

&physics
mp_physcis                          = 4,
ra_lw_physics                       = 1,
ra_sw_physics                       = 1,
radt                                = 30,
sf_sfclay_physics                   = 1,
sf_surface_physics                  = 2,
bl_pbl_physics                      = 1,
bldt                                = 0,
cu_physics                          = 1,
cudt                                = 5,
surface_input_source                = 1,
```

num_soil_layers = 4,
num_land_cat = 28,
mp_zero_out = 2,
/

&dynamics
rk_ord = 3,
diff_6th_opt = 2,
diff_6th_factor = 0.10
w_damping = 1,
diff_opt = 1,
km_opt = 4,
damp_opt = 3,
zdamp = 5000.,
dampcoef = 0.05
khdif = 0,
kvdif = 0,
smdiv = 0.1,
emdiv = 0.01,
epssm = 0.1,
time_step_sound = 0,
h_mom_adv_order = 5,
v_mom_adv_order = 3,
h_sca_adv_order = 5,
v_sca_adv_order = 3,
moist_adv_opt = 1,
scalar_adv_opt = 0,
chem._adv_opt = 0,
tke_adv_opt = 0,
/

&bdy_control
spec_bdy_width = 5,
spec_zone = 1,
relax_zone = 4,
specified = .true.,
/