**The Developmental Testbed Center's
final report on the performance of the Air Force's operational
configuration for version 3.5.1 and proposed configuration for
version 3.8.1 using the Weather Research and Forecasting model**

12 April 2017

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. The Air Force (AF) was interested in assessing an updated configuration using recommended specifications from NCAR's MMM division, including, among other things, the Thompson aerosol-aware microphysics scheme (Thompson and Eidhammer, 2014) and the updated Rapid Radiative Transfer Model (RRTMG; Iacono et al. 2008) long-wave and short-wave radiation schemes, with the latest release of WRF v3.8.1 for possible implementation as an operational configuration (OC) in the future. To address this request, the Developmental Testbed Center (DTC) performed a rigorous test and evaluation which assessed forecast performance between the configuration similar to AF operations using v3.5.1 (previously run by the DTC) and the replacement configuration using v3.8.1 with the Advanced Research WRF (ARW; Skamarock et al. 2008) dynamic core. With other various differences in physics and dynamics between the two configurations, they will simply be referred to as AFv3.5.1 and AFv3.8.1, with AFv3.5.1 used as the baseline.

2. Experiment Design

For this test, the end-to-end forecast system consisted of the WRF Preprocessing System (WPS), 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 is available for dissemination to the user community upon request. The codes utilized were based on the official released versions of WPS (v3.5.1/v3.8.1), WRF (v3.5.1/v3.8.1), UPP (v2.1/v3.1), and MET (v4.1/v5.2). 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 for summer months of 1 July 2011 through 30 September 2011 and winter months of 1 January 2012 through 31 March 2012, consequently creating initialization times including both 00 and 12 UTC, for a total of 123 cases. The forecasts were run out to 48 hours with output files generated every 3 hours.

The table below lists 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 116 cases ran to completion and were used in the verification results.

Missing cycles:

Affected Cycle	Missing data	Reason
2011071712	WRF output	Missing SST input data
2011080112	WRF output	Missing SST input data
2011082400	WRF output	Missing SST input data
2012011012	WRF output	Missing GFS input data
2012011612	WRF output	Missing GFS input data
2012012212	WRF output	Missing GFS input data
2011072500	Verification output	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). 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 lower boundary conditions (topography, soil, vegetation type, etc.) were derived from United States Geological Survey (USGS) input data and were generated through the *geogrid* program of WPS. For AFv3.8.1, the namelist option *use_aero_icbc* was set to true, which when using the Thompson and Eidhammer scheme activates the use of an auxiliary aerosol climatology file that included mass mixing ratios of sulfates, sea salts, organic carbon, dust, and black carbon derived from 7-year global model simulation (Colarco 2010) with 0.5° longitude by 1.25° latitude spacing. The auxiliary file was used in the *metgrid* program of WPS, where the output was then used during the execution of WRF.

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 (*Figure 1*) was selected such that it covers complex terrain, plains, and coastal regions spanning from Central America, north, to northern Canada in order to capture diverse regional effects for worldwide comparability. The domain was 656 x 464 gridpoints, for a total of 304,384 horizontal 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. Note that the vertical levels specified in the WRF namelist are different between the two configurations. Relevant portions of the *namelist.input* file for

v3.8.1 are described in *Appendix A* and *Appendix B* highlights the differences in the *namelist.input* file for v3.5.1.
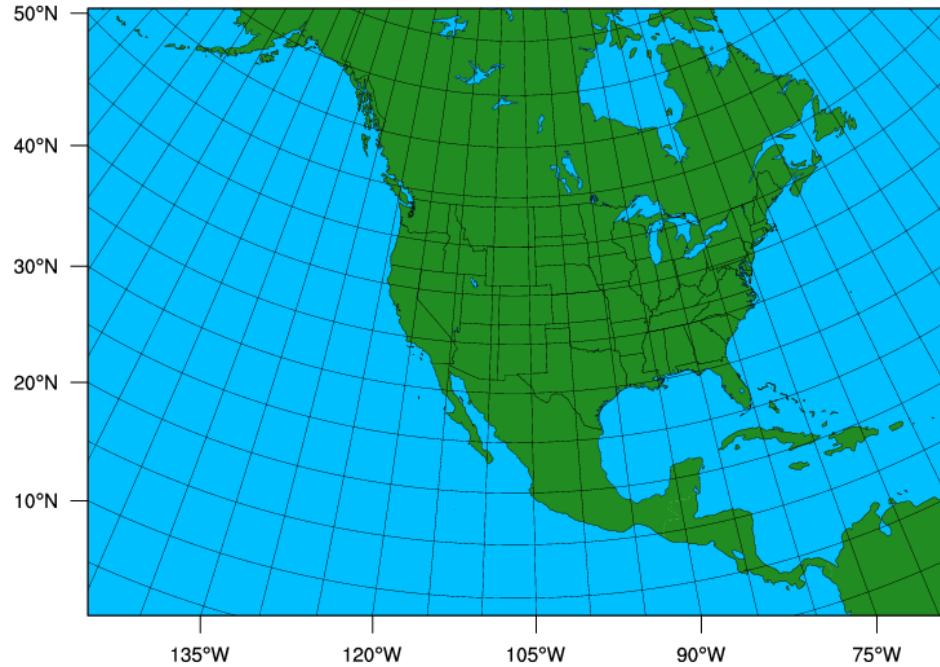

Figure 1. WRF-ARW 15 km computation domain.

2.3.2 Model Configuration

The table below lists the AFv3.5.1 physics suite configuration, similar to the AF's current OC, and the AFv3.8.1 replacement configuration that was used in this testing.

	AFv3.5.1	AFv3.8.1
Microphysics	WRF Single-Moment 5 scheme	Thompson aerosol-aware scheme
Radiation LW and SW	RRTM/Dudhia schemes	RRTMG scheme
Surface Layer	Monin-Obukhov similarity theory	Revised MM5 Monin-Obukhov scheme
Land-Surface Model	Noah	Noah
Planetary Boundary Layer	Yonsei University scheme	Yonsei University scheme
Convection	Kain-Fritsch scheme	Kain-Fritsch scheme

The AFv3.5.1 configuration was run with a long time step of 90 s, while the AFv3.8.1 configuration was run with a long time step of 60 s; both used an acoustic step of 4. While it was initially desired to keep the same long time step between the two configurations, the decision to reduce it for the v3.8.1 configuration stemmed from unstable model runs; through communication with model physics experts [Jimy Dudhia (NCAR/MMM) and Greg Thompson (NCAR/RAL)] it was recommended that the Thompson Eidhammer aerosol-aware scheme not be called with a time step greater than 60 s. At this time step, the model remained stable for all runs conducted. Calls to

3

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 for AFv3.5.1; 0.5 for AFv3.8.1), vertical-velocity damping, and a 5-km-deep diffusive damping layer at the top of the domain (coefficient 0.05 for AFv3.5.1; 0.02 for AFv3.8.1). Positive-definite moisture advection was turned on for both configurations and positive-definite scalar advection was turned on for v3.8.1.

Appendix A provides relevant portions of the *namelist.input* file for v3.8.1 and *Appendix B* highlights the differences in the settings used for v3.5.1.

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. The *copygb* program was used to horizontally interpolate the model forecasts from their native 15-km grid to a smaller, 15-km CONUS grid; the 15-km CONUS domain will be used for the evaluation presented below. 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 quantitative precipitation forecasts (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 regions were verified for the surface and upper air, as well as precipitation variables. Area-averaged results were computed for the CONUS region, CONUS-East

4

and CONUS-West regions, and 14 sub-regions (*Figure 2*). While only a portion of the full results will be discussed in detail for this report, all East, West, and sub-region results are available on the DTC webpage established for this particular test and evaluation activity (http://www.dtcenter.org/eval/afwa_test/wrf_v3.8.1/). 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 initializations individually for surface fields in order to preserve the diurnal signal.
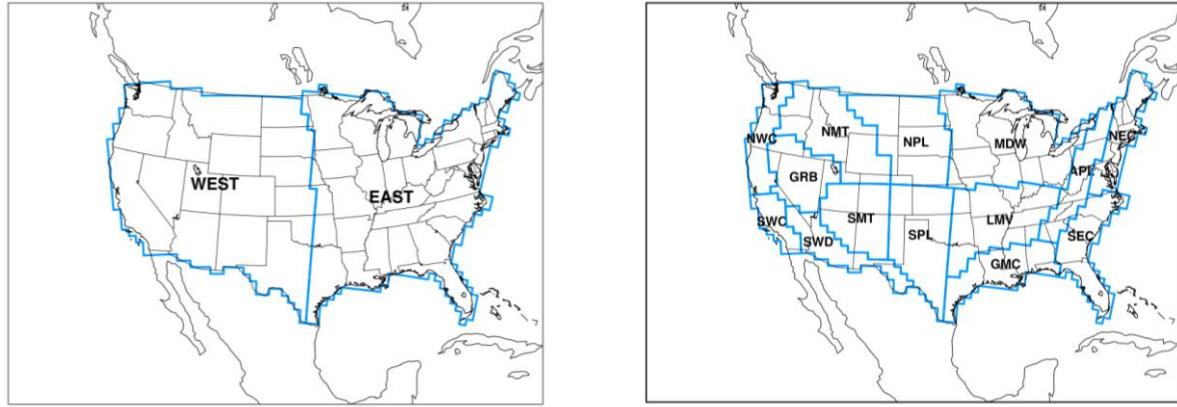

Figure 2. Maps showing the locations of the CONUS-West, CONUS-East (a) and 14 verification regions (b). The outermost blue outline of the verification regions depicts the CONUS verification region.

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 pair-wise difference statistics include zero, the differences are not SS. 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. Thus, a more conservative estimate of SS was employed for frequency bias 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 information. Therefore, in addition to determining SS, the concept of establishing practical significance (PS) was also utilized in 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://library.wmo.int/opac/doc_num.php?explnum_id=3121). 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 in 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

NCEP Stage II analyses were used as the observational dataset of gridded precipitation accumulation. For the QPF verification, a grid-to-grid comparison was made by interpolating the precipitation analyses to the 15-km CONUS domain. For both the precipitation analyses and the post-processed model output, the budget interpolation option was used, which is a method that quasi-conserves area averages. This regridded analysis was then used to evaluate the forecast. Accumulation periods of 3 and 24 hours were examined. Because the 24-hour NCEP Stage II precipitation accumulation analyses were only valid at 12 UTC, the 24-hour QPF were 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 QPF 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 AFv3.5.1 and AFv3.8.1 configurations using the formula:

$$S = 1 - \frac{(RMSE_{NoahMP})^2}{(RMSE_{AFWAOC})^2}$$

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 AFv3.5.1 configuration has higher skill and values greater than one indicate the AFv3.8.1 configuration has higher skill.

4. Verification Results

Due to the substantial changes made to the configuration of the model between AFv3.5.1 and AFv3.8.1 a large number of differences were anticipated at all levels of the atmosphere. As such, the information summarized in this report includes both

surface and upper-air results. In addition, a full suite of verification plots (by type, metric, lead time, threshold, season, etc.), can be accessed on the project webpage provided in Section 3. The first part of the evaluation discusses configuration performance over the CONUS for summer and winter aggregations for the standard verification metrics. In addition to the time series plots shown, further investigation of forecast performance for both configurations over diverse regions of the CONUS is included. The bias at each observation station is presented by surface variable to provide a means to spatially assess the configurations performance respective to the observations. When visualizing the results in this manner, seasonal and temporal differences are apparent, both regionally and between configurations.

Differences between the two configurations are computed by subtracting AFv3.8.1 from AFv3.5.1 (AFv3.5.1-AFv3.8.1). BCRMSE is always a positive quantity, and a perfect score is zero. This results in positive (negative) differences indicating the AFv3.8.1 (AFv3.5.1) 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 and below with a lower limit of $-1/3$. Thus, if the pair-wise difference is negative (positive), the AFv3.5.1 (AFv3.8.1) 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 aggregate value. A breakdown of the configuration with SS and PS better performance by variable, season, statistic, initialization hour, forecast lead time, and vertical level aggregated over the CONUS domain is summarized in Tables 1-8, where the favored configuration is highlighted.

4.1 Temperature BCRMSE and Bias

At the surface, BCRMSE values for both configurations exhibit similar results, including a relatively weak diurnal variation and a small increase with lead time (*Figure 3*). A number of SS and PS pair-wise differences are noted (*Table 1*). Overall, when there is a PS pair-wise difference, AFv3.8.1 is favored. There are fewer PS pair-wise differences between the two configurations for the winter aggregation compared to the summer aggregation and over the West region compared to the East region.

a)

b)
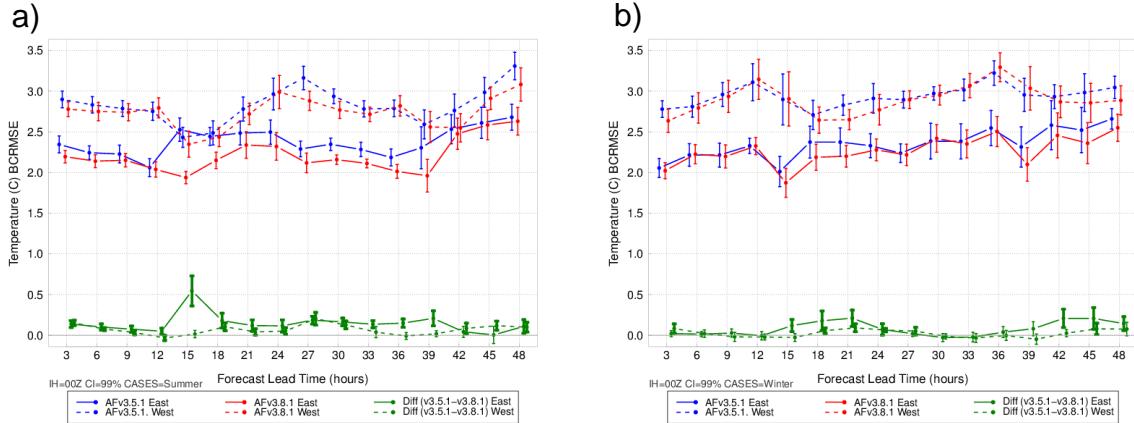
Figure 3. Time series plot of 2 m AGL temperature (°C) median BCRMSE for the 00 UTC initializations aggregated across the (a) summer and (b) winter cases for the East (solid) and West (dashed) verification domains. AFv3.5.1 is in blue, AFv3.8.1 in red, and the differences (AFv3.5.1-AFv3.8.1) in green. The vertical bars attached to the median represent the 99% CIs; the thicker vertical bars denote SS.

When considering 2 m temperature bias, both configurations have a cold bias throughout much of the forecast period, especially over the East region (*Figure 4*). A strong diurnal signal is noted for the winter aggregation, in particular, which is stronger (more amplified) over the West region. The maximum cold bias tends to occur between valid times of about 18 – 03 UTC. For some spatial and temporal aggregations there is a transition to a neutral or warm bias between 09 – 15 UTC (especially over the West). Overall, AFv3.8.1 tends to be warmer than AFv3.5.1; thus, when a cold bias is noted for both configurations, AFv3.8.1 is preferred and as the configurations transition to a warm bias, AFv3.5.1 is preferred (*Table 1*). All pair-wise differences are PS for 2 m temperature bias.

a)

b)
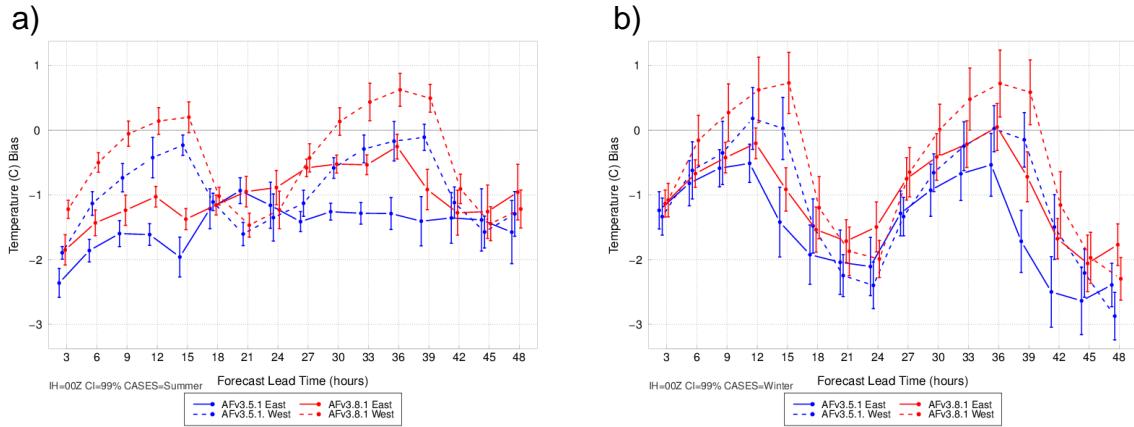
Figure 4. Time series plot of 2 m AGL temperature (°C) median mean error (bias) for the 00 UTC initializations aggregated across the (a) summer and (b) winter cases for the East (solid) and West (dashed) verification domains. AFv3.5.1 is in blue and AFv3.8.1 in red. The vertical bars attached to the median represent the 99% CIs.

The point observation plots of 2 m temperature bias for the summer aggregation (*Figure 5*) at the 36 h forecast lead time for all 00 UTC initializations (00f36; valid 12 UTC) highlight regional differences in performance within a single configuration and between

9

the two configurations. A cold bias for AFv3.5.1 is noted from the Midwest extending eastward to the Atlantic Coast and southward into the Lower Mississippi Valley; cold biases are also observed along the Northwest Coast (*Figure 5*a). On the other hand, a warm bias is noted across a majority of the western portion of the CONUS and for some areas along the SE coast. Overall, AFv3.8.1 is warmer than AFv3.5.1 resulting in smaller, less intense areas of cold bias and increased areas of warm bias (*Figure 5*b). While both configurations have a fairly consistent cold bias across the CONUS at the 48 h lead time (00f48; valid at 00 UTC), the cold bias is largest across the northwest US, including the Northern Rocky Mountains, and Great Basin regions, the southern Plains, and along the Appalachians (*Figure 5*c,d). Again, AFv3.8.1 is generally warmer for many of the regions at the 48 h lead time but still suffers from a cold bias over a majority of the CONUS.

a)

b)

c)

d)
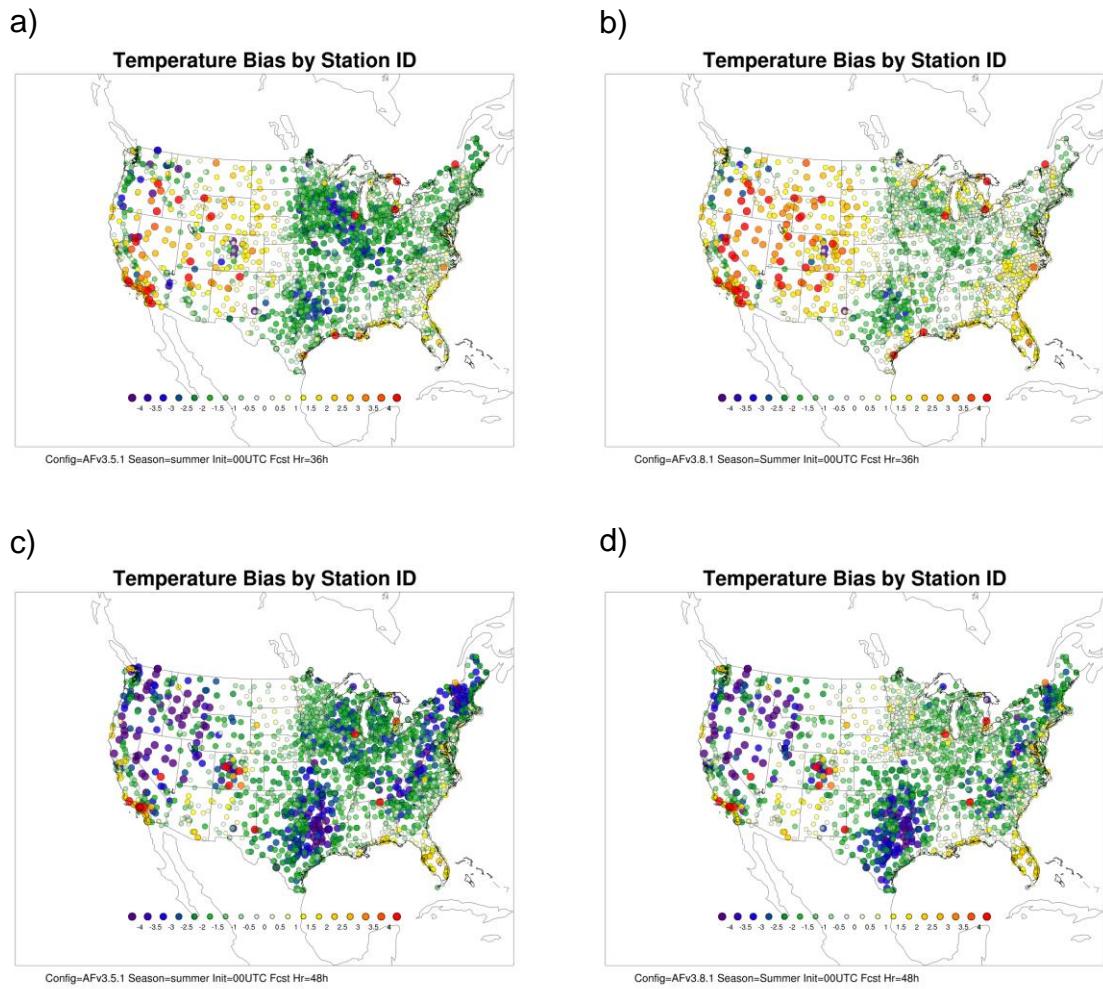
Figure 5. Spatial plot of 2 m AGL temperature (°C) bias by observation station for all 00 UTC initializations in the summer aggregation for (a) AFv3.5.1 at the 36 h forecast lead time, (b) AFv3.8.1 at the 36 h forecast lead time, (c) AFv3.5.1 at the 48 h forecast lead time, and (d) AFv3.8.1 at the 48 h forecast lead time.

For the winter aggregation at the 36 h forecast lead time, a similar pattern as that for the summer aggregation (generally warm across the west and cool across the east) is noted except both configurations are warmer overall leading to smaller cold biases in the East and larger warm biases in the West, especially for AFv3.8.1 (*Figure 6*a,b). A particular regional exception is the northern Rockies region, which exhibits a notable cool bias relative to other regions at the 36 h forecast lead time. For the 48 h forecast lead time, AFv3.5.1 has a severe cold bias that is damped but still present with AFv3.8.1 (*Figure 6*c,d). The difference between the point plots from the 36 to 48 h forecast lead time highlights the large diurnal swing in bias for both configurations indicating neither model captures the maximum temperature during the late afternoon very well.

a)

Temperature Bias by Station ID

Config=AFv3.5.1 Season=winter Init=00UTC Fcst Hr=36h

b)

Temperature Bias by Station ID

Config=AFv3.8.1 Season=Winter Init=00UTC Fcst Hr=36h

c)

Temperature Bias by Station ID

Config=AFv3.5.1 Season=winter Init=00UTC Fcst Hr=48h

d)

Temperature Bias by Station ID

Config=AFv3.8.1 Season=Winter Init=00UTC Fcst Hr=48h
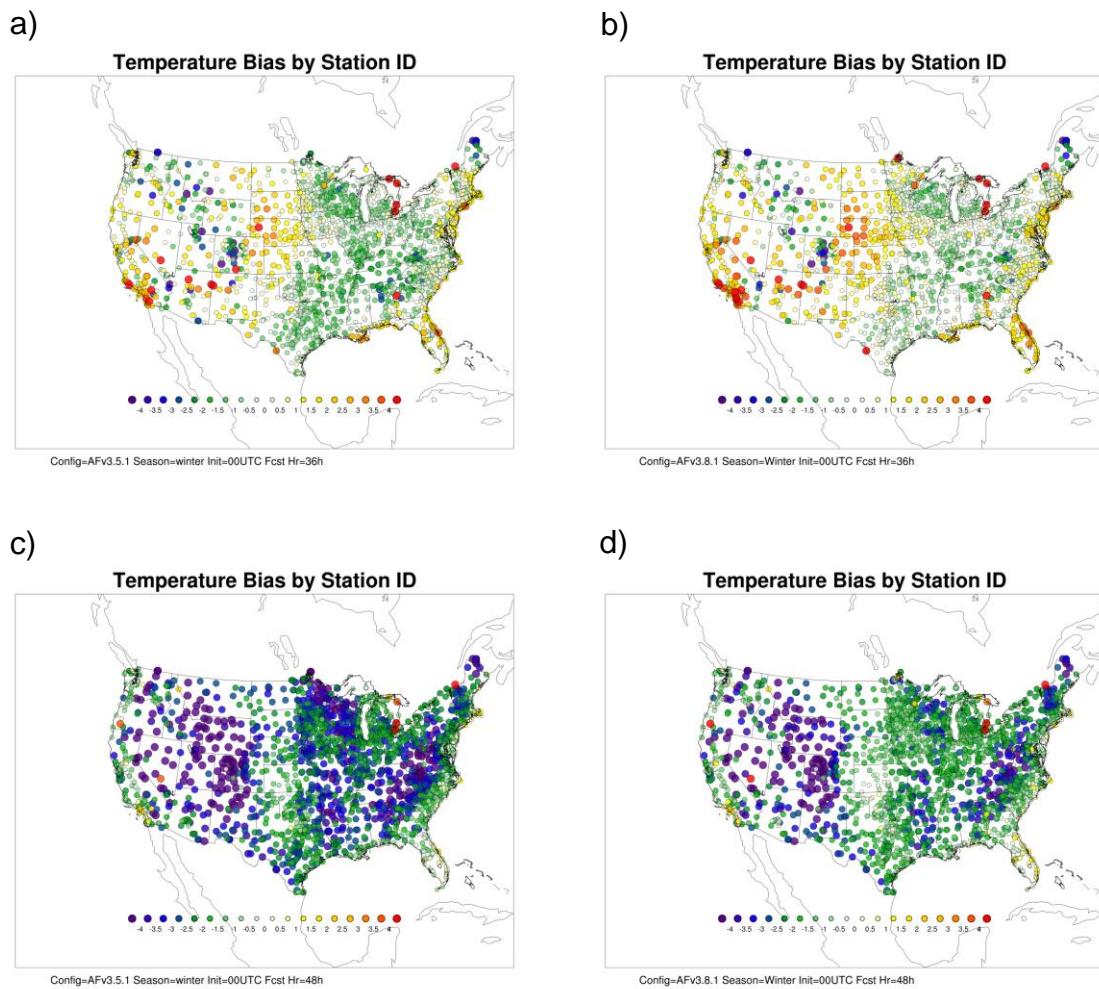
Figure 6. Spatial plot of 2 m AGL temperature (°C) bias by observation station for all 00 UTC initializations in the winter aggregation for (a) AFv3.5.1 at the 36 h forecast lead time, (b) AFv3.8.1 at the 36 h forecast lead time, (c) AFv3.5.1 at the 48 h forecast lead time, and (d) AFv3.8.1 at the 48 h forecast lead time.

Regardless of temporal aggregation or forecast lead time, both configurations have a minimum in temperature BCRMSE values between 500 and 300 hPa, with the largest

error occurring at the lower and upper-levels *(Figure 7*a,b). In general, similar vertical distributions of errors are seen in the East and West regions (not shown); however, the levels at which PS pair-wise differences are noted for each temporal aggregation vary with region (*Table 4*). While there are some SS pair-wise differences that favor AFv3.5.1, all but one (winter in the West) of the PS pair-wise differences, which are generally seen at 850 and 700 hPa, indicate AFv3.8.1 has smaller BCRMSE values.
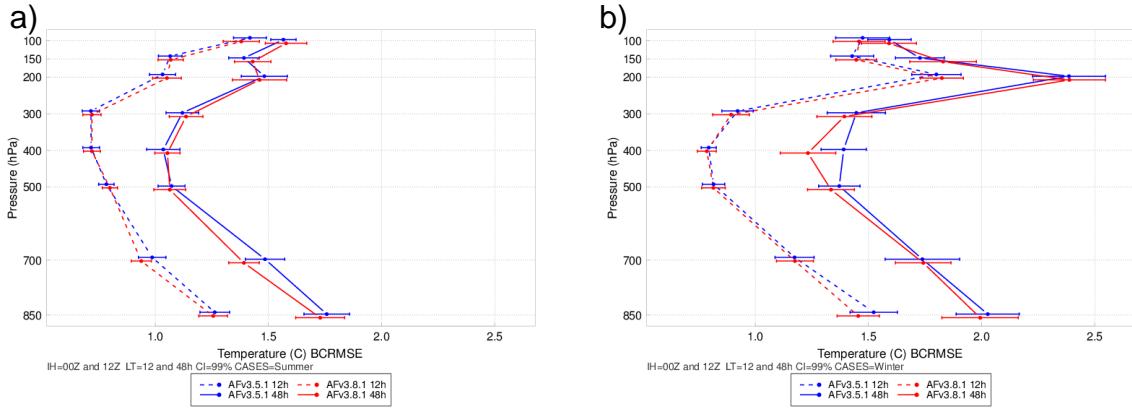
Figure 7. Vertical profile of the median temperature (°C) BCRMSE for the 12- (dashed) and 48-h (solid) forecast lead times aggregated across the CONUS domain over the (a) summer and (b) winter cases. AFv3.5.1 is in blue and AFv3.8.1 in red. The horizontal bars attached to the median represent the 99% CIs.

The shape of the temperature bias distribution is dependent on temporal aggregation, vertical level, and forecast lead time. For the CONUS region summer aggregation, the 12 h forecast lead time exhibits a warm bias at most levels for both configurations, while a cold bias is seen for the 48 h forecast lead time at 850 hPa for both configurations and at mid-levels (400 – 300 hPa) for AFv3.8.1 (*Figure 8*a). For the winter aggregation, a cold bias for both configurations is noted at and below 500 hPa and at 150 hPa, and a warm bias is seen at 300, 200, and 100 hPa (*Figure 8*b). Ultimately, the preferred configuration is highly dependent on level, season, and lead time (*Table 4*a). In general, there are more PS pair-wise differences (more often favoring AFv3.8.1) for the summer aggregation compared to the winter aggregation (where both configurations are favored a similar number of times).
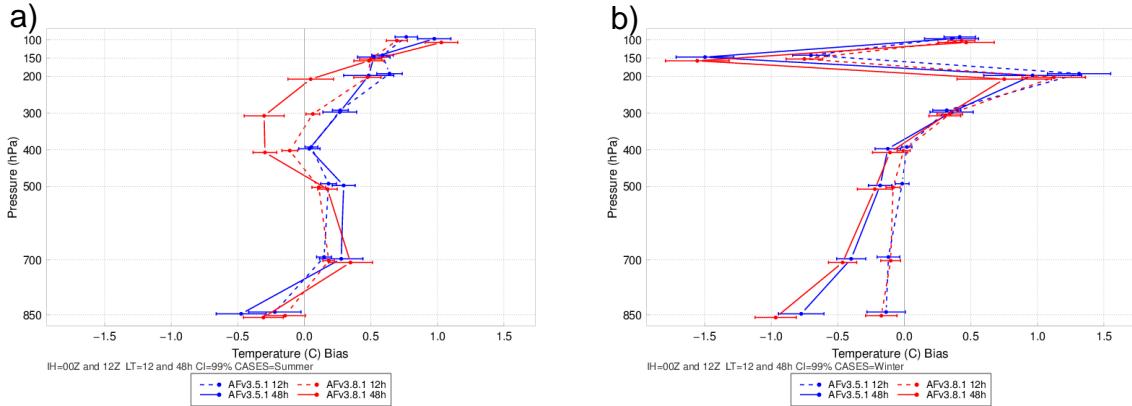
Figure 8. Vertical profile of the median temperature (°C) mean error (bias) for the 12- (dashed) and 48-h (solid) forecast lead times aggregated across the CONUS domain over the (a) summer and (b) winter cases. AFv3.5.1 is in blue and AFv3.8.1 in red. The horizontal bars attached to the median represent the 99% CIs.

When looking at spatial aggregations, the East region has similar behavior to the CONUS results while the West tends to have a larger cold bias at and below 700 hPa and warm bias between 400 and 200 hPa *(Figure 9*a,b). For the summer aggregation, AFv3.8.1 is almost always favored with PS over the West, while AFv3.5.1 is favored for some mid- and upper-levels for the East (*Table 4*). The SS favored configuration for the winter aggregation varies more so than during summer, although both configurations are favored with PS a similar number of times.
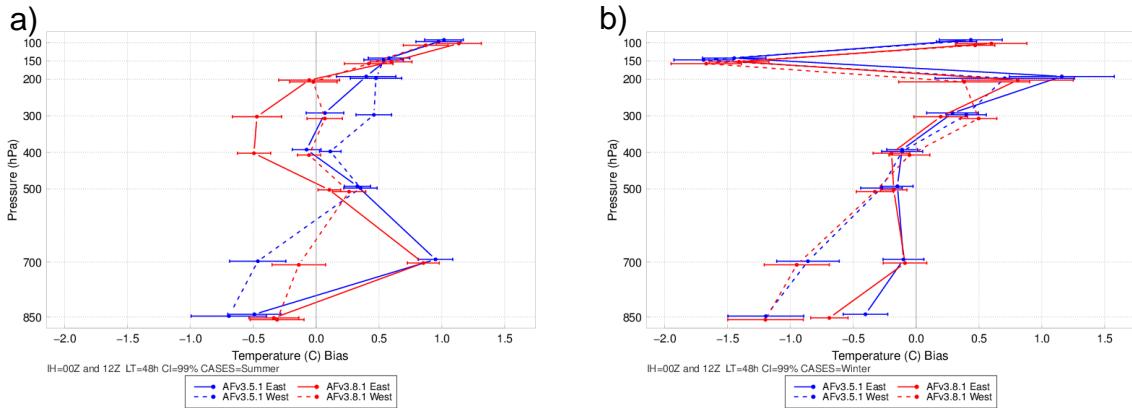
Figure 9. Vertical profile of the median temperature (°C) mean error (bias) at the 48-h lead time aggregated over the East (solid) and West (dashed) regions for the (a) summer and (b) winter cases. AFv3.5.1 is in blue and AFv3.8.1 in red. The horizontal bars attached to the median represent the 99% CIs.

4.2 Dew Point Temperature BCRMSE and Bias

Near surface dew point temperature BCRMSE values in both seasons exhibit a relatively weak diurnal signal with a gentle increase in error as forecast lead time increases (*Figure 10*). Notably, for both seasons, there is an increase in BCRMSE

between the East and West regions, with median errors of approximately 1º C higher in the West region. In the East during the summer, there are many SS/PS pair-wise differences throughout the forecast period, with all differences showing AFv3.8.1 having less error (*Table 2*c); PS pair-wise differences typically occur between 03-12 UTC. In the West, a reversal of performance is seen, with AFv3.5.1 always performing better than AFv3.8.1 and with most differences being PS (*Table 2*b). Due to the dichotomy in the East and West regions, there are overall less differences in the CONUS than in the sub-regions, with most PS pair-wise differences occurring for the 12 UTC initializations (*Table 2*a). In the summer, AFv3.8.1 is exclusively the better performer, while in the winter, AFv3.5.1 has lower error for the 12 UTC initializations.

a)   b)
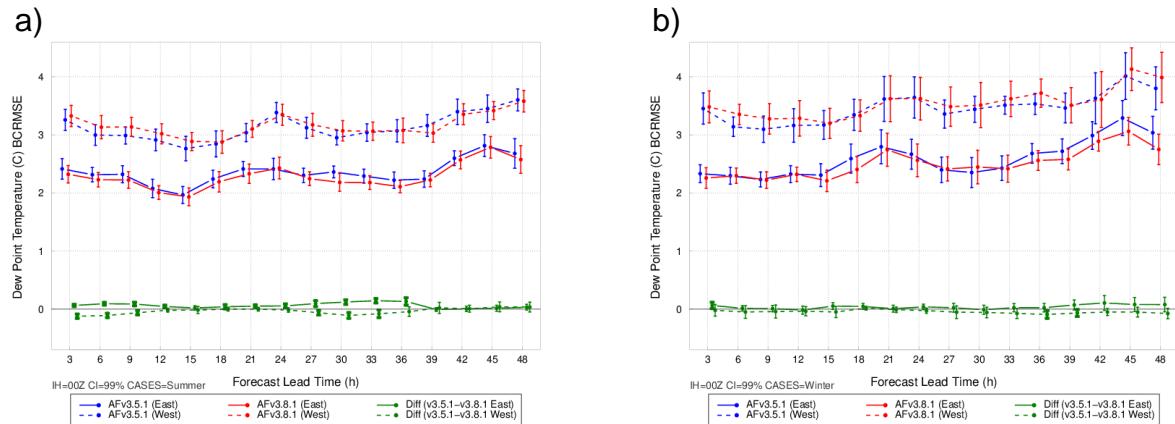
Figure 10. Time series plot of 2 m AGL dew point temperature (°C) median BCRMSE for the 00 UTC initializations aggregated across the (a) summer and (b) winter cases for the East (solid) and West (dashed) verification domains. AFv3.5.1 is in blue, AFv3.8.1 in red, and the difference (AFv3.5.1-AFv3.8.1) in green. The vertical bars attached to the median represent the 99% CIs; the thicker vertical bars denote SS.

A diurnal signal in 2 m dew point temperature bias is seen in the East, West, and CONUS (not shown) in both the summer and winter seasons (*Figure 11*). In the summer, a moist bias is typically observed between 15-00 UTC in the East, with a slight shift later in the West from 21-06 UTC; a neutral-to-dry bias is seen around 06-12 UTC (09-18 UTC) in the East (West) region (*Figure 11*a-b). AFv3.8.1 has a smaller diurnal variation than AFv3.5.1 in the East and West regions, often resulting in smaller magnitude moist and dry biases; therefore, in the summer, most PS pair-wise differences indicate AFv3.8.1 as the better performing configuration (*Table 2*b,c). It is also noted that while AFv3.8.1 is most often favored in the West and East, there are more PS pair-wise differences favoring AFv3.5.1 in the West compared to the East. Similar to summer, a winter season diurnal signal is noted in the East and West regions, with minimum bias around 15-18 UTC and maximum bias 21-00 UTC (*Figure 11*c-d). A moist bias is most often observed for AFv3.5.1, while AFv3.8.1 typically displays neutral-to-moist biases in both regions. Due to this behavior, AFv3.8.1 is typically the better performer throughout the model integration in the winter, regardless of region (*Table 2*).
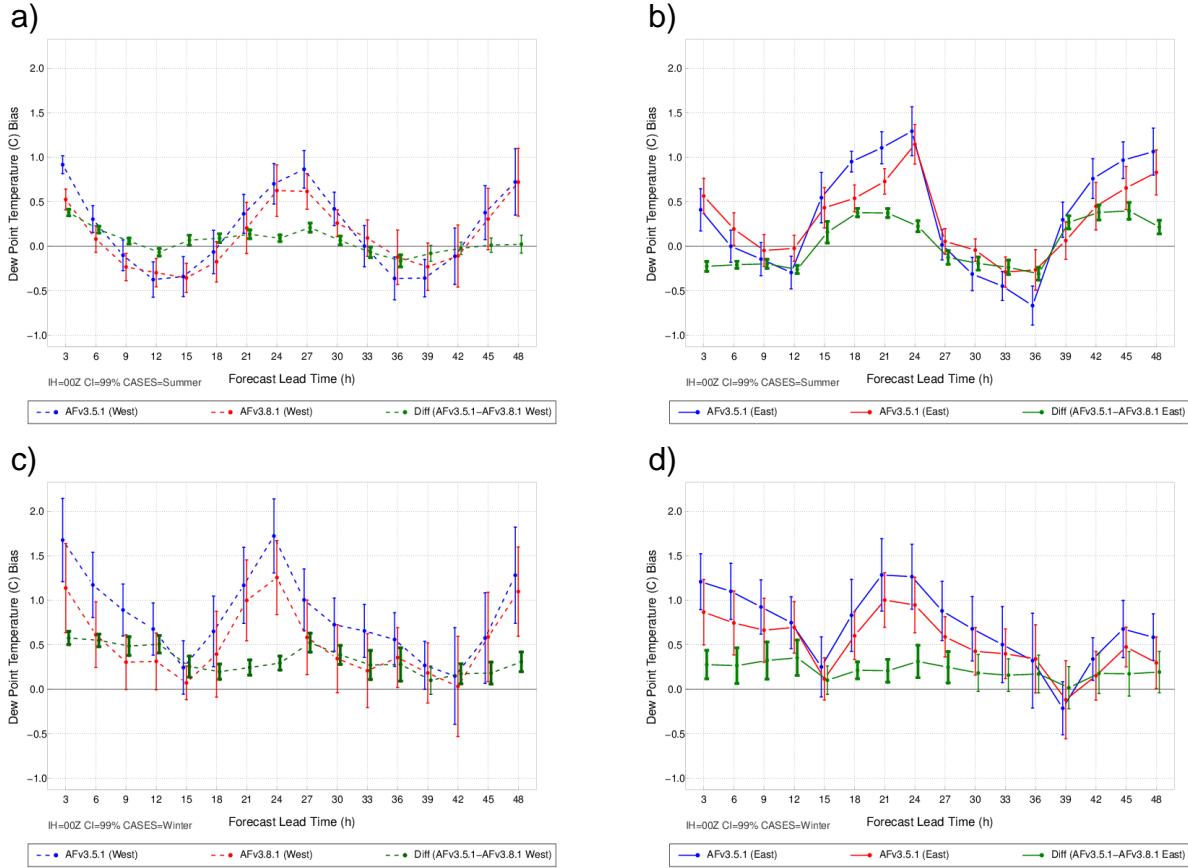
Figure 11. Time series plot of 2 m AGL dew point temperature (°C) median mean error (bias) for the 00 UTC initializations aggregated across the (a) summer cases over the West region, (b) summer cases over the East region, (c) winter cases over the West region, and (d) winter cases over the East region. AFv3.5.1 is in blue and AFv3.8.1 in red, and the difference (AFv3.5.1-AFv3.8.1) in green. The vertical bars attached to the median represent the 99% CIs; the thicker vertical bars denote SS.

Point observation plots of 2 m dew point temperature bias for the 36 h and 48 h hour forecasts (*Figure 12*) highlight the diurnal signal seen in the summary bias statistics for summer (*Figure 11*a-b). At the 36 h forecast lead time, both configurations have a dry bias over the Midwest, Appalachians, Northeast Coast, portions of the Lower Mississippi Valley, the Rocky Mountains, and California. While the spatial patterns are similar between AFv3.5.1 and AFv3.8.1, AFv3.5.1 has overall larger dry biases than AFv3.8.1, corresponding with the summary plots described above. Both configurations have a wet bias over the Southern Plains, Great Basin, and areas near the Southeast Coast. By the 48 h forecast, a general shift toward larger moist biases across the CONUS is seen for both configurations (*Figure 12*c-d). The moist bias noted at 36 h over the Southern Plains strengthens with a large area having biases at or in excess of 4º C. With exception to areas in the Northeast Coast, both configurations have a moist bias over most of the Eastern region, with AFv3.8.1 having slightly less of an overall moist bias. In the Western region of the CONUS, a number of regions see an increase in moist bias from the 36 h to 48 h forecast lead time, particularly areas along the Northwest Coast and areas within the Northern and Southern Rocky Mountains.

15

a)

Dew Point Temperature Bias by Station ID

Config=AFv3.5.1 Season=summer Init=00UTC Fcst Hr=36h

b)

Dew Point Temperature Bias by Station ID

Config=AFv3.8.1 Season=Summer Init=00UTC Fcst Hr=36h

c)

Dew Point Temperature Bias by Station ID

Config=AFv3.5.1 Season=summer Init=00UTC Fcst Hr=48h

d)

Dew Point Temperature Bias by Station ID

Config=AFv3.8.1 Season=Summer Init=00UTC Fcst Hr=48h
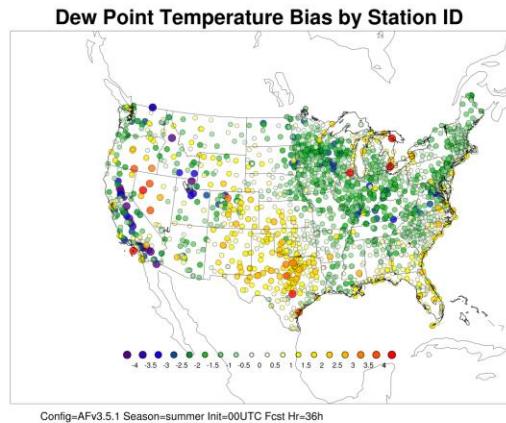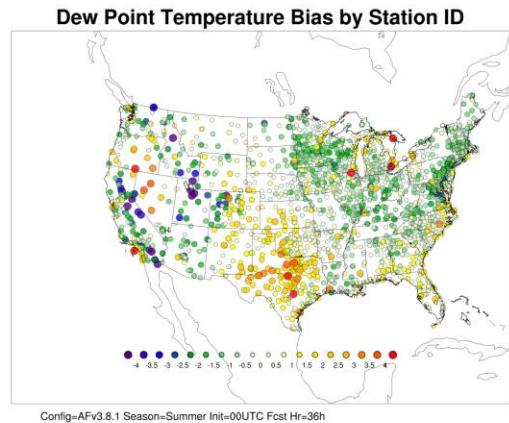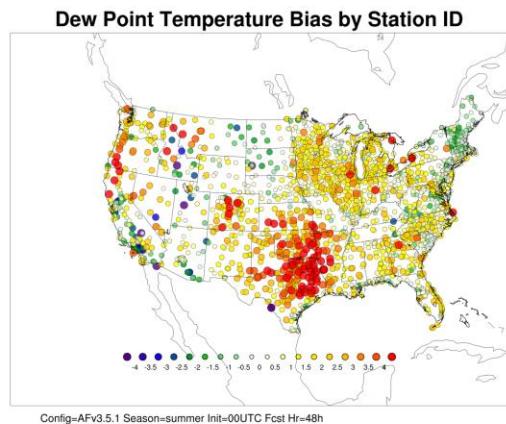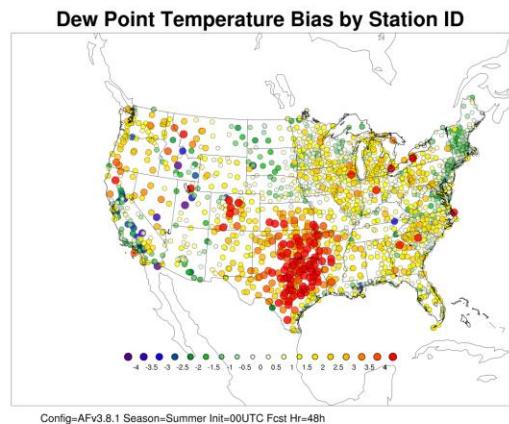
Figure 12. Spatial plot of 2 m AGL dew point temperature (°C) bias by observation station for all 00 UTC initializations in the summer aggregation for (a) AFv3.5.1 at the 36 h forecast lead time, (b) AFv3.8.1 at the 36 h forecast lead time, (c) AFv3.5.1 at the 48 h forecast lead time, and (d) AFv3.8.1 at the 48 h forecast lead time.

In the winter, the point observation bias plots show an overall moist bias at the 36 h forecast lead time for both configurations, with the Midwest, Northern Pains, Northeast Coast, Great Basin, and Southern and Northern Rocky Mountains contributing the most to moist bias (*Figure 13*a-b). Overall, in these areas, AFv3.5.1 has higher 2 m dew point temperature biases than AFv3.8.1; in particular, several stations over the Rocky Mountains have large differences, with AFV3.8.1 having a strong dry bias where AFv3.5.1 has a near neutral bias. While there is an overall moist bias at the 36 h forecast lead time, areas along the West Coast, Southern Plains, and within the Southeast Coast have a neutral/dry bias, with AFv3.5.1 again having higher bias values. At the 48 h forecast lead time, most areas in the West see an overall moistening from the 36 h forecast lead time (*Figure 13*c-d). In addition, several regions in the East that were drier at the 36 h forecast lead time, see a transition to a more neutral or slightly wet bias at 48 h (Southern Plains, areas in the Gulf Coast and Southeastern regions).

a)

Dew Point Temperature Bias by Station ID

Config=AFv3.5.1 Season=winter Init=00UTC Fcst Hr=36h

b)

Dew Point Temperature Bias by Station ID

Config=AFv3.8.1 Season=Winter Init=00UTC Fcst Hr=36h

c)

Dew Point Temperature Bias by Station ID

Config=AFv3.5.1 Season=winter Init=00UTC Fcst Hr=48h

d)

Dew Point Temperature Bias by Station ID

Config=AFv3.8.1 Season=Winter Init=00UTC Fcst Hr=48h

Figure 13. Same as Figure 12, but for the winter aggregation.

Minimal differences are noted in the East and West regions when considering upper level BCRMSE; therefore, focus is placed on the CONUS domain (*Figure 14*). In the summer season, dew point temperature BCRMSE grows with increasing forecast lead time and decreasing vertical level (*Figure 14*a). No SS/PS pair-wise differences are noted over the CONUS, East, and West regions (*Table 5*). For both configurations, in the winter season, increasing BCRMSE is noted with increasing forecast lead time but not vertical level (*Figure 14*b). A few PS pair-wise differences are noted in the CONUS and East regions, always favoring AFv3.8.1 (*Table 5*).

a)

b)
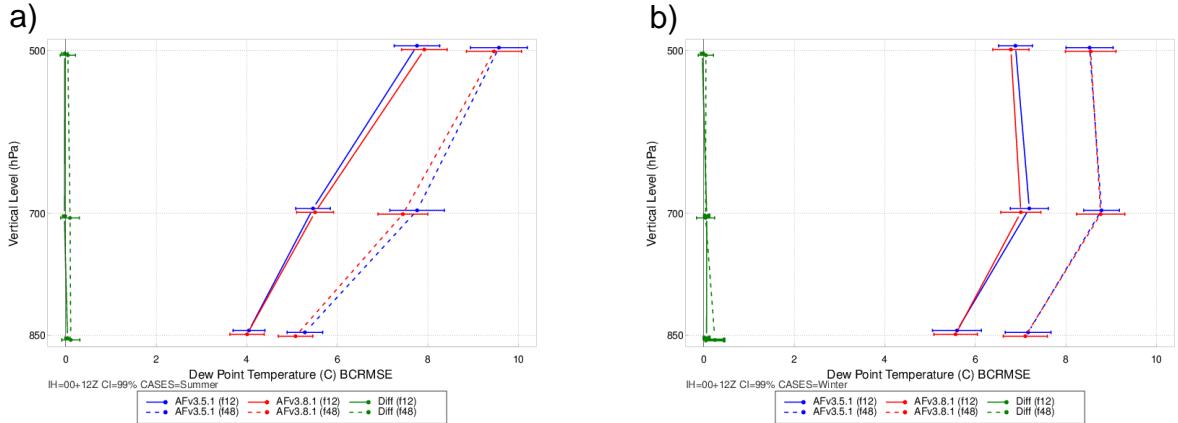
Figure 14. Vertical profile of the median dew point temperature (°C) BCRMSE for the 12 (solid) and 48 h (dashed) forecast lead times aggregated across the CONUS domain over the (a) summer and (b) winter cases. AFv3.5.1 is in blue and AFv3.8.1 in red, and the difference (AFv3.5.1-AFv3.8.1) in green. The horizontal bars attached to the median represent the 99% CIs; the thicker horizontal bars denote SS.

More differences in regional bias distribution are noted in dew point temperature bias than in BCRMSE (*Figure 15*). In the summer, the East and West regions have differing bias profiles (*Figure 15*a-b). For both configurations, the West region varies little with vertical level and with forecast lead time, with a consistent moist bias of ~1º C. In the West, AFv3.5.1 has a slight drying trend with forecast lead time at 850-700 hPa, while AFv3.8.1 displays a minimal moistening trend at those levels. This behavior leads to AFv3.5.1 having several PS pair-wise differences in the West, with all at or above 700 hPa (*Table 5*b). In the East, both configurations have similar bias profiles, with a dry bias at 850 hPa, which further dries at 700 hPa before moistening to a neutral-to-moist bias at 500 hPa. All PS pair-wise differences in the East favor AFv3.8.1 and occur after the 12 h forecast lead time (*Table 5*c). The bias curves in the winter are similar between the two configurations, with median dew point temperature bias generally increasing with vertical level; this is observed in both the East and West regions at all forecast lead times (*Figure 15*c-d). In addition, as forecast lead time increases, the differences between the two configurations also increases. In the West region, both AFv3.5.1 and AFv3.8.1 have a moist bias at all vertical levels, with AFV3.5.1 having a smaller moist bias, indicating better performance; this is also reflected in PS pair-wise differences, where AFv3.5.1 is always the favored configuration in the West (*Table 5*b). In the East during the winter, at 850 hPa, AFv3.8.1 is unbiased at all forecast lead times (note the median values transition from moist to dry), while AFv3.5.1 has a small moist bias at the beginning of the forecast that dries with increasing time. Both configurations have a moist bias at 700-500 hPa throughout the model integration period. At all levels, AFv3.8.1 typically has higher median bias values than AFv3.5.1.This translates to AFv3.8.1 being PS better at 850 hPa, where there is a dry bias or unbiased forecast, and AFv3.5.1 being better at 500 hPa, where there is a moist bias (*Table 5*c). Overall, across the full CONUS region, difference in performance between the two configurations is dominated by the West results, where all but one PS pair-wise difference favors AFv3.5.1 (*Table 5*a).

18
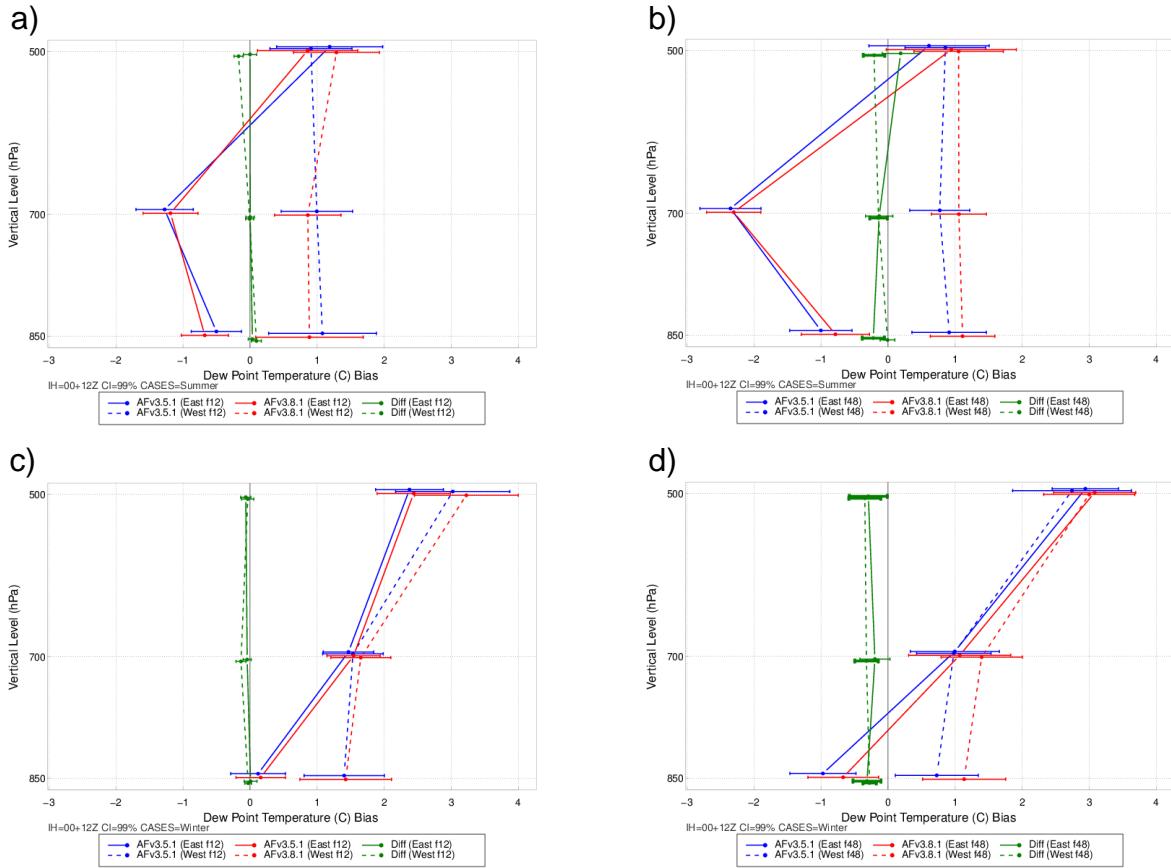
Figure 15. Vertical profile of the median dew point temperature (°C) mean error (bias) for the (a) 12 h forecast lead time over the summer cases, (b) 48 h forecast lead times over the summer cases, (c) 12 h forecast lead time over the winter cases, (d) 48 h forecast lead times over the winter cases aggregated across the East (solid) and West (dashed) regions. AFv3.5.1 is in blue and AFv3.8.1 in red, and the difference (AFv3.5.1-AFv3.8.1) in green. The horizontal bars attached to the median represent the 99% CIs; the thicker horizontal bars denote SS.

4.3 Wind BCRMSE and Bias

For 10 m wind speed BCRMSE, a weak diurnal signature, with a maximum around valid time 00 UTC, is observed during the summer (*Figure 16*a-b). During winter, the diurnal signal is less pronounced. A number of SS pair-wise differences favored AFv3.8.1 during both summer and winter regardless of spatial aggregation; however, none of the differences were PS (*Table 3*). The daytime hours correspond to fewer of the SS differences compared with nighttime hours independent of initialization aggregation.

a)

b)
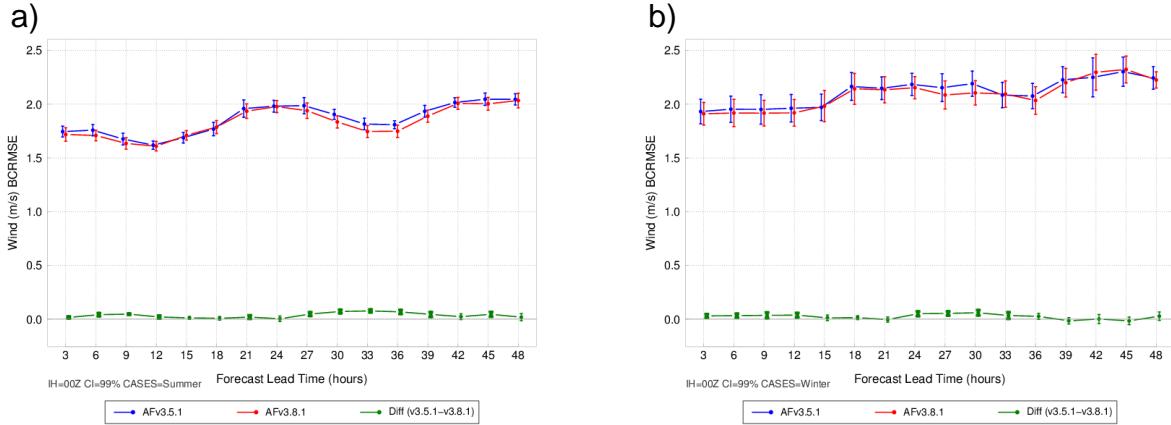
Figure 16. Time series plot of 10 m wind speed (m s⁻¹) median BCRMSE for the 00 UTC initializations aggregated across the (a) summer and (b) winter cases for the CONUS verification domain. AFv3.5.1 is in blue, AFv3.8.1 in red, and the differences (AFv3.5.1-AFv3.8.1) in green. The vertical bars attached to the median represent the 99% CIs; the thicker vertical bars denote SS.

A prominent diurnal signal in bias is seen for all temporal aggregations and both initializations for 10 m wind speed bias, with the largest high bias seen at times valid during overnight hours and smaller-to-neutral wind speed bias during the daytime for all verification domains (*Figure 17*, 00 UTC shown). The wind speed bias is systematically higher during winter compared to summer for both configurations. For the summer aggregation, AFv3.8.1 has a larger high bias than AFv3.5.1 during nighttime hours, while both configurations have similar bias magnitude during the daytime. While the diurnal bias distribution is similar to that for the summer aggregation, the AFv3.8.1 wind speed bias for the winter aggregation is consistently larger than AFv3.5.1 (with a difference of ~0.5 m/s) during most times of day, especially for the CONUS and East regions. For the West region, the consistent winter difference tenuously dissipates to mimic the summer wind speed bias behavior during the daytime periods. The wind speed bias for the CONUS and East regions generally SS favored AFv3.5.1 (*Table 3*). For the West region during summer, the AFv3.8.1 is occasionally favored for some forecast lead times. The only PS pair-wise differences are noted for the winter aggregation over the East region as well as a couple persisting to the CONUS region, influenced by the East region, all of which favor the AFv3.5.1 configuration.
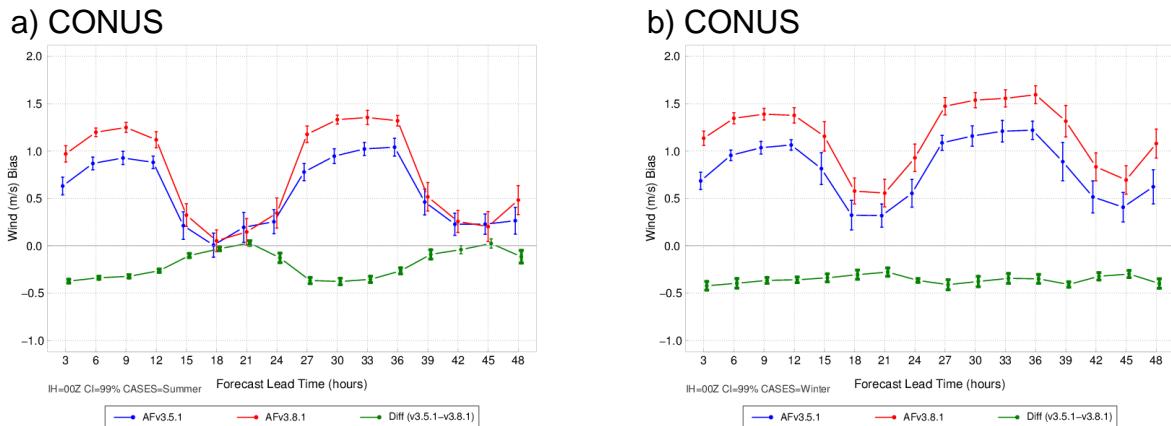
a) CONUS
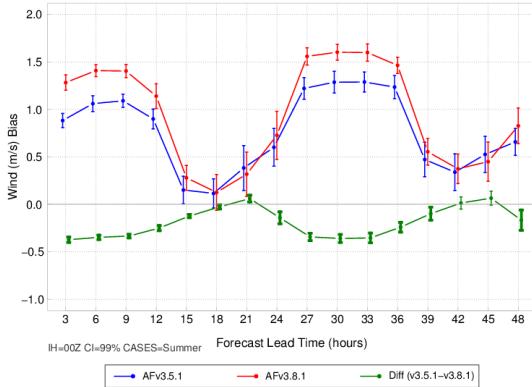
b) CONUS

c) CONUS-East

d) CONUS-East
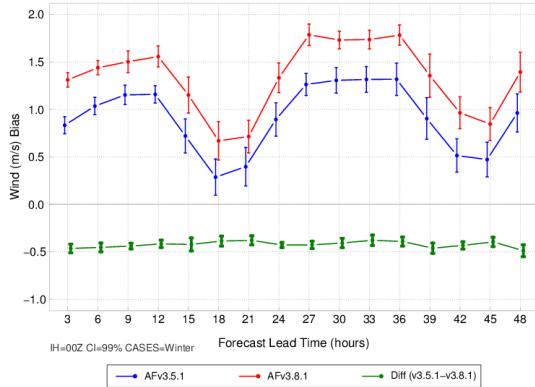
e) CONUS-West
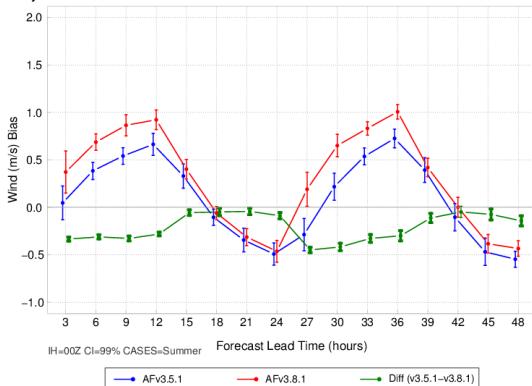
f) CONUS-West
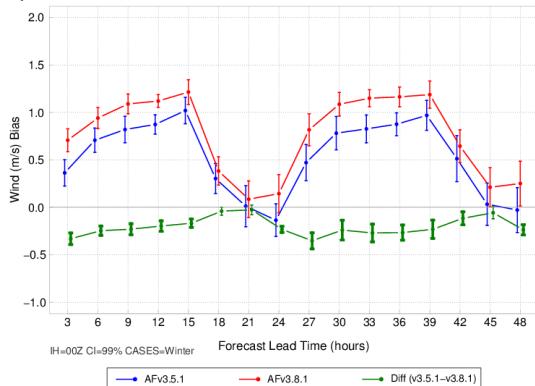
Figure 17. Time series plot of 10 m wind speed (m s^{-1}) median mean error (bias) for the 00 UTC initializations aggregated across the summer (left column) and winter (right column) cases for the CONUS (top row), CONUS-East (center row), and CONUS-West (bottom row) verification domains. AFv3.5.1 is in blue, AFv3.8.1 in red, and the differences (AFv3.5.1-AFv3.8.1) in green. The vertical bars attached to the median represent the 99% CIs; the thicker vertical bars denote SS.

To highlight the diurnal signal, spatial wind speed bias distributions by observation station at 00f18 and 00f30 (valid at 18 and 06 UTC, respectively) for summer (*Figure 18*) and winter (*Figure 19*) are shown. In general, the Plains region exhibits small bias values at numerous observation stations, while high wind speed biases are predominantly located along the Pacific and Atlantic Coasts and across the Upper Midwest. The summer time wind speed bias magnitudes are similar at 00f18 for both versions. Overall, the bias magnitudes increased for the winter aggregation when compared to summer, regardless of forecast lead time. A positive shift in wind speed bias for AFv3.8.1 compared to AFv.3.5.1 CONUS-wide is also noted, consistent with *Figure 17*.

a)

Wind Bias by Station ID

Config=AFv3.5.1 Season=summer Init=00UTC Fcst Hr=18h

b)

Wind Bias by Station ID

Config=AFv3.8.1 Season=Summer Init=00UTC Fcst Hr=18h

c)

Wind Bias by Station ID

Config=AFv3.5.1 Season=summer Init=00UTC Fcst Hr=30h

d)

Wind Bias by Station ID

Config=AFv3.8.1 Season=Summer Init=00UTC Fcst Hr=30h
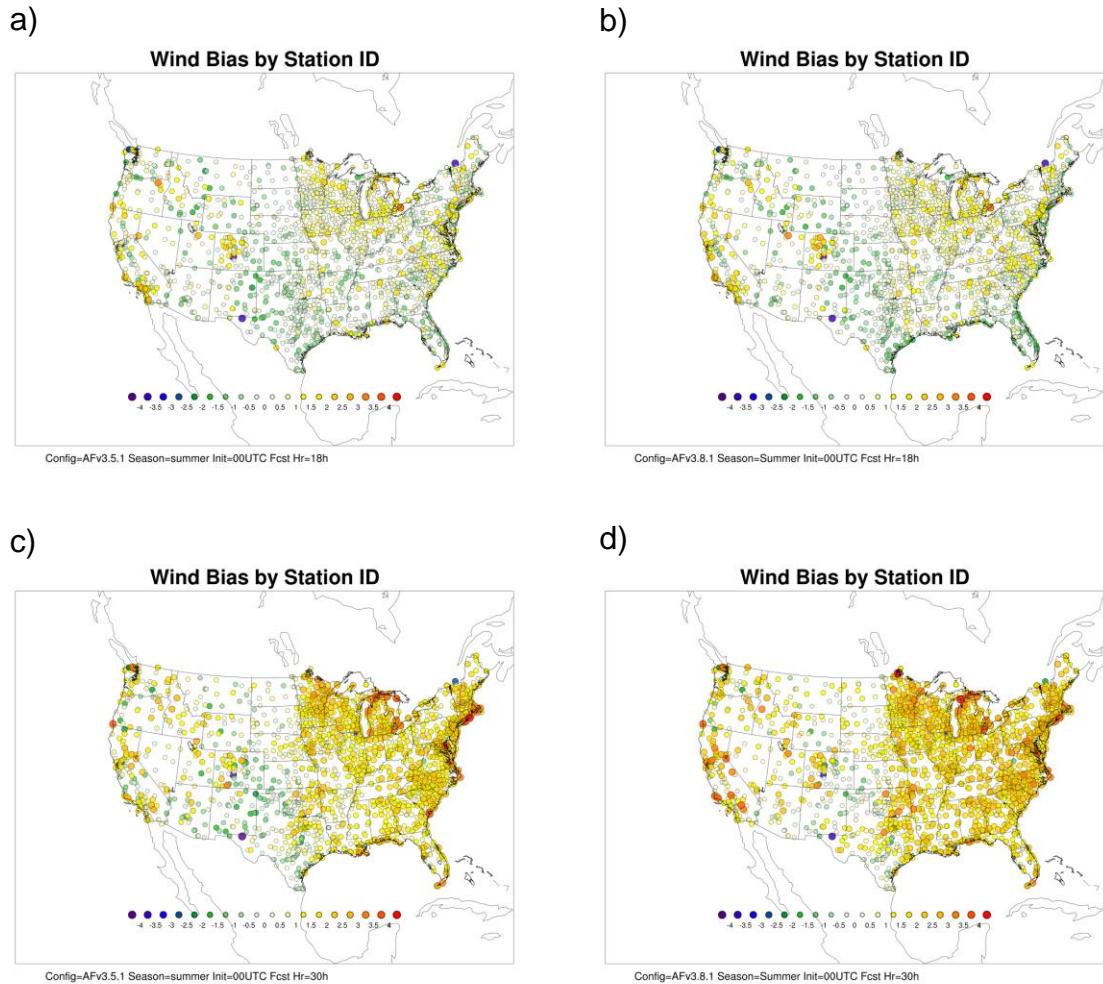
Figure 18. Spatial plot of 10 m wind speed (m s^{-1}) bias by observation station for all 00 UTC initializations in the summer aggregation for (a) AFv3.5.1 at the 18 h forecast lead time, (b) AFv3.8.1 at the 18 h forecast lead time, (c) AFv3.5.1 at the 30 h forecast lead time, and (d) AFv3.8.1 at the 30 h forecast lead time.

a)

Wind Bias by Station ID

Config=AFv3.5.1 Season=winter Init=00UTC Fcst Hr=18h

b)

Wind Bias by Station ID

Config=AFv3.8.1 Season=Winter Init=00UTC Fcst Hr=18h

c)

Wind Bias by Station ID

Config=AFv3.5.1 Season=winter Init=00UTC Fcst Hr=30h

d)

Wind Bias by Station ID

Config=AFv3.8.1 Season=Winter Init=00UTC Fcst Hr=30h
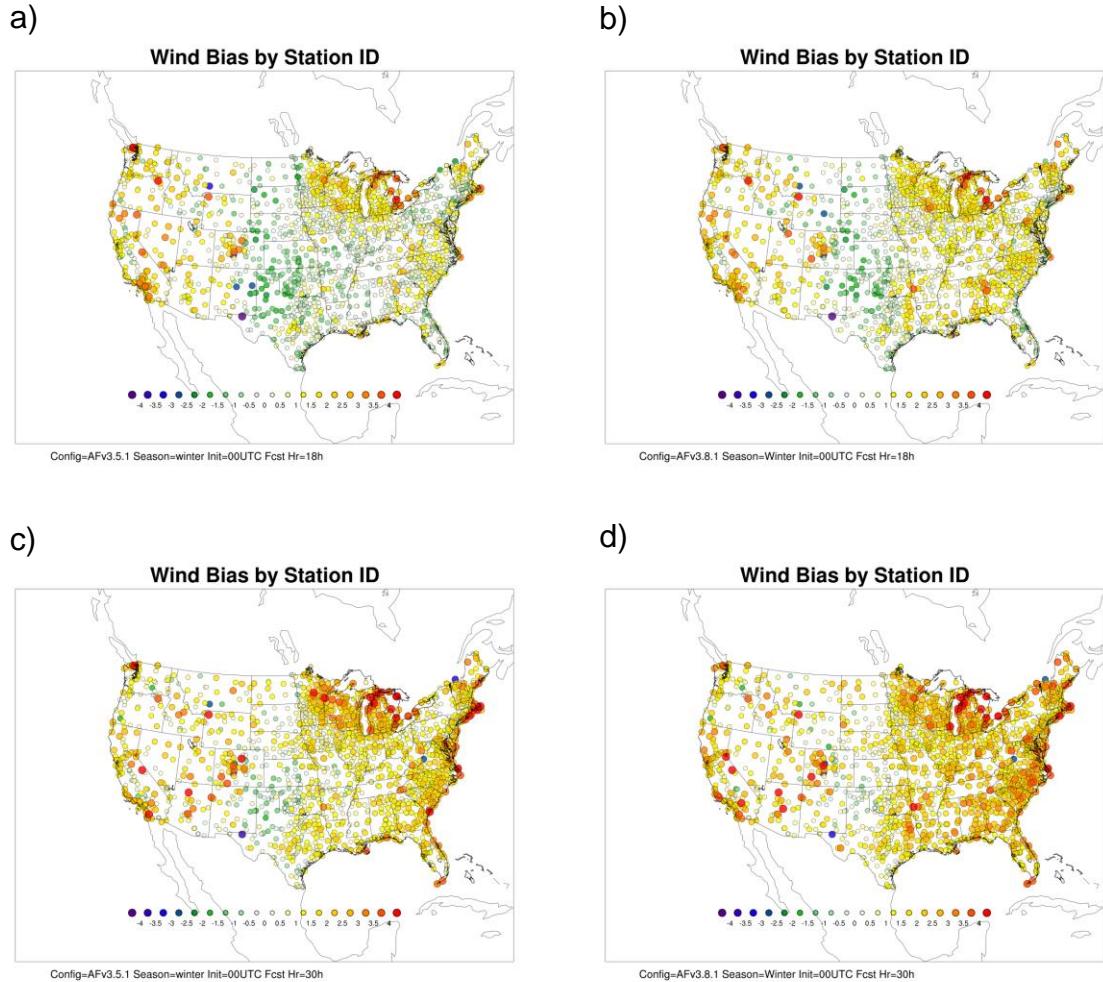
Figure 19. Spatial plot of 10 m wind speed (m s^{-1}) bias by observation station for all 00 UTC initializations in the winter aggregation for (a) AFv3.5.1 at the 18 h forecast lead time, (b) AFv3.8.1 at the 18 h forecast lead time, (c) AFv3.5.1 at the 30 h forecast lead time, and (d) AFv3.8.1 at the 30 h forecast lead time.

For summer and winter, both configurations have a maximum in upper level wind speed BCRMSE values between 300 and 200 hPa, which also serves as the vertex point for a shift from generally increasing to decreasing values (*Figure 20*a,b). The BCRMSE also tends to increase with forecast lead time. The majority of the pair-wise differences were not SS, and none were PS (*Table 6*). When SS differences were noted, they favored AFv3.8.1 at most levels except 200 and 150 hPa where they favored AFv3.5.1.
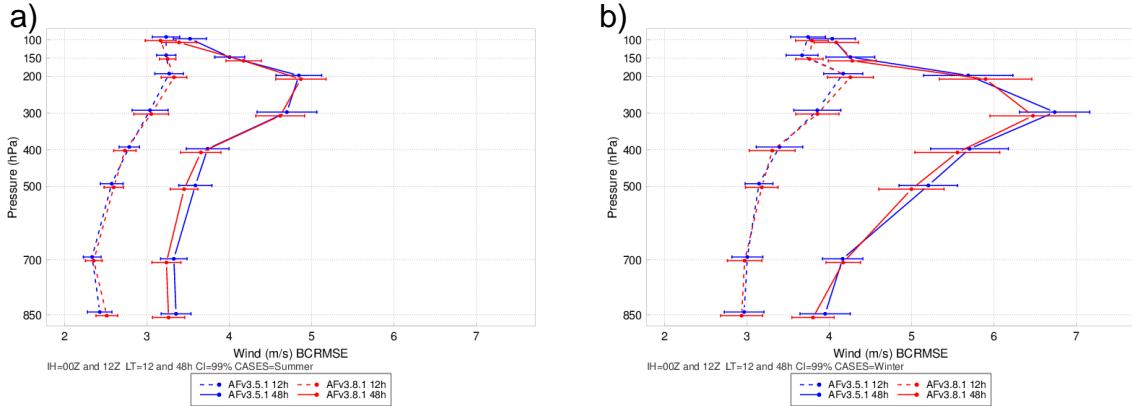
Figure 20. Vertical profile of the median wind speed (m s⁻¹) BCRMSE for the 12- (dashed) and 48-h (solid) forecast lead times aggregated across the CONUS domain over the (a) summer and (b) winter cases. AFv3.5.1 is in blue and AFv3.8.1 in red. The horizontal bars attached to the median represent the 99% CIs.

For both versions and referencing all the upper air wind profiles (*Figure 21*), the wind speed bias was generally negative during the summer but transitions between positive (lower- and upper-most levels) and negative (mid- to upper-levels) wind speed biases during winter. The SS differences generally favor AFv3.8.1 from 850 to 400 hPa and again at 100 hPa (*Table 6*). From 300 to 150 hPa, AFv3.5.1 is favored, and a few of these differences are PS for certain lead times and levels.
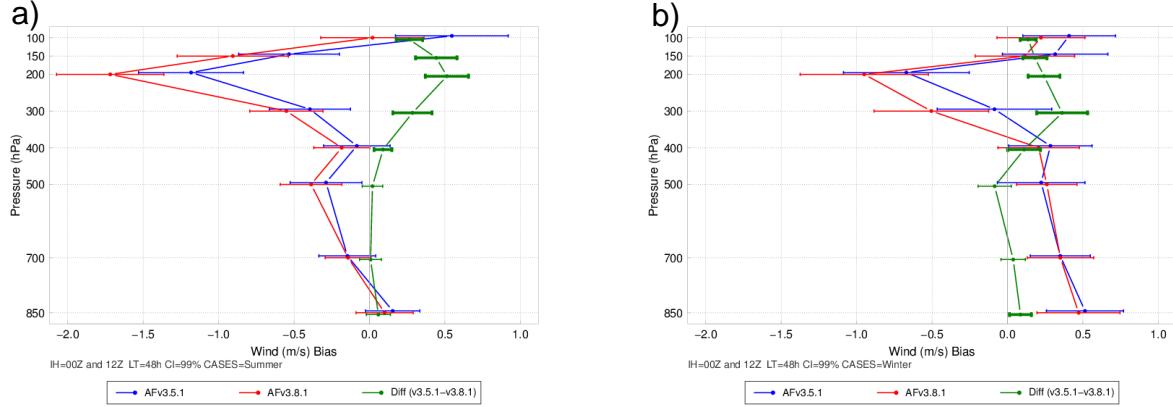
Figure 21. Vertical profile of the median wind speed (m s⁻¹) mean error (bias) at the 48-h lead time aggregated over CONUS region for the (a) summer and (b) winter cases. AFv3.5.1 is in blue, AFv3.8.1 is in red, and the difference (AFv3.5.1-AFv3.8.1) is in green. The horizontal bars attached to the median represent the 99% CIs; the thicker horizontal bars denote SS.

4.4 3-h Precipitation Accumulation GSS and Frequency Bias

In general, with increasing threshold, there is a decrease in 3 h accumulated precipitation GSS, which is more noticeable in the summer aggregation; this accompanies a decrease in the base rate, the ratio of total observed grid-box events to the total number of grid boxes summed over all cases (*Figure 22*a,b). A decrease in skill is also noted with increased forecast lead time (not shown). During the summer months, GSS is lower for both the East and West aggregations than during the winter.

Differences between the two configurations are small, with very few SS pair-wise differences, all favoring AFv3.8.1 (*Table 7*). All SS differences are located over the West region and are for thresholds >1, for which there are very few observed events.

When examining frequency bias for 3 h accumulated precipitation, an over-estimate is often seen at the first few thresholds, decreasing with increasing threshold (*Figure 22*c,d). During the summer, this tends to lead to an under-estimate at the higher thresholds, while during the winter, the decrease is more subtle and the aggregate bias often remains above 1; it is worth noting that the CIs, especially in the winter season are often wide and encompass 1. AFv3.8.1 is most often seen as having the lower bias; however, very few differences are SS, likely due to the large width of the CIs (*Table 7*). All but one of the SS differences occur in the summer aggregation at mid-thresholds for the East and CONUS regions (which also shared similar plot characteristics).
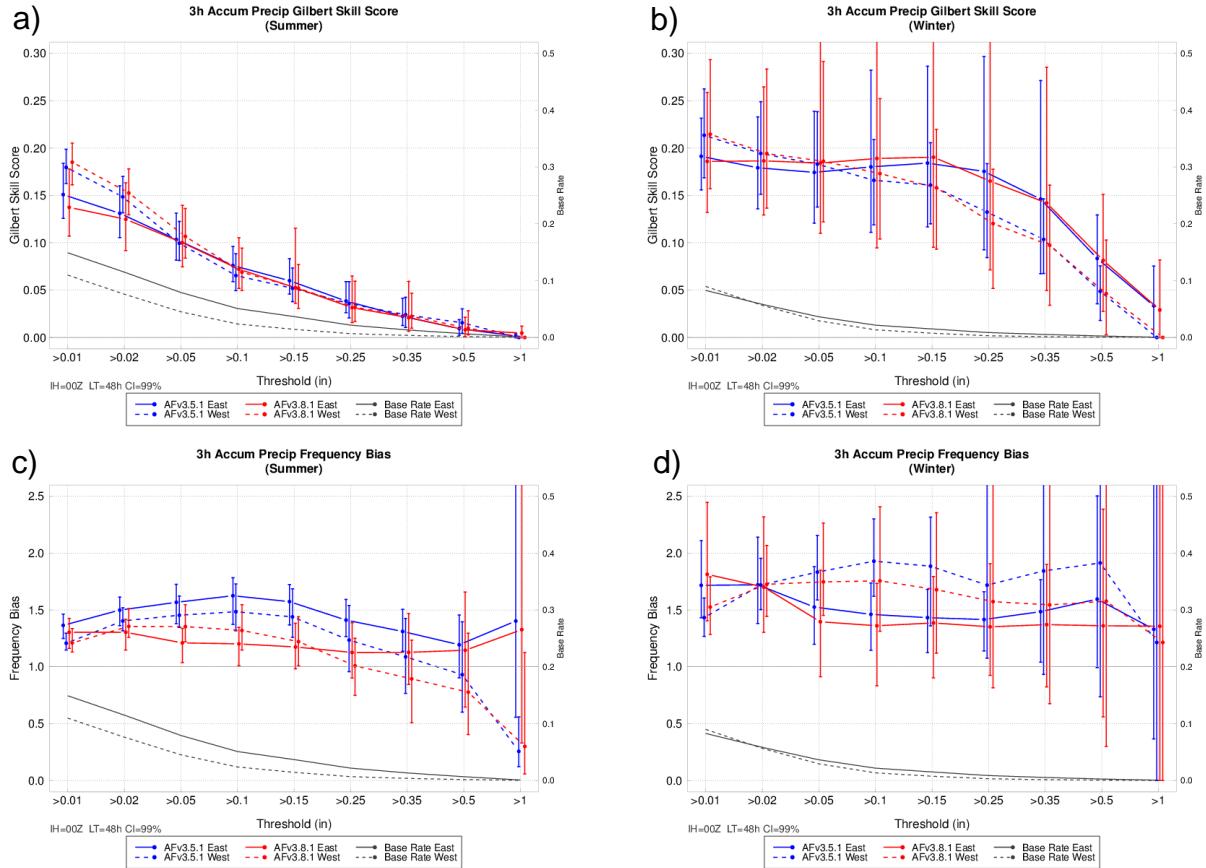
Figure 22. Threshold plots of 3-hrly accumulated precipitation (in) for median GSS (top row) and frequency bias (bottom row) for the 00 UTC initializations and the 48 h forecast aggregated across the (left column) summer and (right column) winter cases for the East (solid) and West (dashed) verification regions. AFv3.5.1 is in blue, AFv3.8.1 is in red, and base rate is in gray. The vertical bars attached to the median represent the 99% CIs.

4.5 Daily Accumulated Precipitation GSS and Frequency Bias

Similar to the 3 h QPF, the daily accumulated precipitation generally shows a decrease in GSS with an increase in threshold and forecast lead time, with exceptions in the East for the 36-hour forecast at the larger thresholds; a decrease in base rate is also observed, with the most dramatic decrease seen between the first two thresholds (*Figure 23*a,b). Similar to 3 h QPF GSS, the summer aggregation has lower skill than during the winter, at the mid-range thresholds, in particular. Differences between the two configurations are small with very few being SS and favoring AFv3.5.1 (*Table 8*).

During the summer, daily-accumulated QPF frequency bias is relatively unbiased in the West for all thresholds (*Figure 23*c). In the East, aggregate frequency bias and accompanying CIs for AFv3.5.1 generally indicate a high bias while AFv3.8.1 is more frequently unbiased, especially for the mid to lower thresholds. During the winter, the bias generally remains above 1 for all thresholds (*Figure 23*d); as threshold increases, the width of the CIs also increase. Minimal SS pair-wise difference are observed typically occurring during the summer at the mid-thresholds and favoring AFv3.8.1, with one exception (*Table 8*).
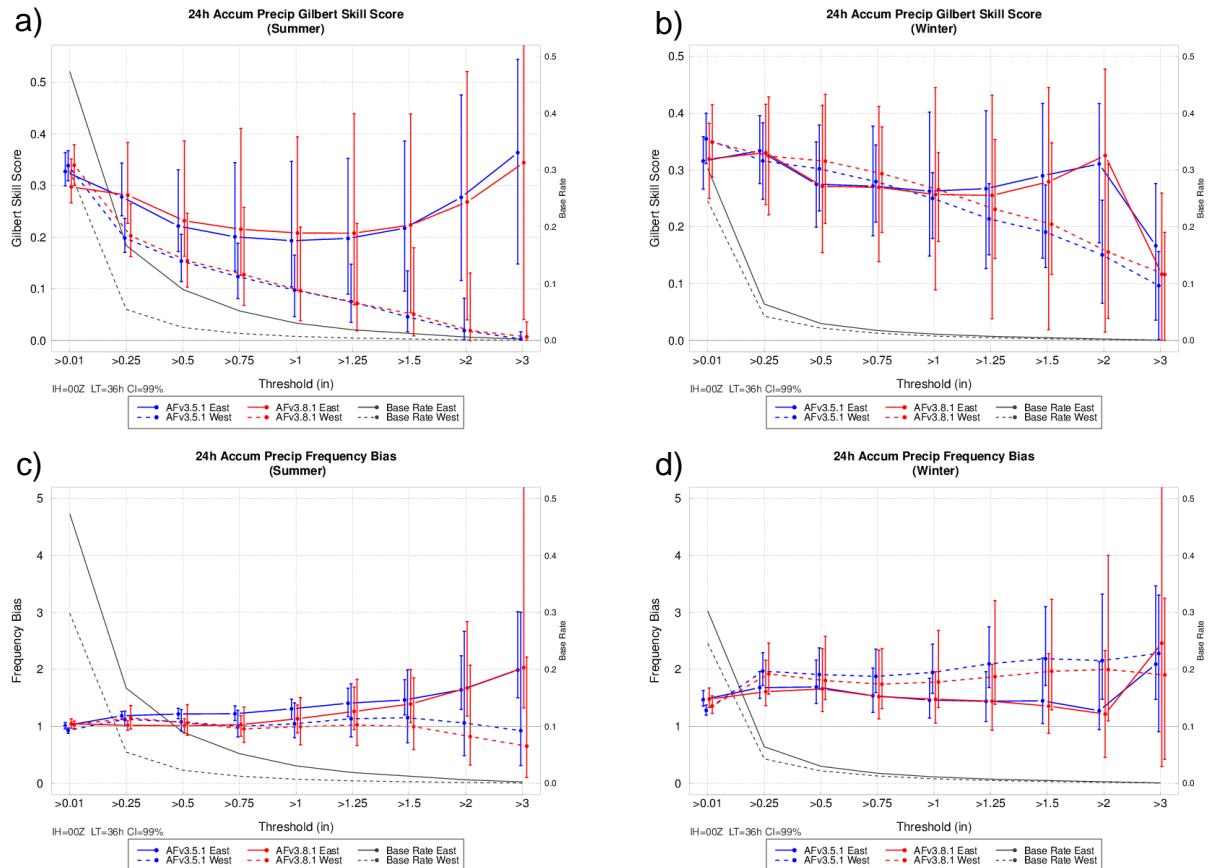
Figure 23. Threshold plots of daily-accumulated precipitation (in) for median GSS (top row) and frequency bias (bottom row) for the 00 UTC initializations and the 36 h forecast aggregated across the (left column) summer and (right column) winter cases for the East (solid) and West (dashed) verification regions.

4.6 GO Index

For both 00 and 12 UTC initializations during the summer and winter seasons, the median GO Index values and associated CIs (estimated by the width of the notches about the median on the boxplot) are less than 1, indicating AFv3.5.1 is the better performer (*Figure 24*a). Given the results discussed above, the components of the GO Index were further investigated to determine which variables and levels may be influencing the degradation of the AFv3.8.1 performance in the GO Index calculation. *Figure 24*b shows the GO Index by season and initialization without including the RMSE and associated weights for 400 hPa height. By excluding 400 hPa height from the GO Index calculation, a reversal in performance is seen at all permutations except the 00 UTC initializations during the winter, where no SS difference in performance is seen. While the GO Index is a fixed calculation, by investigating the components, it can help pinpoint strengths and weaknesses of the configurations; in the case of AFv3.8.1, weaknesses are seen with the 400 hPa height errors.
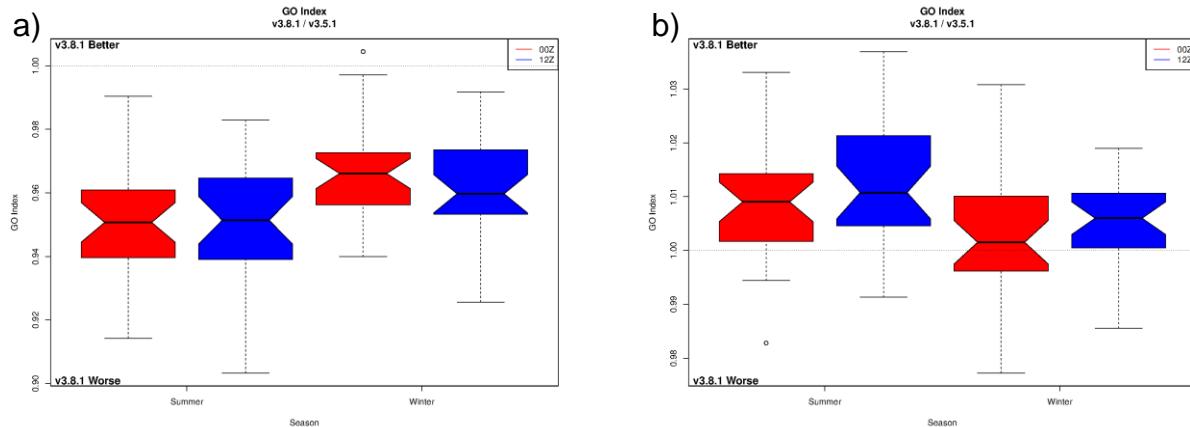
Figure 24. Boxplot of GO Index values aggregated across the summer and winter season, stratified by initialization time, where 00 UTC is in red and 12 UTC is in blue for (a) the standard GO Index calculation and (b) the GO Index calculation without 400hPa height. The median value is the thick black line 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.

5. Summary

An end-to-end sensitivity test was conducted to test and evaluate forecast performance when substituting the Air Force's current operational configuration using WRF v3.5.1 with a proposed configuration for v3.8.1. Changes were made to physics parameterizations (i.e., microphysics, radiation and surface layer schemes), along with other various parameters (e.g., eta levels, time step) in the configuration settings. Each configuration was run over the same set of cases spanning three months each during the summer and winter of 2011-12, and a full objective verification analysis was

performed comparing the two configurations. The goal of this testing was to assess the proposed configuration as part of the process for consideration of implementation into the AF's operational NWP infrastructure.

The testing methodology allowed for pair-wise differences to be computed, including an assessment of both SS and PS pair-wise differences. As was expected with substantial changes to the configuration, a large number of SS and PS pair-wise differences were observed at both the surface and upper air. A number of PS pair-wise differences were seen in 2 m temperature and dew point temperature; very few of the differences in 10 m wind speed were PS. For 2 m temperature, BCRMSE results favored AFv3.8.1 regardless of region. A cold temperature bias at 2 m was generally observed for both configurations; however, AFv3.8.1 was the preferred configuration in the East and in the West, with a few exceptions. For 2 m dew point temperature, PS pair-wise differences in BCRMSE favored AFv3.8.1 in the East and AFv3.5.1 in the West, while PS pair-wise differences in bias generally favored AFv3.8.1 for all regions. While very few PS pair-wise differences were noted for 10 m wind speed, the SS pair-wise differences in BCRMSE favored AFv3.8.1, while SS pair-wise differences in bias favored AFv3.5.1. In terms of upper air verification results, temperature BCRMSE had few PS pair-wise differences but generally favored AFv3.8.1. While upper-air temperature bias showed AFv3.8.1 as the preferred configuration over the West during the summer; results were more mixed over the West during the winter and in the East for both seasons. Upper-air dew point temperature had very few PS pair-wise differences in BCRMSE; however, bias results showed AFv3.5.1 was favored in the West, while in the East, AFv3.8.1 was favored during the summer with more mixed results in the winter. With upper-air wind, there were no PS differences for BCRMSE. Wind bias differences generally favored AFv3.5.1 in the summer with some PS pair-wise differences, while winter was mixed with only one instance of PS pair-wise differences.

When examining the GO Index, AFv3.5.1 was shown as the better performer for both the 00 and 12 UTC initializations during the summer and winter seasons. Based on the overall results observed in the analysis, an investigation into the cause of the AFv3.8.1 degradation as determined by the GO Index was conducted and showed that removal of RMSE for 400 hPa height from the calculation resulted in a reversal of performance, where AFv3.8.1 was preferred for all but the 00 UTC winter aggregation.

6. References

Iacono, M.J., J.S. Delamere, E.J. Mlawer, M.W. Shephard, S.A. Clough, and W.D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113, D13103.

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.

Thompson, G. and T. Eidhammer, 2014: A Study of Aerosol Impacts on Clouds and Precipitation Development in a Large Winter Cyclone. *J. Atmos. Sci.,* **71**, 3636–3658,doi: 10.1175/JAS-D-13-0305.1.

a) CONUS

		Surface Temperature	f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *
		Winter	v3.8.1	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
	12 UTC Inits	Summer	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1	v3.8.1 *	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	v3.5.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--
Bias	00 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
	12 UTC Inits	Summer	v3.5.1 *	v3.5.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *

b) CONUS-West

		Surface Temperature	f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits	Summer	v3.8.1 *	v3.8.1	v3.8.1	v3.5.1	--	v3.8.1 *	v3.8.1	v3.8.1	v3.8.1 *	v3.8.1 *	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	--	--	--	--	v3.8.1	v3.8.1 *	v3.8.1	v3.8.1	--	--	--	--	--	v3.8.1 *	--
	12 UTC Inits	Summer	--	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.5.1	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--
		Winter	--	v3.8.1	v3.8.1 *	v3.8.1	--	--	--	--	--	--	--	--	--	--	v3.5.1 *	v3.5.1
Bias	00 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
	12 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *
		Winter	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *

c) CONUS-East

		Surface Temperature	f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *
		Winter	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
	12 UTC Inits	Summer	v3.8.1 *	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--
Bias	00 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *
		Winter	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
	12 UTC Inits	Summer	--	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *

Table 1. SS (light shading) and PS (dark shading) pair-wise differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (where the highlighted configuration is favored) for **2 m temperature** BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

a) CONUS

Surface Dew Point Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits	Summer	--	v3.8.1	v3.8.1	--	--	v3.8.1	v3.8.1	v3.8.1	--	v3.8.1	v3.8.1 *	v3.8.1 *	--	v3.8.1	v3.8.1	--
		Winter	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	--	--	--	--	v3.8.1	v3.8.1	v3.8.1 *	v3.8.1 *	--	v3.8.1	v3.8.1	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	--	--	--	--	--	v3.5.1 *	v3.5.1 *	v3.5.1 *	--	--	--	--	v3.5.1 *	v3.5.1 *	v3.5.1 *	--
Bias	00 UTC Inits	Summer	v3.8.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
	12 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	--	v3.8.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--

b) CONUS-West

Surface Dew Point Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits	Summer	v3.5.1 *	v3.5.1 *	v3.5.1 *	--	--	--	--	--	v3.5.1 *	v3.5.1 *	v3.5.1 *	--	--	--	--	--
		Winter	--	--	--	--	--	--	--	--	--	--	--	v3.5.1 *	v3.5.1 *	--	--	--
	12 UTC Inits	Summer	v3.5.1 *	v3.5.1 *	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1 *	--	--	--	--	--	--	--	--	--
		Winter	--	--	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	--	--	--	v3.5.1 *	--	--	v3.5.1 *	v3.5.1 *
Bias	00 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.5.1	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.8.1 *	--	--	--	--
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.5.1 *	v3.8.1 *
	12 UTC Inits	Summer	v3.5.1	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	--	--	--	v3.8.1 *	--	v3.5.1 *	v3.8.1 *
		Winter	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *

c) CONUS-East

Surface Dew Point Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits	Summer	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--
		Winter	v3.8.1 *	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	v3.8.1	--	v3.8.1	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	00 UTC Inits	Summer	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--	--	--	--
	12 UTC Inits	Summer	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
		Winter	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	--	--

Table 2. SS (light shading) and PS (dark shading) pair-wise differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (where the highlighted configuration is favored) for **2 m dew point** BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

a) CONUS

Surface Wind Speed				f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits		Summer	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	v3.8.1	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	v3.8.1	--
			Winter	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--
	12 UTC Inits		Summer	--	v3.8.1	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1
			Winter	--	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	v3.8.1	v3.8.1	--	v3.8.1
Bias	00 UTC Inits		Summer	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	--	v3.5.1
			Winter	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1
	12 UTC Inits		Summer	v3.5.1	--	--	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	--	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1
			Winter	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1 *	v3.5.1	v3.5.1

b) CONUS-West

Surface Wind Speed				f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits		Summer	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--
			Winter	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--
	12 UTC Inits		Summer	--	--	--	--	--	--	v3.8.1	v3.8.1	--	--	--	--	--	--	v3.8.1	v3.8.1
			Winter	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1
Bias	00 UTC Inits		Summer	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	v3.8.1	v3.8.1
			Winter	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	--	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	v3.5.1
	12 UTC Inits		Summer	--	--	--	v3.8.1	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.8.1	v3.8.1	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1
			Winter	v3.5.1	v3.5.1	--	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1

c) CONUS-East

Surface Wind Speed				f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Inits		Summer	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	v3.8.1	v3.8.1
			Winter	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	v3.8.1
	12 UTC Inits		Summer	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1
			Winter	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	v3.5.1	v3.8.1	v3.8.1	--	v3.8.1	--
Bias	00 UTC Inits		Summer	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	--	v3.5.1
			Winter	v3.5.1 *	v3.5.1 *	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1	v3.5.1	v3.5.1 *
	12 UTC Inits		Summer	--	--	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	v3.8.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.5.1
			Winter	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1	v3.5.1

Table 3. SS (light shading) and PS (dark shading) pair-wise differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (where the highlighted configuration is favored) for **10 m wind speed** BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

a) CONUS

Upper Air Temperature		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	v3.8.1	--	v3.8.1 *	v3.8.1	v3.8.1	--	v3.8.1 *
	700	--	v3.8.1	v3.8.1 *	v3.8.1 *	--	--	--	--
	500	--	--	--	--	--	--	v3.8.1	--
	400	--	v3.5.1	v3.5.1	--	--	--	--	v3.8.1
	300	--	v3.5.1	--	--	--	v3.8.1	v3.8.1	v3.8.1 *
	200	--	--	--	--	--	v3.8.1	--	--
	150	--	--	--	--	--	v3.5.1	v3.5.1	v3.5.1
	100	v3.8.1	--	--	--	--	--	--	--
Bias	850	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.5.1	v3.5.1 *	v3.5.1
	700	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1 *	v3.8.1 *	v3.5.1	--	v3.5.1 *
	500	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1	v3.5.1	v3.5.1 *	--
	400	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.8.1	--	--	--
	300	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1	--	--	--
	200	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
	150	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.5.1	v3.8.1	v3.5.1	--
	100	v3.8.1	v3.5.1	--	--	--	v3.5.1	v3.5.1 *	v3.5.1 *

Table 4. SS (light shading) and PS (dark shading) pair-wise differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (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 a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

b) CONUS-West

Upper Air Temperature		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	--	--	v3.8.1 *	v3.8.1 *	--	--
	700	--	--	v3.8.1	v3.8.1	--	v3.8.1	--	--
	500	--	--	--	--	--	--	--	--
	400	--	--	--	--	--	v3.8.1	--	--
	300	--	--	--	--	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *
	200	--	--	v3.8.1	--	--	--	--	--
	150	--	--	--	--	--	v3.5.1	--	v3.5.1 *
	100	v3.8.1	--	v3.8.1	--	--	--	v3.8.1	--
Bias	850	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	--
	700	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1	--	v3.5.1 *
	500	v3.8.1	v3.8.1	v3.8.1	--	v3.5.1	v3.5.1 *	--	--
	400	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.8.1 *	--	--	--	--
	300	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	--	--
	200	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
	150	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1	--
	100	v3.8.1 *	v3.8.1	v3.8.1 *	v3.8.1 *	--	v3.5.1	v3.5.1	v3.5.1 *

c) CONUS-East

Upper Air Temperature		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	v3.8.1	v3.8.1	v3.8.1 *	--	--	--	--
	700	--	--	v3.8.1	v3.8.1 *	--	--	--	--
	500	--	--	--	--	--	--	v3.8.1	v3.8.1
	400	--	--	--	--	--	--	--	v3.8.1 *
	300	--	--	--	--	--	v3.8.1	v3.8.1	v3.8.1
	200	--	--	--	v3.8.1	v3.8.1	--	--	--
	150	--	--	--	--	--	--	--	v3.5.1
	100	v3.8.1	--	--	--	--	--	--	--
Bias	850	--	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *
	700	v3.8.1	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1	--	--	--
	500	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1	--	v3.5.1	--
	400	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	--	--
	300	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	--	--	v3.8.1 *	v3.8.1 *
	200	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.8.1 *
	150	v3.8.1 *	--	--	--	v3.5.1	--	--	--
	100	v3.8.1	v3.5.1 *	v3.5.1 *	v3.5.1 *	--	v3.5.1	v3.5.1 *	v3.5.1 *

a) CONUS

Upper Air Dew Point Temperature		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	--	--	v3.8.1 *	--	--	v3.8.1 *
	700	--	--	--	--	v3.8.1 *	--	--	--
	500	--	--	--	--	--	--	--	--
Bias	850	--	--	v3.8.1 *	v3.8.1 *	--	v3.5.1 *	v3.5.1 *	v3.8.1 *
	700	--	v3.8.1 *	v3.8.1 *	v3.8.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *
	500	v3.5.1 *	v3.5.1 *	--	--	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *

Table 5. SS (light shading) and PS (dark shading) pair-wise differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (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 a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

b) CONUS-West

Upper Air Dew Point Temperature		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	--	--	--	--	--	--
	700	--	--	--	--	--	--	--	--
	500	--	--	--	--	--	--	--	--
Bias	850	v3.8.1 *	--	--	--	--	--	v3.5.1 *	v3.5.1 *
	700	--	--	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *
	500	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *	v3.5.1 *

c) CONUS-East

Upper Air Dew Point Temperature		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	--	--	v3.8.1 *	--	--	v3.8.1 *
	700	--	--	--	--	--	--	--	--
	500	--	--	--	--	--	--	--	--
Bias	850	--	--	v3.8.1 *	v3.8.1 *	--	v3.8.1 *	v3.8.1 *	v3.8.1 *
	700	--	v3.8.1 *	--	--	--	--	--	--
	500	--	v3.8.1 *	v3.8.1 *	--	--	v3.5.1 *	v3.5.1 *	v3.5.1 *

a) CONUS

Upper Air Wind Speed		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	v3.8.1	v3.8.1	v3.8.1	--	v3.8.1	v3.8.1
	700	--	--	--	--	--	--	--	--
	500	--	--	--	--	--	--	v3.8.1	v3.8.1
	400	--	--	--	--	--	--	v3.8.1	--
	300	v3.8.1	--	v3.8.1	v3.8.1	--	--	--	--
	200	--	--	--	--	v3.5.1	--	v3.5.1	--
	150	--	v3.5.1	v3.5.1	--	v3.5.1	--	v3.5.1	v3.5.1
	100	--	--	--	--	--	--	--	--
Bias	850	--	v3.8.1	--	--	v3.8.1	--	v3.8.1	v3.8.1
	700	--	--	--	--	--	--	--	--
	500	v3.8.1	--	--	--	v3.8.1	v3.8.1	--	--
	400	v3.8.1	--	--	v3.5.1	v3.8.1	--	v3.5.1	v3.8.1
	300	v3.5.1	v3.5.1	v3.5.1	v3.5.1	--	v3.5.1	v3.5.1	v3.5.1 *
	200	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1	v3.5.1	v3.5.1	v3.5.1
	150	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.8.1	v3.5.1	v3.8.1	v3.8.1
	100	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.5.1	--	v3.8.1

Table 6. SS (light shading) and PS (dark shading) pair-wise differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (where the highlighted configuration is favored) for **upper air wind speed** BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

b) CONUS-West

Upper Air Wind Speed		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	--	--	--	--	v3.8.1	v3.8.1
	700	--	--	--	--	--	--	--	--
	500	v3.5.1	--	--	--	--	--	--	--
	400	--	--	--	--	--	v3.8.1	--	--
	300	--	--	--	v3.8.1	--	--	--	--
	200	--	--	--	--	--	--	--	--
	150	--	--	v3.5.1	v3.5.1	--	v3.5.1	--	v3.5.1
	100	--	--	--	--	--	--	--	--
Bias	850	v3.8.1	--	v3.5.1	--	--	v3.8.1	v3.8.1	--
	700	--	--	v3.8.1	--	v3.5.1	--	--	v3.5.1
	500	v3.5.1	--	--	--	v3.8.1	v3.8.1	--	--
	400	--	--	--	--	--	--	--	v3.8.1
	300	--	v3.5.1	v3.5.1	v3.5.1	--	v3.5.1	v3.5.1	v3.5.1
	200	--	v3.5.1	v3.5.1	v3.5.1 *	--	--	--	--
	150	v3.5.1	v3.5.1	v3.5.1	v3.5.1	v3.8.1	--	--	v3.8.1
	100	--	--	v3.8.1	v3.8.1	v3.8.1	--	--	v3.8.1

c) CONUS-East

Upper Air Wind Speed		Summer				Winter			
		f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	v3.8.1
	700	--	v3.8.1	--	--	--	--	v3.8.1	--
	500	--	--	--	--	--	--	--	--
	400	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1
	300	--	--	v3.8.1	--	--	--	--	v3.8.1
	200	--	--	--	--	--	v3.5.1	v3.5.1	--
	150	--	v3.5.1	--	--	--	--	--	--
	100	v3.8.1	--	--	--	--	--	--	--
Bias	850	v3.5.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	--
	700	--	--	--	--	--	v3.8.1	--	--
	500	--	--	--	--	v3.8.1	--	--	--
	400	v3.8.1	--	--	v3.5.1	--	--	--	v3.8.1
	300	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	--	v3.5.1	v3.5.1	v3.5.1 *
	200	v3.5.1	v3.5.1	v3.5.1 *	v3.5.1 *	v3.5.1	--	v3.5.1	v3.5.1
	150	v3.5.1	v3.5.1	v3.5.1	v3.5.1 *	v3.8.1	v3.5.1	v3.5.1	v3.5.1
	100	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1

a) CONUS

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	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Frequency Bias	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--
		f24	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

b) CONUS-West

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	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	v3.8.1
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	v3.8.1	--	--	--	--	--	--	--	--	v3.8.1
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Frequency Bias	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	v3.8.1	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

c) CONUS-East

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	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Frequency Bias	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--
		f24	--	--	v3.8.1	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table 7. SS differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (where the highlighted configuration is favored) for **3 h QPF** GSS and frequency bias by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

a) CONUS

		Daily QPF		>0.01	>0.25	>0.5	>0.75	>1	>1.25	>1.5	>2	>3
Gilbert Skill Score	00 UTC Inits	Summer	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	f24	v3.5.1	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
Frequency Bias	00 UTC Inits	Summer	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	f24	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--
			f48	v3.5.1	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--

b) CONUS-West

		Daily QPF		>0.01	>0.25	>0.5	>0.75	>1	>1.25	>1.5	>2	>3
Gilbert Skill Score	00 UTC Inits	Summer	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
Frequency Bias	00 UTC Inits	Summer	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	f24	--	--	--	--	--	--	--	--	--
			f48	v3.5.1	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--

c) CONUS-East

		Daily QPF		>0.01	>0.25	>0.5	>0.75	>1	>1.25	>1.5	>2	>3
Gilbert Skill Score	00 UTC Inits	Summer	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	f24	v3.5.1	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	v3.5.1
Frequency Bias	00 UTC Inits	Summer	f36	--	v3.8.1	v3.8.1	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
	12 UTC Inits	Summer	f24	--	v3.8.1	v3.8.1	v3.8.1	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--

Table 8. SS differences for the AF configuration run with WRF v3.5.1 and WRF v3.8.1 (where the highlighted configuration is favored) for **24 h QPF** GSS and frequency bias by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the a) CONUS, b) CONUS-West, and c) CONUS-East verification domain.

Appendix A: Subset of the WRF v3.8.1 *namelist.input* used for this test

```
&time_control
run_hours                 = 48,
interval_seconds          = 10800,
history_interval          = 180,
frames_per_outfile        = 1,
restart                   = .false.,
io_form_history           = 2,
/

&fdda
/

&domains
time_step                 = 60,
time_step_fract_num       = 0,
time_step_fract_den       = 1,
max_dom                   = 1,
e_we                      = 656,
e_sn                      = 464,
e_vert                    = 57,
num_metgrid_levels        = 27,
num_metgrid_soil_levels   = 4,
dx                        = 15000,
dy                        = 15000,
p_top_requested           = 1000,
interp_type               = 2,
lowest_lev_from_sfc       = .false.,
lagrange_order            = 1,
force_sfc_in_vinterp      = 6,
zap_close_levels          = 500,
adjust_heights            = .false.,
eta_levels        = 1.0000, 0.9970, 0.9920, 0.9850, 0.9780, 0.9690, 0.9600, 0.9500,
                    0.9380, 0.9250, 0.9100, 0.8940, 0.8760, 0.8570, 0.8350, 0.8120,
                    0.7870, 0.7600, 0.7310, 0.7000, 0.6680, 0.6350, 0.6000, 0.5650,
                    0.5300, 0.4940, 0.4580, 0.4108, 0.3674, 0.3278, 0.2914, 0.2582,
                    0.2278, 0.2004, 0.1752, 0.1526, 0.1324, 0.1148, 0.0992, 0.0856,
                    0.0738, 0.0634, 0.0542, 0.0464, 0.0394, 0.0332, 0.0278, 0.0230,
                    0.0190, 0.0154, 0.0122, 0.0094, 0.0070, 0.0048, 0.0030, 0.0014,
                    0.0000,
rh2qv_method              = 2,
/
```

```
&physics
mp_physics              = 28,
ra_lw_physics           = 4,
ra_sw_physics           = 4,
radt                    = 30,
sf_sfclay_physics       = 1,
sf_surface_physics      = 2,
bl_pbl_physics          = 1,
bldt                    = 0,
cu_physics              = 1,
cudt                    = 5,
icloud                  = 3,
aer_opt                 = 3,
swint_opt               = 1,
ysu_topdown_pblmix      = 1,
surface_input_source    = 1,
num_soil_layers         = 4,
mp_zero_out             = 0,
num_land_cat            = 21,
fractional_seaice       = 1,
seaice_threshold        = 0.0,
tice2tsk_if2cold        = .true.,
do_radar_ref            = 1,
use_aero_icbc           = .true.,
/

&dynamics
rk_ord                  = 3,
diff_6th_opt            = 2,
diff_6th_factor         = 0.10,
w_damping               = 1,
diff_opt                = 2,
km_opt                  = 4,
damp_opt                = 3,
zdamp                   = 5000.,
dampcoef                = 0.20,
khdif                   = 0,
kvdif                   = 0,
smdiv                   = 0.1,
emdiv                   = 0.01,
epssm                   = 0.5,
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           = 1,
chem_adv_opt             = 0,
tke_adv_opt              = 0,
use_baseparam_fr_nml     = .true.,
/


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

Appendix B: Original settings in the WRF v3.5.1 *namelist.input* that were modified when testing v3.8.1 (as described in Appendix A).

```
&domains
time_step                = 90,
interp_type              = 1,
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_physics               = 4,
ra_lw_physics            = 1,
ra_sw_physics            = 1,
icloud                   = 1,
aer_opt                  [not set for v3.5.1]
swint_opt                [not set for v3.5.1]
ysu_topdown_pblmix       [not set for v3.5.1]
num_land_cat             [not set for v3.5.1]
seaice_threshold         = 271,
```

```
do_radar_ref              [not set for v3.5.1]
use_aero_icbc             [not set for v3.5.1]
/


 &dynamics
diff_opt                  = 1,
dampcoef                  = 0.05,
epssm                     = 0.1,
scalar_adv_opt            = 0,
/
```