The Developmental Testbed Center WRFv3.1.1+ QNSE Test and Evaluation Final Report

Point of Contact: Jamie Wolff
April 30, 2010
Updated: August 22, 2011

Executive Summary

The Weather Research and Forecasting (WRF) model is a state-of-the-art numerical weather prediction system that is highly configurable and suitable for a broad range of weather applications. Given the numerous options available, it is important to rigorously test configurations to assess the performance of select configurations for specific applications. To assess the performance of the newly available Quasi-Normal Scale Elimination (QNSE) PBL and surface layer schemes in WRF, the Developmental Testbed Center (DTC) performed testing and evaluation with the Advanced Research WRF (ARW) dynamic core for two physics suite configurations at the request of the sponsor, the Air Force Weather Agency (AFWA). One configuration was based on AFWA's Operational Configuration, which now provides a baseline for testing and evaluating new options available in the WRF system. The second configuration substituted AFWA's current operational PBL and surface layer schemes with the newly available QNSE planetary boundary layer (PBL) and surface layer schemes. 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). A coding error was discovered after the comprehensive testing was complete related to the 2m and 10m diagnostic fields, the impact of which is difficult to assess until a rerun can be conducted. In spite of this, surface fields were evaluated and discussed in this report. The following points summarize the SS differences seen in the verification results between the AFWA and QNSE configurations:

- In general, the relative magnitudes of the SS pair-wise differences favoring the AFWA configuration are larger than those favoring the QNSE configuration.
- For vertical profiles of temperature BCRMSE, the SS pair-wise differences at and below 400 hPa and above 150 hPa favor the AFWA configuration, while those at 200 and 300 hPa favor the QNSE configuration; for bias the QNSE configuration is favored at 850, 700, and 200 hPa, while those between 500 and 300 hPa favor the AFWA configuration.
- All SS pair-wise differences for vertical profiles of dew point temperature indicate the AFWA configuration has a lower BCRMSE, while for bias the SS pair-wise differences generally favor the QNSE configuration.
- Similar to dew point temperature, the vertical profile for wind speed BCRMSE show that all SS pair-wise differences correspond to the AFWA configuration having smaller values, regardless of level, lead time or temporal aggregation; however, for bias the QNSE configuration is favored for all SS pair-wise differences above 700 hPa.
- For surface temperature BCRMSE and bias, nearly all SS pair-wise differences, regardless of lead time or temporal aggregation, favor the AFWA configuration.
- The SS pair-wise differences for surface dew point temperature BCRMSE tend to favor the AFWA configuration, with some exceptions during the summer aggregation; for bias the favored configuration is sensitive to initialization time, lead time, and season aggregation.
- The SS pair-wise differences of BCRMSE for surface wind speed generally favor the AFWA configuration, except around 00 UTC valid times, while for surface wind speed bias the pair-wise differences indicate that the QNSE configuration has SS smaller bias during the overnight hours and the AFWA configuration has SS smaller bias during the daytime hours.

- Overall, very few SS pair-wise differences are noted in the QPF analysis; with no exceptions, the AFWA configuration is favored in for both the 3-hour and 24-hour QPF.

1. Introduction

The Weather Research and Forecasting (WRF) model is a state-of-the-art numerical weather prediction system that is highly configurable and suitable for a broad range of weather applications. Given the numerous options available, it is important to rigorously test configurations to assess the performance of select configurations for specific applications. The Air Force Weather Agency (AFWA) is interested in improvements in the characterization of the planetary boundary layer (PBL) and surface layer. The Quasi-Normal Scale Elimination (QNSE) PBL and surface layer schemes developed by Sukoriansky, Galperin and Perov, (Sukoriansky et al. 2005) are new features available since WRF version 3.1 with the goal of addressing these issues. To assess the performance of these new schemes, the Developmental Testbed Center (DTC) performed testing and evaluation with the Advanced Research WRF (ARW) dynamic core (Skamarock et al. 2008) for two physics suite configurations at the request of the sponsor, AFWA. One configuration was based on AFWA's Operational Configuration, which now provides a baseline for testing and evaluating new options available in the WRF system. The second configuration substituted AFWA's current operational PBL and surface layer schemes with the QNSE schemes. Forecast verification statistics were computed for the two configurations and the analysis was based on the objective statistics of the model output. Both of these configurations will be designated DTC Reference Configurations (RCs) and the results made available to the WRF community.

2. Experiment Design

The end-to-end forecast system employed the WRF Preprocessing System (WPS), WRF, WRF Postprocessor (WPP) and graphics generation using NCL. Post-processed forecasts were verified using the Model Evaluation Tools (MET). In addition, the full data set was archived and made available for dissemination. The codes utilized were based on the official released versions of WPS (v3.1.1), WPP (v3.1), and MET (v3.0.1). Both WPP and MET included relevant bug fixes that were checked into the respective code repositories prior to testing. For WRF, a tag from the repository was also used, which was based on v3.1.1 with a considerable number of updates. The requirement to use code that was checked into the WRF repository ensures the code changes have been vetted through the WRF Developers Committee and run through regression testing. Tagged code is also easily traceable through time and reflects capabilities that will eventually be part of a public release, making these results more meaningful and relevant to the WRF user community.

2.1 Forecast Periods

Forecasts were initialized every 36 hours from 2 June 2008 through 31 May 2009, automatically creating a default of initialization times including both 00 and 12 UTC, for a total of 243 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 lists the forecast initializations that failed to complete the end-to-end process due to the noted reasons described in the table. All missing forecasts were due to missing or bad input data sets, not model crashes.

Missing forecasts:

Affected Case	Missing data	Reason
2008071000	WRF output	Missing SST input data
2008091512	WRF output	Bad SST input data
2008101512	WRF output	Bad SST input data
2008101700	WRF output	Bad SST input data
2008101812	WRF output	Bad SST input data
2008102112	WRF output	Missing AGRMET input data
2008121112	WRF output	Bad SST input data
2009030100	WRF output	Missing SST input data
2009040112	WRF output	Bad SST input data
2008042212	WRF output	Bad SST input data
2009052400	WRF output	Missing SST input data
2009052512	WRF output	Missing SST input data

Missing verification:

Affected Case	Missing data	Reason
2008071300	Incomplete sfc. and upper air verification beyond 33h	Missing Prepbufr data
2008071412	Incomplete sfc. and upper air verification before 24h	Missing Prepbufr data
2008101400	Incomplete sfc. and upper air verification beyond 39h	Missing Prepbufr data
2009012700	Incomplete sfc. and upper air verification for 24-27h	Missing Prepbufr data

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 Agricultural Meteorological Modeling (AGRMET) System was utilized for the lower boundary conditions (LoBCs) in addition to a daily, real-time sea surface temperature product from Fleet Numerical Meteorology and Oceanography Center (FNMOC), which was used to initialize the sea surface temperature (SST) field for the forecasts. Finally, the time-invariant components of the LoBCs (topography, soil and vegetation type etc.) were derived from United States Geological Survey (USGS) input data. WPS was run once and the same output data were used for initializing both configurations of the model.

2.3 Model Configuration Specifics

2.3.1 Domain Configuration

A 15-km contiguous U.S. (CONUS) grid was employed in this test. The domain (Fig. 1) 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 two physics suite configurations used for each model configuration in this test are described in the table below. The QNSE PBL and surface layer are two separate schemes; however, they are directly related and should be specified together. The model configuration based on AFWA's Operational Configuration will be referred to as AFWA, while the companion configuration will be referred to as QNSE.

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

Relevant portions of the *namelist.input* can be found in Appendix B.

2.4 Post-processing

The *wrfpost* program within WPP was used to destagger the forecasts, to generate derived meteorological variables, including mean sea level pressure, 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

Objective model verification statistics were generated using the MET package. MET is comprised of grid-to-point comparisons, which were utilized to compare gridded surface and upper-air model data to point observations, as well as grid-to-grid comparisons, which were 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 full domain, as well as the 14 sub-domains shown in Fig. 2. Only the full domain is described in detail for this report, however, all sub-domain results are available on the DTC website (http://verif.rap.ucar.edu/eval/afwa_rc_test/). In addition to the regional area 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 pair-wise verification statistics include zero the differences are not statistically significant. Because frequency bias is not amenable to a pair-wise difference comparison due to the nonlinear attributes of this metric, 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.

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, RAOBS, and buoy data) within the NCEP North American Data Assimilation System (NDAS) prepbufr files. Objective model verification statistics were then generated for surface (using METAR and buoy observations) and upper air (using RAOBS) temperature, dew point temperature, and wind. Because shelter-level variables are not realistic at the initial model 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 for the surface and upper air statistics of temperature, dew point temperature and wind 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 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. The observational datasets used were the NCEP Stage II analysis for the 3-hour accumulation and the NCEP/CPC daily gauge analysis for the 24-hour accumulation. Because the 24-hour accumulation observations are 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 verification metrics computed included the frequency bias and the equitable threat score, or Gilbert skill score, (GSS). In addition, two state-of-the-art verification techniques were also computed and evaluated in an exploratory manner; Method for Object-based Diagnostic Evaluation (MODE) and Fractional Skill Score (FSS). The results of these new techniques will not be discussed here, but are available on the DTC website. 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 RMSE for both the AFWA and QNSE configurations by:

$$S = 1 - \frac{(\mathrm{RMSE}_{\mathrm{QNSE}})^2}{(\mathrm{RMSE}_{\mathrm{AFWA}})^2}$$

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 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 AFWA configuration has higher skill and values greater than one indicate the QNSE configuration has higher skill.

4. Verification Results

Differences are computed between the two configurations by subtracting the QNSE configuration from the AFWA configuration. BCRMSE is always a positive quantity and a perfect score is zero. Given these properties, differences that are negative (positive) indicate the AFWA (QNSE) configuration has a lower BCRMSE. For GSS, the perfect score is one and the no-skill forecast is zero. Thus, if the pair-wise difference is positive (negative) the AFWA (QNSE) configuration has a higher GSS. The properties of bias (which has a perfect score of zero) and frequency bias (which has a perfect score of one) are not as conducive to generalized statements such as those that can be made for BCRMSE and GSS. Both of these metrics can have positive or negative values. Given this, when looking at the pair-wise differences it is important to also note the magnitude of the bias in relation to the perfect score for each individual configuration to know which configuration has a smaller bias. A breakdown of the configuration with statistically significant (SS) 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. Please note, 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

The overall distribution for temperature BCRMSE for both the AFWA and QNSE configurations show a minimum error between 500 and 300 hPa (Fig. 3). As expected, the BCRMSE increases with forecast lead time. The pair-wise differences for the annual aggregation at all forecast lead times indicate all SS differences at and below 400 hPa, as well as those at and above 150 hPa, favor the AFWA configuration (see Table 1). Conversely, the SS pair-wise differences at 200 and 300 hPa favor the QNSE configuration. It is worth noting, however, that the relative magnitudes of the SS pair-wise differences favoring the AFWA configuration are larger than those favoring the QNSE configuration. All of the seasonal aggregations have similar distributions, except for a tendency for the summer season to produce more SS pair-wise differences at mid-levels that favor the QNSE configurations (see Table 1).

Both configurations produce a temperature bias that transitions from cold at lower levels to warm at upper levels. The level at which this transition occurs varies slightly with lead time and the time period of the aggregation (Fig. 4). For the summer and fall aggregations, there is a general increase in bias with height. A distinct minimum is noted at 850 hPa for all temporal aggregations. The SS pair-wise differences (see Table 1) indicate the QNSE configuration tends to produce the smallest temperature bias at 850 hPa and 200 hPa, whereas the AFWA configuration tends to produce the smallest bias between 500 and 300 hPa. However for the winter season the number of SS pair-wise differences at mid-levels decreases markedly. The configuration with the smallest bias at 150 and 100 hPa depends on the aggregation period, with QNSE producing the smallest bias in summer and for earlier lead times in fall and AFWA producing the smallest bias in winter and spring.

4.1.2 Dew Point Temperature BCRMSE and bias

The dew point temperature BCRMSE increases as the pressure decreases for both configurations during all temporal aggregations except winter. BCRMSE gradually increases with increasing lead time (Fig. 5). All SS pair-wise differences for the pair-wise comparison correspond to the AFWA configuration having a lower BCRMSE. This result also holds for all of the seasonal aggregations (See Table 2).

Both configurations tend to produce a positive dew point temperature or moist bias at all levels and lead times for the annual aggregation (Fig. 6). The magnitude of the bias is fairly consistent and actually decreases slightly for the longer lead times. This trend generally holds for all seasons (not shown), except summer. For the summer aggregation, the AFWA configuration has no SS dew point temperature bias at 850 and 700 hPa for the 12 hour lead time, and at 850 hPa for the 48 hour lead time. The QNSE configuration has no SS dew point temperature bias at 850 hPa for the 12 hour lead time that transitions to a low, or dry, bias by the 48 hour lead time. The SS pair-wise differences generally favor the QNSE configuration, except for the summer aggregation for which the configuration with the smallest moisture bias depends on both the level and the lead time (see Table 2).

4.1.3 Wind BCRMSE and bias

The vertical distribution of wind speed BCRMSE for both configurations exhibits the same general properties for all lead times and temporal aggregations. The distribution increases to a maximum between 300 and 200 hPa and then decreases aloft (Fig. 7). All SS pair-wise differences correspond to the AFWA configuration having smaller errors then the QNSE configuration regardless of level, lead time or temporal aggregation (see Table 3). For the seasonal aggregations, the pressure level at which these SS pair-wise differences occur vary with lead time (see Table 3) and as the lead time increases the number of SS pair-wise differences also

increases across the different levels. Overall, the summer season has the fewest SS pair-wise differences of any season.

Vertical profiles of wind speed bias indicate the winds for the AFWA configuration are non-biased at 850 hPa, whereas the winds for the QNSE configuration are too strong (Fig. 8). The wind speed bias for both configurations transitions to winds that are too light at upper levels, namely at 200 hPa. For this metric, the QNSE configuration has a consistent SS bias towards higher wind speeds as compared to the AFWA configuration at all levels below 400 hPa. This translates to the QNSE configuration having SS smaller bias when the overall wind speed bias is too light (at and above 700 hPa) and the AFWA configuration having SS smaller bias at levels where the overall wind speed bias is too fast (generally, below 850 hPa). This relationship holds for the seasonal aggregations (not shown), with only a few exceptions by season, forecast lead time, and level (see Table 3).

4.2 Surface

Following the completion of the extensive testing on the comprehensive set of cases, the developers of the QNSE scheme uncovered a bug in the code (based on preliminary results from one month of testing provided by the DTC) leading to a significant misrepresentation of surface (2m and 10m diagnostic) fields only, the upper air results remain unaffected. Because of the late date of this discovery, it was not feasible for the DTC to rerun after a fix had been checked into the WRF repository. All 2m and 10m diagnostic fields discussed in the next section contain this known bug. The exact impact of the bug fix is difficult to assess until a comprehensive rerun can be conducted.

4.2.1 Temperature BCRMSE and bias

The surface temperature BCRMSE for both configurations undergoes a slight increase with lead time for both 00 and 12 UTC initializations and for all temporal aggregations (Fig. 9). Diurnal variations are also evident for all temporal aggregations, with a weak signal noted for the winter season. The lowest error values occur around 06-09 UTC daily and the maximum BCRMSE values occur around 00 UTC. The SS pair-wise differences for all lead times and temporal aggregations generally indicate the AFWA configuration has lower BCRMSE then the QNSE configuration, except for one lead time for the fall aggregation (see Table 4). Generally, similar results are seen for the winter and spring seasonal aggregations, while for the summer aggregation, there is an overall decrease in the number of SS pair-wise differences.

Time series of surface temperature bias exhibit a diurnal cycle for both configurations. For both configurations, the cold surface temperature bias is strongest during daytime hours, with a weaker cold bias during the overnight hours. During the winter season, for the AFWA configuration, there is a weak warm bias during the overnight hours (Fig. 10), which indicates the configuration is under-predicting the amplitude of the diurnal temperature cycle. The magnitude and the sign of the bias are dependent on the phase of the diurnal cycle but the amplitude does not increase with increasing lead time. Conversely, the QNSE configuration produces a cold bias for all forecast lead times, during all temporal aggregations, where the magnitude of the bias is largest in the afternoon hours (~18 UTC) and undergoes a slight overall increasing trend with lead time. The SS pair-wise differences indicate the AFWA configuration produces the smallest surface temperature bias (see Table 4). For all temporal aggregations and for all lead times, the AFWA configuration has a SS smaller bias than the QNSE configuration.

4.2.2 Dew Point Temperature BCRMSE and bias

As was seen in the surface temperature BCRMSE results, the surface dew point temperature BCRMSE also exhibits a diurnal cycle, where the lowest BCRMSE values occur during the overnight hours and the highest values occur in the late afternoon into the evening hours (Fig. 11). In contrast, however, the surface dew point temperature BCRMSE for both configurations

undergoes a slight overall increase with lead time, rather than only the QNSE configuration exhibiting this behavior. The seasonal aggregations display similar trends as seen for the annual aggregation, with the exception of winter, where the QNSE configuration exhibits a SS higher bias and a larger increase with lead time. All SS pair-wise differences for the annual, winter, and spring aggregations correspond to the AFWA configuration having smaller errors (see Table 5). For the annual aggregation, SS pair-wise differences occur between 21 UTC and 03 UTC for both initialization times, regardless of the lead time. For winter, SS pair-wise differences occur for nearly all forecast lead times and more often during the daytime hours for the spring season. For the summer and fall aggregations the QNSE configuration has smaller errors for some of the overnight valid times, but the remaining SS pair-wise differences correspond to smaller errors for the AFWA configuration

More so than any other variable and metric evaluated, the bias for surface dew point temperature is sensitive to the forecast initialization, lead time, and seasonal aggregation, in particular, for the AFWA configuration. However, there are some general trends worth discussing. Both configurations exhibit a diurnal cycle that is consistent between the 00 UTC and 12 UTC initializations (Fig. 12). The overall amplitude of the diurnal cycle for the QNSE configuration is higher than the AFWA configuration leading to a larger high bias during the late afternoon into evening hours. The amplitude of these variations noted for the QNSE configuration do vary with season, where the summer season exhibits the largest amplitude, while the winter exhibits the smallest. Also for the winter season, the QNSE configuration transitions to a smaller high bias during the overnight into morning hours. The SS pair-wise differences for the annual and spring aggregations favor the AFWA configuration except for some valid times between 12-15 UTC where the QNSE configuration has a smaller bias (see Table 5). For the summer and fall seasons, the SS pair-wise differences favor the QNSE configuration for some valid times corresponding to 03-15 UTC, while other valid times favor the AFWA configuration. Finally, for the winter aggregation, the SS pair-wise differences all favor the QNSE configuration.

4.2.3 Wind BCRMSE and bias

The BCRMSE for surface wind speed shows a weak diurnal signal and an overall slight increase in error with longer lead times for both configurations (Fig. 13). The largest wind speed errors occur around 00 UTC and the smallest around 12 UTC for both configurations. The SS pair-wise differences for the annual aggregations favor the AFWA configuration, except between 00-12 UTC valid times regardless of season, especially for the 00 UTC initializations, which favor the QNSE configuration (see Table 6).

Both configurations produce a high wind speed bias at the surface for all forecast lead times and temporal aggregations (Fig. 14). The AFWA configuration produces a much more distinct diurnal cycle with biases that are larger than those for the QSNE configuration during the overnight hours and smaller during the daytime hours. This relationship leads to the general statement for the pair-wise differences that the QNSE configuration has a SS smaller bias during the overnight hours (between about 00 UTC and 12 UTC) and the AFWA configuration has a SS smaller bias during the daytime hours. This trend is seen for all season aggregations as well (see Table 6).

4.2.4 3-hourly QPF GSS and bias

When evaluating the GSS for precipitation it is important to know the number of observations that make up a particular distribution of values for each threshold. The base rate, indicating the ratio of observed grid box events to the total number of grid boxes in the domain, is shown on each precipitation plot by threshold. As the base rate decreases, the number of cases observed decreases and the event becomes infrequent. With this decreasing base rate is often an increase in the size of the CIs as well, indicating more spread and less confidence in the median value.

When examining the GSS values for the 3-hour QPF, it is seen that the highest GSS values occur at the lowest precipitation threshold of 0.01" and steadily decrease to near-zero for thresholds greater than 1.0" (Fig. 15). The number of observed events by threshold has a similar trend. The base rate for the 00 UTC 12-hour forecast is lower than the 12 UTC 12-hour forecast, likely due to the increased precipitation potential in the late afternoon with the heating cycle. For the seasonal aggregations, the base rate for the higher thresholds is the largest during the summer, as would be expected (Fig. 16), leading to smaller overall CIs. In the analysis presented here, configurations pair-wise differences favor the AFWA configuration with only three exceptions and these differences are generally seen for the lower thresholds (see Table 7). More SS pair-wise differences at higher thresholds are noted for the 00 UTC valid times, regardless of initialization time. The fewest number of SS pair-wise differences occur during the summer season.

With few exceptions, both configurations have a SS high bias for thresholds less than 0.25" regardless of initialization time or temporal aggregation (Fig. 17). For the annual and summer aggregations, above 0.25" the general trend is a decreasing bias where in many cases the CIs encompass one (perfect score for frequency bias) for the 0.35" threshold and then transition to a SS low bias for higher thresholds. In contrast, the CIs associated with the other seasonal aggregations generally encompass one for all thresholds above 0.25". Again, the largest CIs are noted during the winter season where the base rate is lowest. When applying the more conservative approach for assessing SS between the two configurations, the AFWA configuration is favored in all cases where no overlap of the CIs is observed. SS differences are generally noted for the spring and summer seasons, along with the overall annual average, for the lowest thresholds from forecasts valid at 00 UTC, regardless of the initialization time (see Table 7).

4.2.5 Daily Precipitation GSS and bias

The base rate for the 24-hour QPF is over 30% for the lowest threshold but the decrease in GSS values as the threshold increases is similar to that shown for the 3-hour QPF (Fig. 18). Without exception, the SS pair-wise differences favor the AFWA configuration for the 24-QPF field (see Table 8). These differences are generally noted for thresholds less than one inch; however, some higher threshold SS pair-wise differences should be acknowledged, especially for the 12 UTC initializations.

The overall magnitude of the 24-hour accumulation biases for the 00 and 12 UTC initializations are similar up to the 1" threshold, and reveal a general SS high bias for both configurations for all seasons, with a few exceptions where the AFWA configuration encompasses one during the fall and winter seasons (Fig. 19). For the largest accumulation thresholds (greater than 1.5" or 2") the CIs are very large (especially for the fall and spring seasons), encompassing one for all except the summer aggregation, and are, therefore, classified as nonbiased due to low confidence in the actual magnitude or sign of the bias. Once again, when using the more conservative method for assessing SS between the two configurations, all favor the AFWA configuration, occur at the lowest thresholds, and are seen only for the summer and annual aggregations (see Table 8).

4.3 GO Index

The computation of the GO Index, as defined in section 3.3, indicates that the AFWA configuration has higher skill (i.e., GO Index skill score less than one) as compared to the QNSE configuration, regardless of the temporal aggregation and initialization time (Fig. 20). In general, the overall variability is largest for the annual aggregation, while it is smallest for the summer aggregation. In general, skill scores are not sensitive to initialization time.

5. Summary

Two WRF-ARW configurations were comprehensively tested and evaluated to assess the impact of the new QNSE PBL and surface layer schemes available in WRF, using AFWA's Operational Configuration as a baseline. Because both configurations were run for the same cases, pair-wise differences were computed for standard verification metrics between the two configurations, and an assessment of the statistical significance (SS) was included. In general, the AFWA configuration was favored more often than the QNSE configuration. However, for some metrics and certain levels, lead times, thresholds, or temporal aggregations, QNSE was favored. It may be noted, though, that the relative magnitudes of the SS differences favoring the AFWA configuration are generally larger than those favoring the QNSE configuration.

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.

Sukoriansky, S., B. Galperin, and V. Perov, 2005: Application of a new spectral theory of stably stratified turbulence to the atmospheric boundary layer over sea ice. *Boundary-Layer Meteorol.,* **117,** 231-257.

Table 1. Statistically significant (SS) pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted 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 full integration domain.

		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	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	700	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	500	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA
	400	AFWA	--	AFWA	AFWA	--	QNSE	--	--	--	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA	AFWA
	300	--	QNSE	--	--	--	QNSE	QNSE	QNSE	QNSE	--	--	--	--	AFWA	AFWA	--	--	--	--	--
	200	QNSE	QNSE	QNSE	--	QNSE	QNSE	QNSE	--	QNSE	QNSE	--	--	QNSE	--	--	--	QNSE	QNSE	QNSE	QNSE
	150	AFWA	--	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	--	--	AFWA	AFWA	AFWA	--	--	AFWA	--
	100	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--	--	--	--	--	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA
Bias	850	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	--	QNSE	--	--	--	AFWA	QNSE	QNSE	--	--
	700	QNSE	QNSE	QNSE	QNSE	QNSE	QSNE	QNSE	QNSE	QNSE	QNSE	--	QNSE	--	--	--	--	QNSE	QNSE	QNSE	--
	500	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	--	AFWA	AFWA	--	AFWA	--	AFWA	AFWA	AFWA	AFWA
	400	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	--	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA
	300	AFWA	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA	--	QNSE	--	--	--	AFWA	--	--	--
	200	QNSE	QNSE	QNSE	QNSE	--	--	--	AFWA	QNSE	QNSE	QNSE	QNSE	--	--	QNSE	--	QNSE	QNSE	QNSE	QNSE
	150	--	--	--	--	QNSE	QNSE	QNSE	QNSE	QNSE	--	QNSE	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--
	100	QNSE	QNSE	--	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	--	--	--	--	AFWA	AFWA	--	AFWA	--	AFWA

Table 2. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted 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 full integration domain.

		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	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	700	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	500	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA
Bias	850	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	QNSE	QNSE	QNSE	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE
	700	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	AFWA	QNSE	--	QNSE	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	--
	500	QNSE	QNSE	QNSE	QNSE	QNSE	--	--	QNSE	QNSE	QNSE	--	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	--	QNSE

Table 3. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted 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 full integration domain.

		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	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	700	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	500	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA
	400	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA	--	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	300	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--	--	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA
	200	--	--	AFWA	AFWA	--	--	--	--	--	--	--	--	--	--	AFWA	AFWA	--	--	--	--
	150	AFWA	AFWA	AFWA	AFWA	--	--	--	--	AFWA	--	AFWA	--	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA
	100	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--	AFWA	AFWA	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--
Bias	850	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	700	QNSE	QNSE	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	QNSE	QNSE	AFWA	QNSE	QNSE	AFWA	AFWA	AFWA
	500	QNSE	QNSE	QNSE	QNSE	QNSE	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE
	400	QNSE	QNSE	QNSE	QNSE	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	QNSE	QNSE	QNSE	AFWA
	300	QNSE	QNSE	QNSE	QNSE	--	QNSE	--	QNSE	QNSE	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE
	200	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE
	150	QNSE	QNSE	QNSE	QNSE	QNSE	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	QNSE	QNSE	--	--
	100	QNSE	--	--	--	QNSE	--	QNSE	QNSE	--	AFWA	--	--	--	--		--	--	--	--	--

Table 4. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted 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 full integration domain.

			f03	f06	f09	f12	f15	f18	f21	f24	F27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA	AFWA	AFWA	AFWA
		Summer	AFWA	--	--	--	AFWA	AFWA	AFWA	--	--	--	--	--	AFWA	AFWA	AFWA	--
		Fall	AFWA	--	--	QNSE	--	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA
		Winter	AFWA	--	AFWA	--	--	AFWA	AFWA	AFWA	--	AFWA	AFWA	--	--	AFWA	AFWA	AFWA
		Spring	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA
	12 UTC Initializations	Annual	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Summer	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--	AFWA
		Fall	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--
		Winter	--	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
Bias	00 UTC Initializations	Annual	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Summer	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Fall	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Winter	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
	12 UTC Initializations	Annual	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Summer	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Fall	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Winter	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA

Table 5. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted 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 full integration domain.

			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	--	--	QNSE	--	--	--	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	AFWA
		Summer	--	QNSE	QNSE	QNSE	QNSE	--	AFWA	AFWA	QNSE	QNSE	QNSE	--	--	--	AFWA	AFWA
		Fall	--	QNSE	QNSE	--	--	--	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	AFWA
		Winter	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA
	12 UTC Initializations	Annual	AFWA	--	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--
		Summer	--	--	AFWA	AFWA	QNSE	QNSE	QNSE	--	--	AFWA	AFWA	AFWA	QNSE	QNSE	--	--
		Fall	--	AFWA	AFWA	AFWA	--	--	--	--	--	AFWA	AFWA	AFWA	--	--	--	AFWA
		Winter	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA
Bias	00 UTC Initializations	Annual	--	--	--	--	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	--	--	AFWA	AFWA
		Summer	QNSE	QNSE	QNSE	--	AFWA	--	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA
		Fall	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--	--	AFWA	AFWA
		Winter	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	--	--	QNSE	QNSE	QNSE	QNSE	QNSE	QNSE	--	--
		Spring	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	--	AFWA	AFWA	AFWA
	12 UTC Initializations	Annual	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	--	--	AFWA	AFWA	AFWA	AFWA	--	--
		Summer	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE
		Fall	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	--	--	AFWA	AFWA	QNSE	QNSE	QNSE	--
		Winter	QNSE	--	--	--	--	--	--	QNSE	QNSE	QNSE	--	--	QNSE	QNSE	QNSE	QNSE
		Spring	QNSE	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	--

Table 6. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted 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 full integration domain.

			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Summer	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	AFWA	QNSE	--	--	--	AFWA	AFWA	AFWA	AFWA	--
		Fall	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Winter	--	QNSE	QNSE	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	--	--	--	--	AFWA	AFWA	AFWA	--	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	--
	12 UTC Initializations	Annual	AFWA	AFWA	AFWA	QNSE	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Summer	AFWA	AFWA	AFWA	QNSE	--	--	--	AFWA	AFWA	AFWA	AFWA	QNSE	--	AFWA	AFWA	AFWA
		Fall	AFWA	AFWA	AFWA	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Winter	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
		Spring	AFWA	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA
Bias	00 UTC Initializations	Annual	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE
		Summer	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	QNSE
		Fall	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE
		Winter	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	QNSE
		Spring	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	QNSE
	12 UTC Initializations	Annual	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--
		Summer	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--
		Fall	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--
		Winter	--	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	QNSE	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	--	--
		Spring	AFWA	AFWA	AFWA	QNSE	QNSE	QNSE	QNSE	--	AFWA	AFWA	AFWA	--	QNSE	QNSE	QNSE	--

Table 7. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted is favored) for 3-hour QPF GSS and frequency bias by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the full integration domain.

			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
GSS	Annual	f12	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--
		f24	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	--	--	--	--	--	--
		f36	AFWA	AFWA	--	--	--	--	--	AFWA	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--
		f48	AFWA	AFWA	AFWA	--	AFWA	AFWA	--	--	--	AFWA	AFWA	AFWA	--	--	--	--	--	--
	Summer	f12	AFWA	AFWA	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--
		f24	--	--	--	--	--	AFWA	--	AFWA	--	AFWA	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	AFWA	--	--	AFWA	AFWA	AFWA	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Fall	f12	AFWA	AFWA	AFWA	--	--	--	--	--	QNSE	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--
		f24	AFWA	AFWA	--	--	--	--	--	--	--	AFWA	--	--	--	--	--	--	--	--
		f36	AFWA	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--
		f48	AFWA	AFWA	AFWA	--	--	--	--	--	--	AFWA	--	--	--	--	--	--	--	--
	Winter	f12	AFWA	--	--	--	--	--	--	--	--	AFWA	--	--	--	--	--	--	AFWA	--
		f24	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	--	AFWA	AFWA	AFWA	--	--	--	--	--	QNSE
		f36	AFWA	--	AFWA	--	AFWA	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	QNSE
		f48	AFWA	AFWA	--	--	--	--	--	--	--	AFWA	AFWA	--	--	--	--	--	--	--
	Spring	f12	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	--
		f24	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--	AFWA	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--
		f48	AFWA	AFWA	AFWA	--	AFWA	--	AFWA	--	--	--	--	--	--	--	AFWA	--	--	--

Frequency Bias																				
	Annual	f12	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--
		f24	AFWA	AFWA	AFWA	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	--	--	--	--	--	--
		f48	AFWA	AFWA	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Summer	f12	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--
		f24	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	AFWA	AFWA	AFWA	--	--	--	--	--	--
		f48	AFWA	AFWA	AFWA	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Fall	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Spring	f12	--	--	--	--	--	--	--	--	--	AFWA	--	--	--	--	--	--	--	--
		f24	AFWA	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	AFWA	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table 8. SS pair-wise differences between the AFWA and QNSE configuration run with WRF v3.1.1+ (where the version highlighted is favored) for 24-hour QPF GSS and frequency bias by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the full integration domain.

				>0.01	>0.25	>0.5	>0.75	>1	>1.25	>1.5	>2	>3
GSS	00 UTC Initializations	Annual	f36	AFWA	AFWA	AFWA	AFWA	--	--	--	--	AFWA
		Summer	f36	--	AFWA	--	--	--	--	--	--	--
		Fall	f36	AFWA	AFWA	--	--	--	--	--	AFWA	AFWA
		Winter	f36	AFWA	--	--	AFWA	--	--	--	--	--
		Spring	f36	AFWA	AFWA	AFWA	AFWA	--	--	--	--	--
	12 UTC Initializations	Annual	f24	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--
			f48	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--
		Summer	f24	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--
			f48	AFWA	AFWA	AFWA	AFWA					
		Fall	f24	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	--
			f48	AFWA	AFWA	--	--	--	--	--	--	--
		Winter	f24	AFWA	AFWA						AFWA	AFWA
			f48	AFWA	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--
		Spring	f24	AFWA	AFWA	AFWA	AFWA	AFWA	--	--	--	--
			f48	AFWA	AFWA	--	AFWA	--	--	AFWA	--	--
Frequency Bias	00 UTC Initializations	Annual	f36	AFWA	--	--	--	--	--	--	--	--
		Summer	f36	AFWA	AFWA							
		Fall	f36	--	--	--	--	--	--	--	--	--
		Winter	f36	--	--	--	--	--	--	--	--	--
		Spring	f36	--	--	--	--	--	--	--	--	--
	12 UTC Initializations	Annual	f24	AFWA	AFWA	--	--	--	--	--	--	--
			f48	AFWA	--	--	--	--	--	--	--	--
		Summer	f24	AFWA	AFWA	--	--	--	--	--	--	--
			f48	AFWA	--	--	--	--	--	--	--	--
		Fall	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Spring	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--

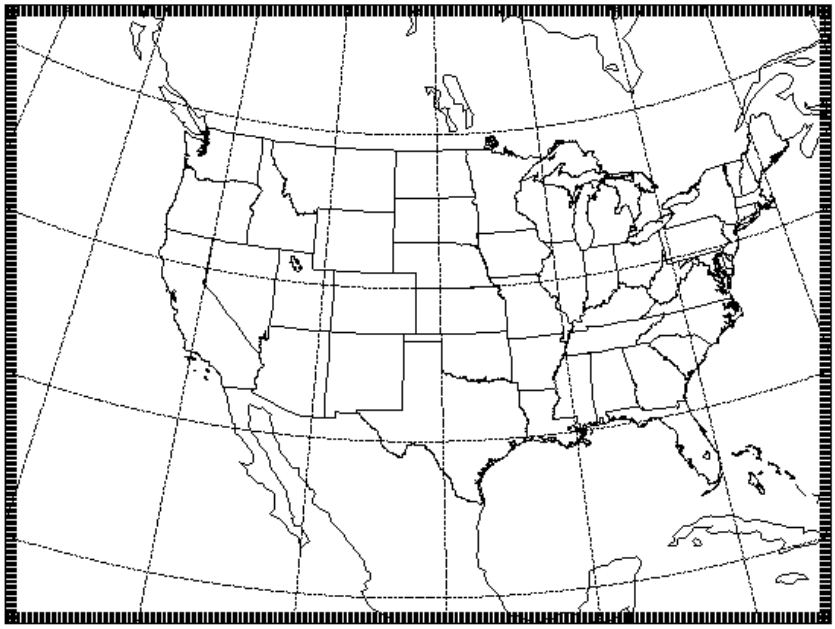

Figure 1. Map showing the boundary of the WRF-ARW computational domain.

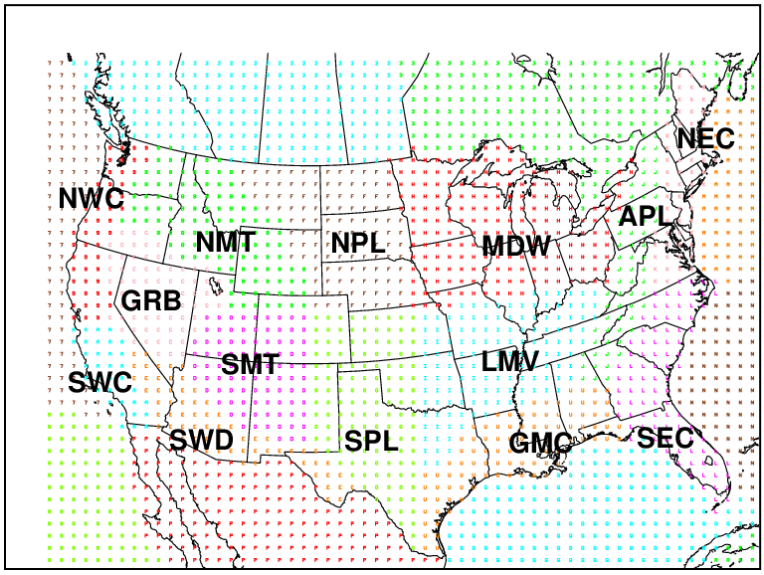

Figure 2. Map showing the locations of the 14 regional verification domains.

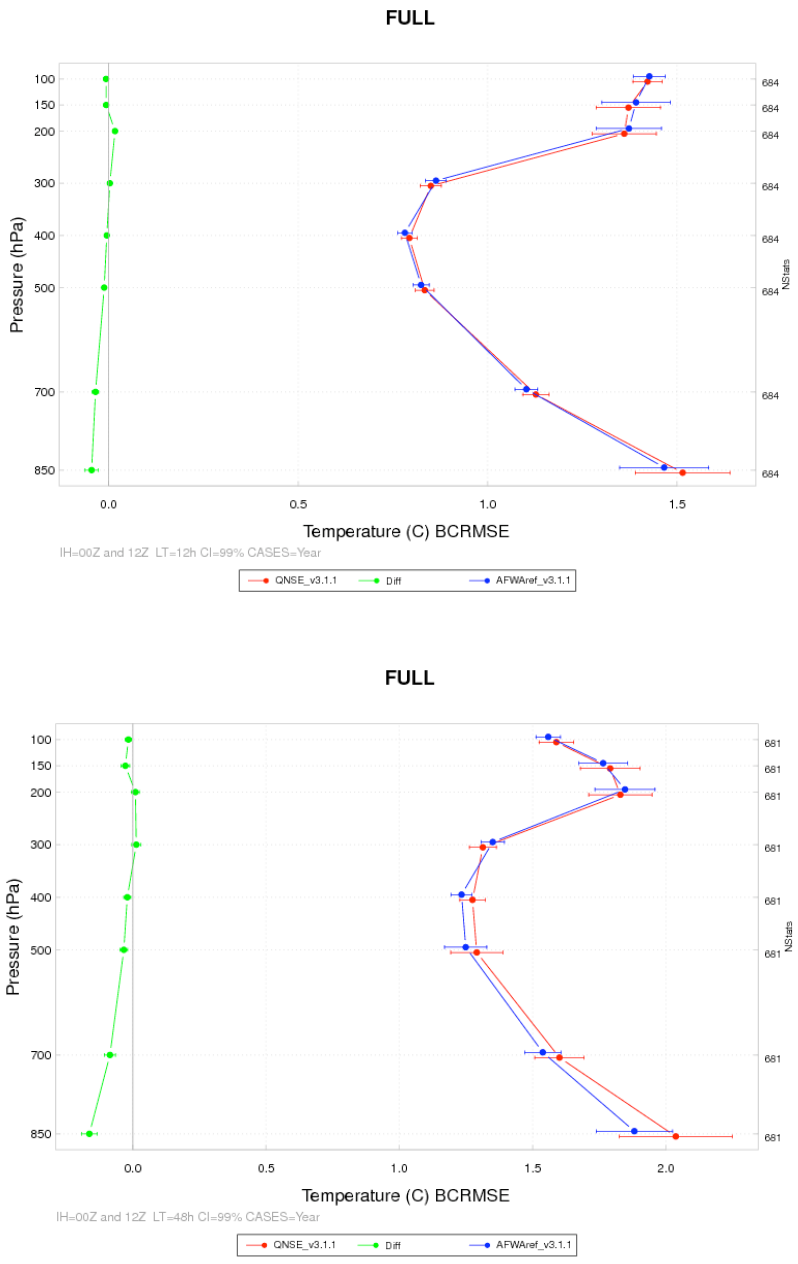

Figure 3. Vertical profile of the median BCRMSE for temperature (C) for the full integration domain aggregated across the entire year of cases (annual) for the 12-hour (top) and 48-hour (bottom) lead times. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The horizontal bars represent the 99% CIs.

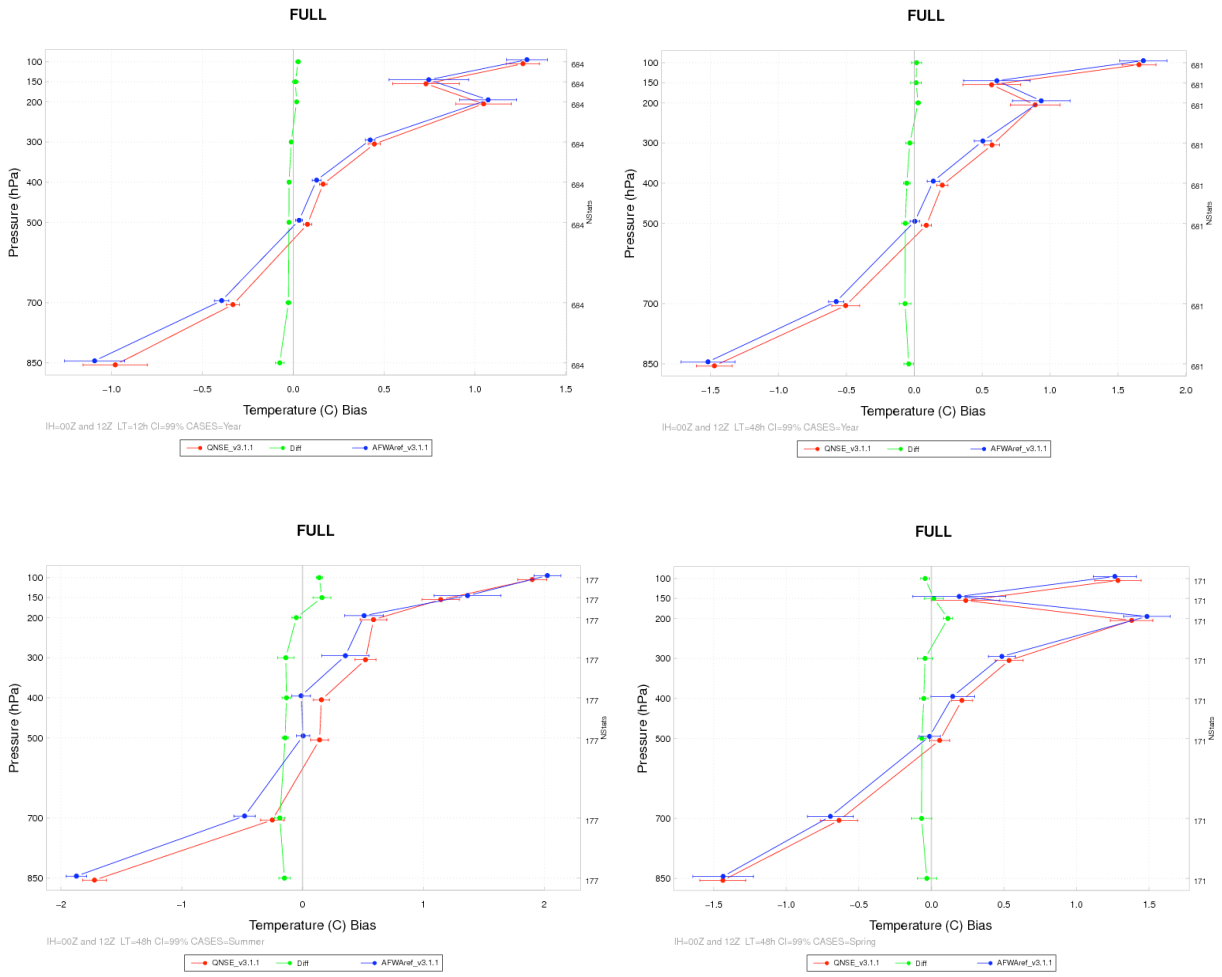

Figure 4. Vertical profile of the median bias for temperature (C) for the full integration domain aggregated across the entire year of cases for the 12-hour (top left) and 48-hour (top right) lead times, and for the 48-hour lead time for the summer season (bottom left) and spring season (bottom right). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The horizontal bars represent the 99% CIs.

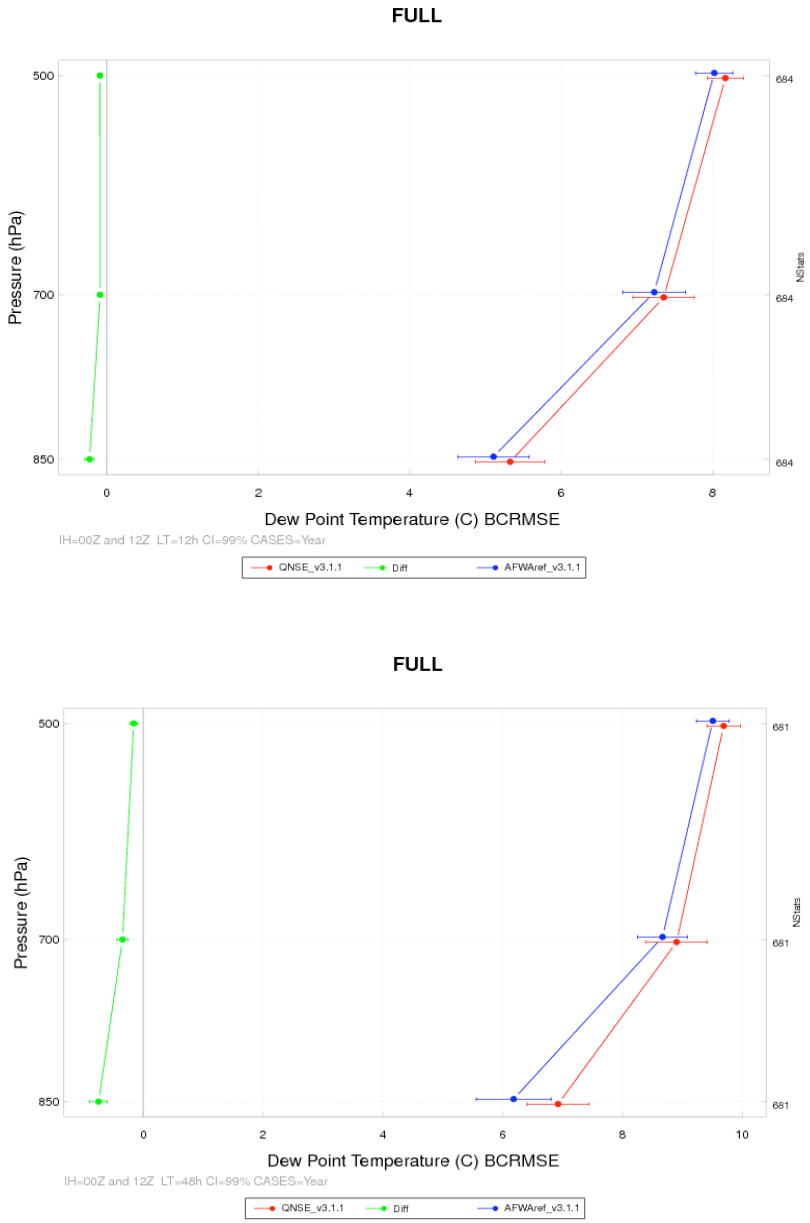

Figure 5. Vertical profile of the median BCRMSE for dew point temperature (C) for the full integration domain aggregated across the entire year of cases (annual) for the 12-hour (top), 48-hour (bottom) lead times. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The horizontal bars represent the 99% CIs.

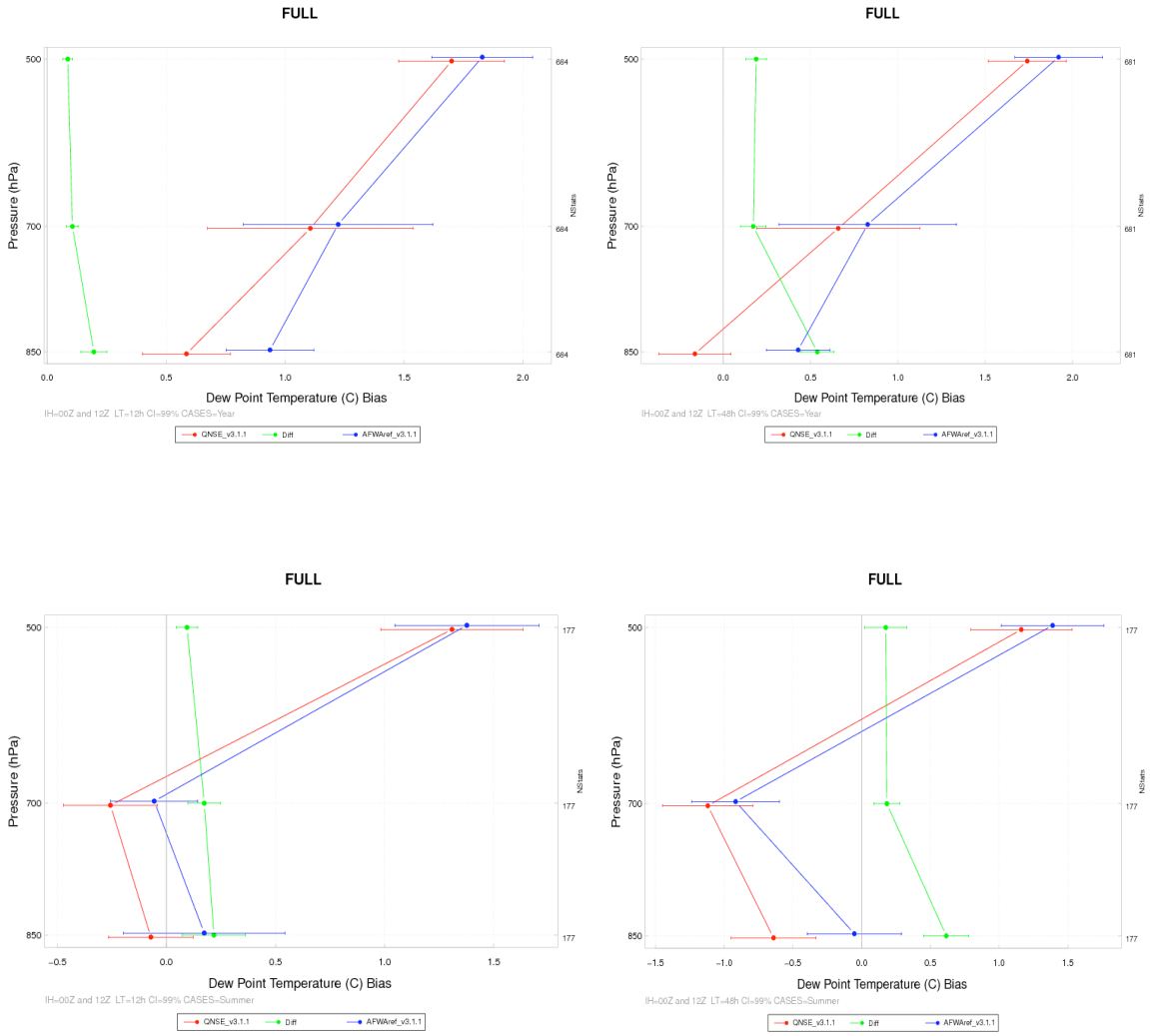

Figure 6. Vertical profile of the median bias for dew point temperature (C) for the full integration domain at the 12-hour (left) and 48-hour (right) lead times aggregated across the entire year of cases (top), and for the summer season (bottom). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The horizontal bars represent the 99% CIs.

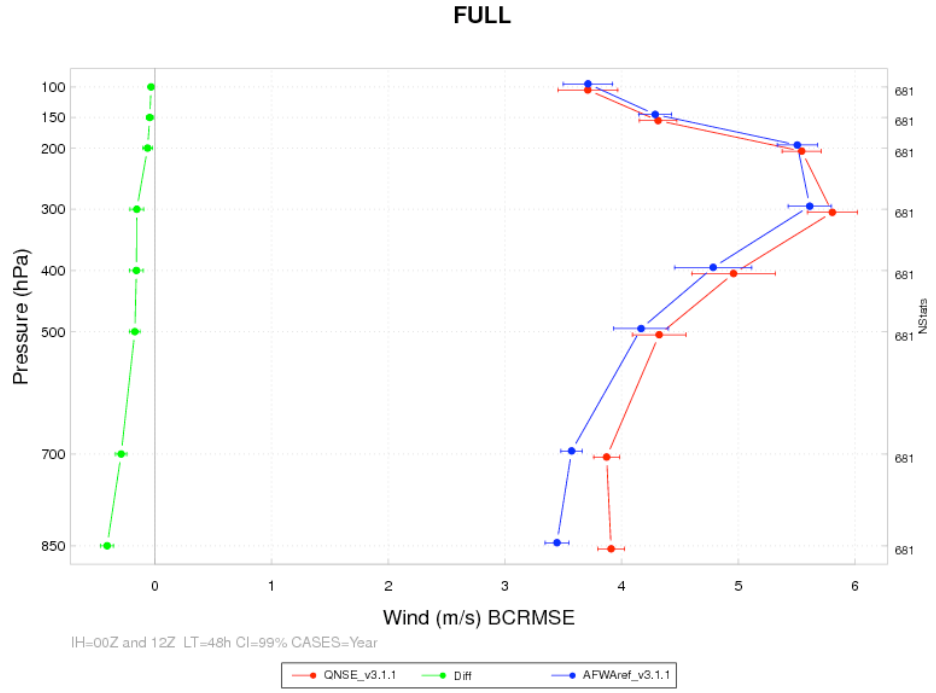

Figure 7. Vertical profile of the median BCRMSE of wind speed (m/s) for the full integration domain at the 48-hour lead time aggregated across the entire year of cases. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The horizontal bars represent the 99% CIs.

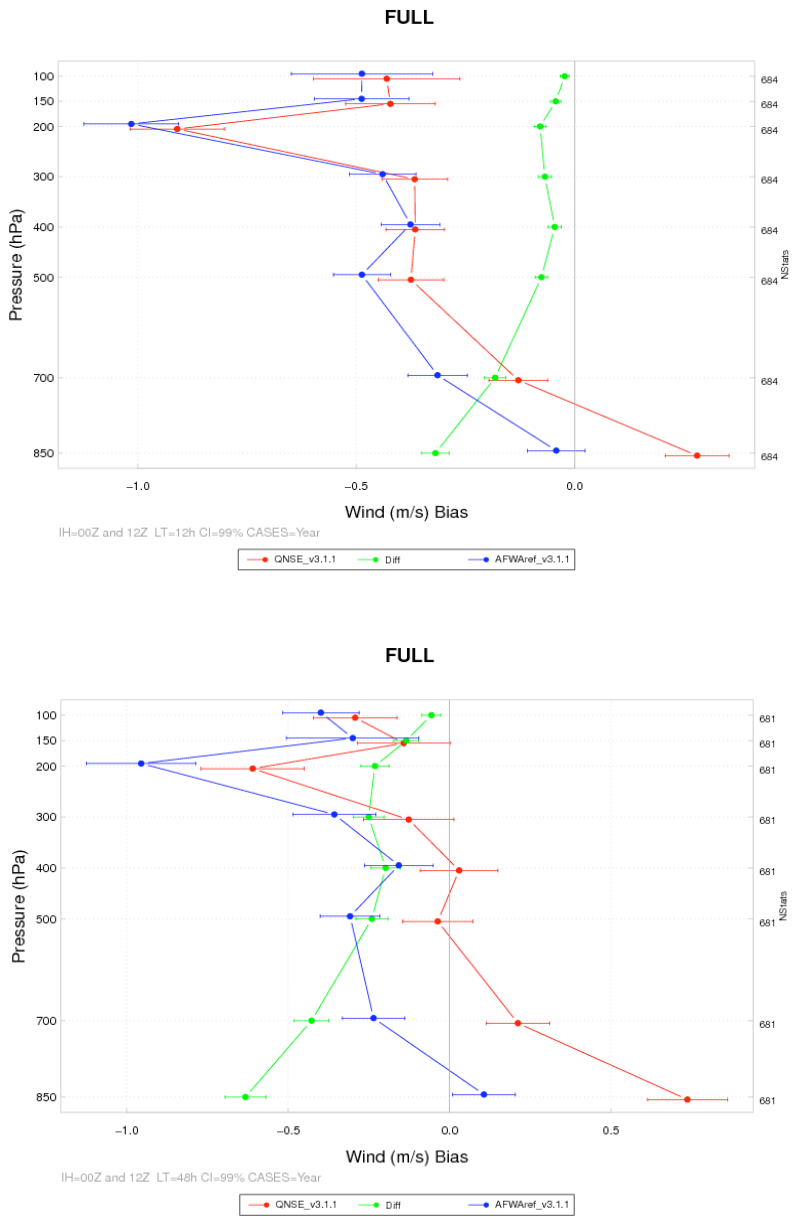

Figure 8. Vertical profile of the median bias of wind speed (m/s) for the full integration domain aggregated across the entire year of cases (annual) for the 12-hour (top) and 48-hour (bottom) lead times. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The horizontal bars represent the 99% CIs.

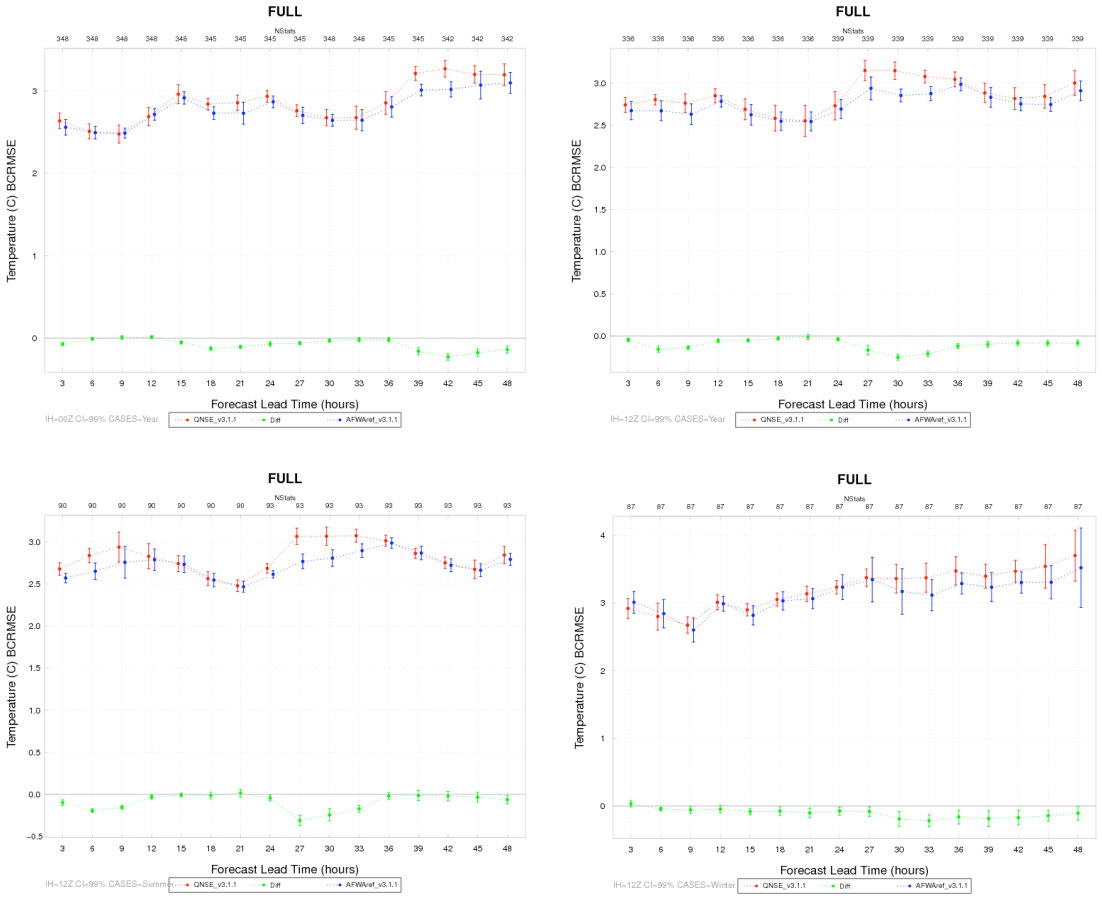

Figure 9. Time series plot of 2m AGL temperature (C) for median BCRMSE for the 00 UTC initializations only (top left) and 12 UTC initializations only (top right) aggregated across the entire year of cases, for the summer season 12 UTC initializations only (bottom left), and for the winter season 12 UTC initializations only (bottom right). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs.

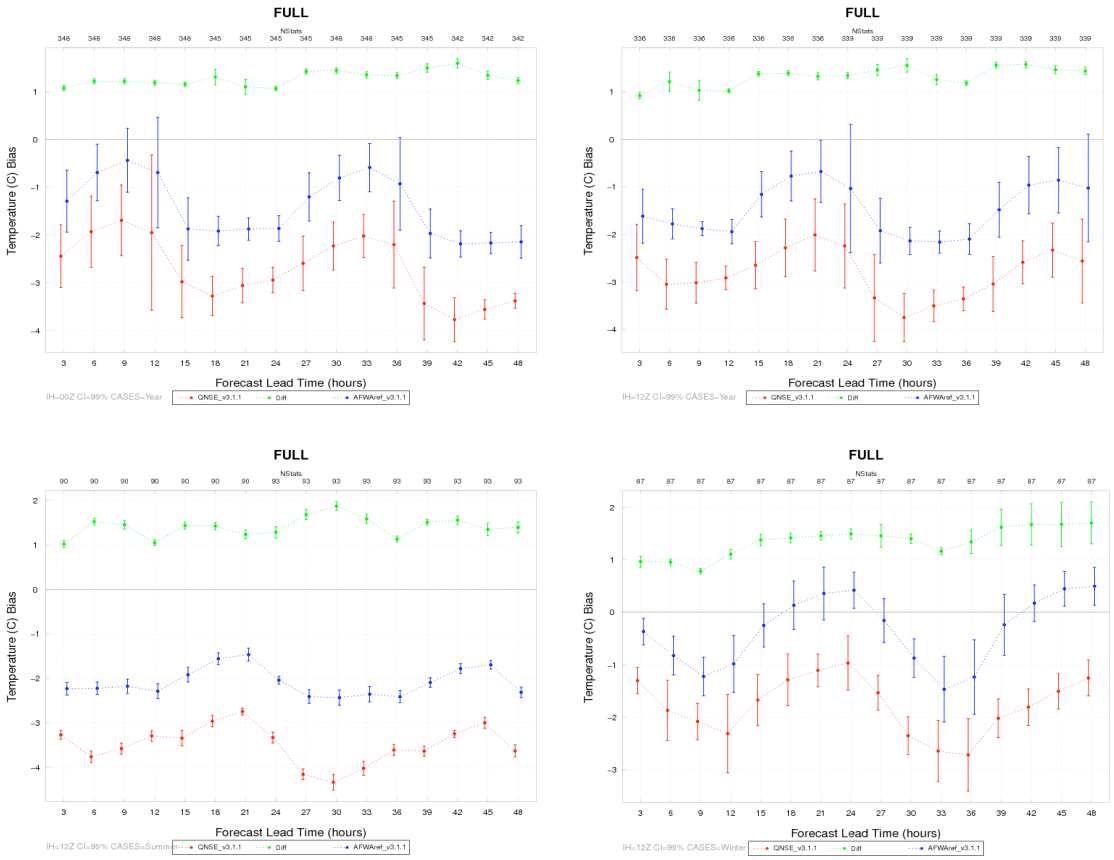

Figure 10. Time series plot of 2m AGL temperature (C) for median bias for the 00 UTC initializations only (top left) and 12 UTC initializations only (top right) aggregated across the entire year of cases, for the summer season 12 UTC initializations only (bottom left), and for the winter season for the 12 UTC initializations only (bottom right). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs.

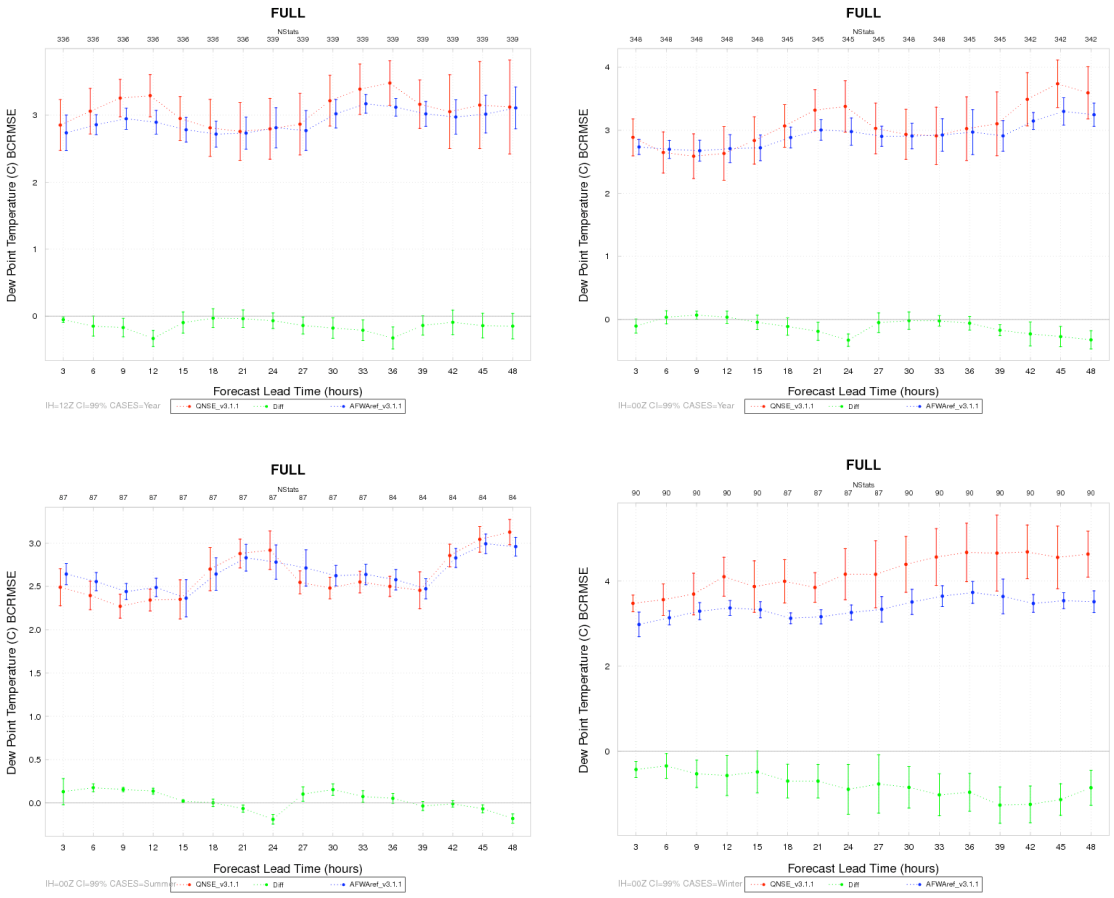

Figure 11. Time series plot of 2m AGL dew point temperature (C) for median BCRMSE for the 12 UTC initializations only (top left) and 00 UTC initializations only (top right) aggregated across the entire year of cases, for the summer season 00 UTC initializations only (bottom left), and for the winter season 00 UTC initializations only (bottom right). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs.

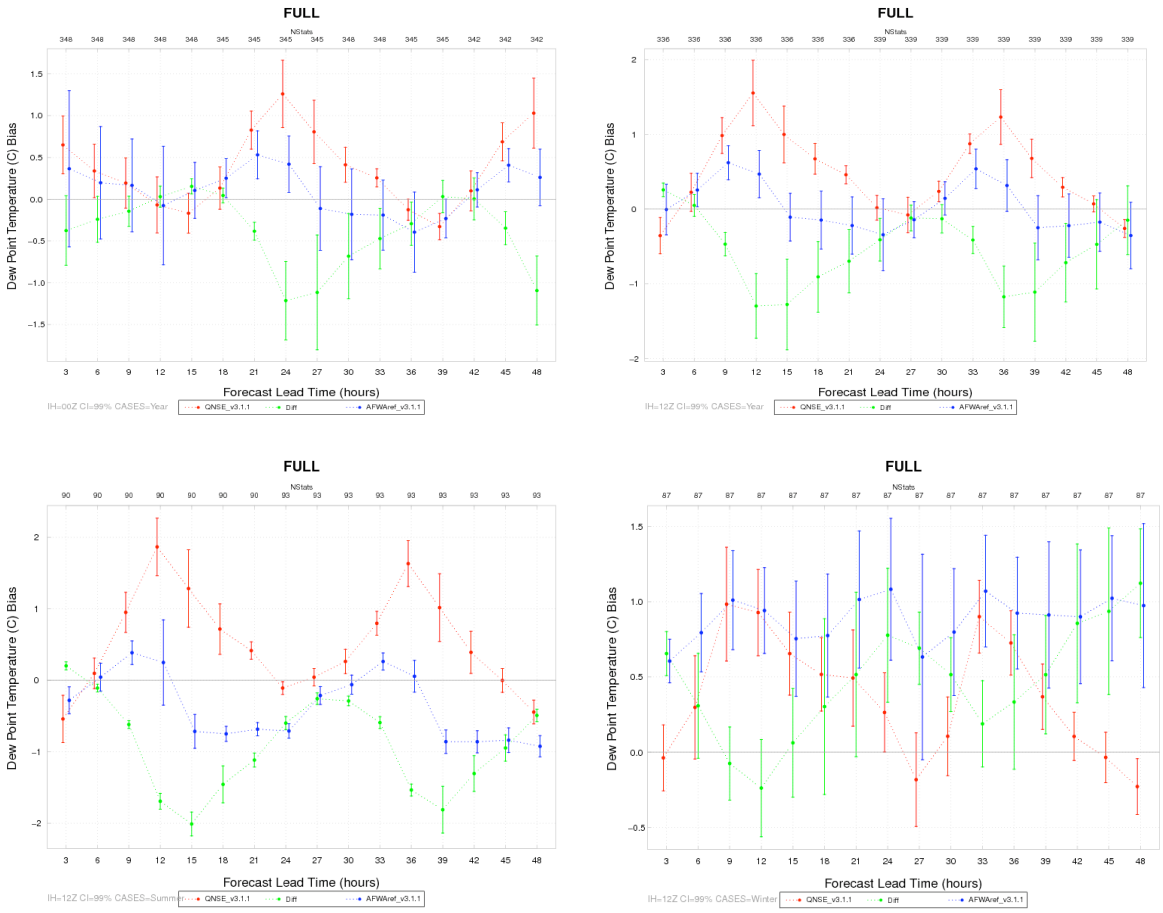

Figure 12. Time series plot of 2m AGL dew point temperature (C) for median bias for the 00 UTC initializations only (top left) and 12 UTC initializations only (top right) aggregated across the entire year of cases, for the summer season 12 UTC initializations only (bottom left), and for the winter season 12 UTC initializations only (bottom right). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs.

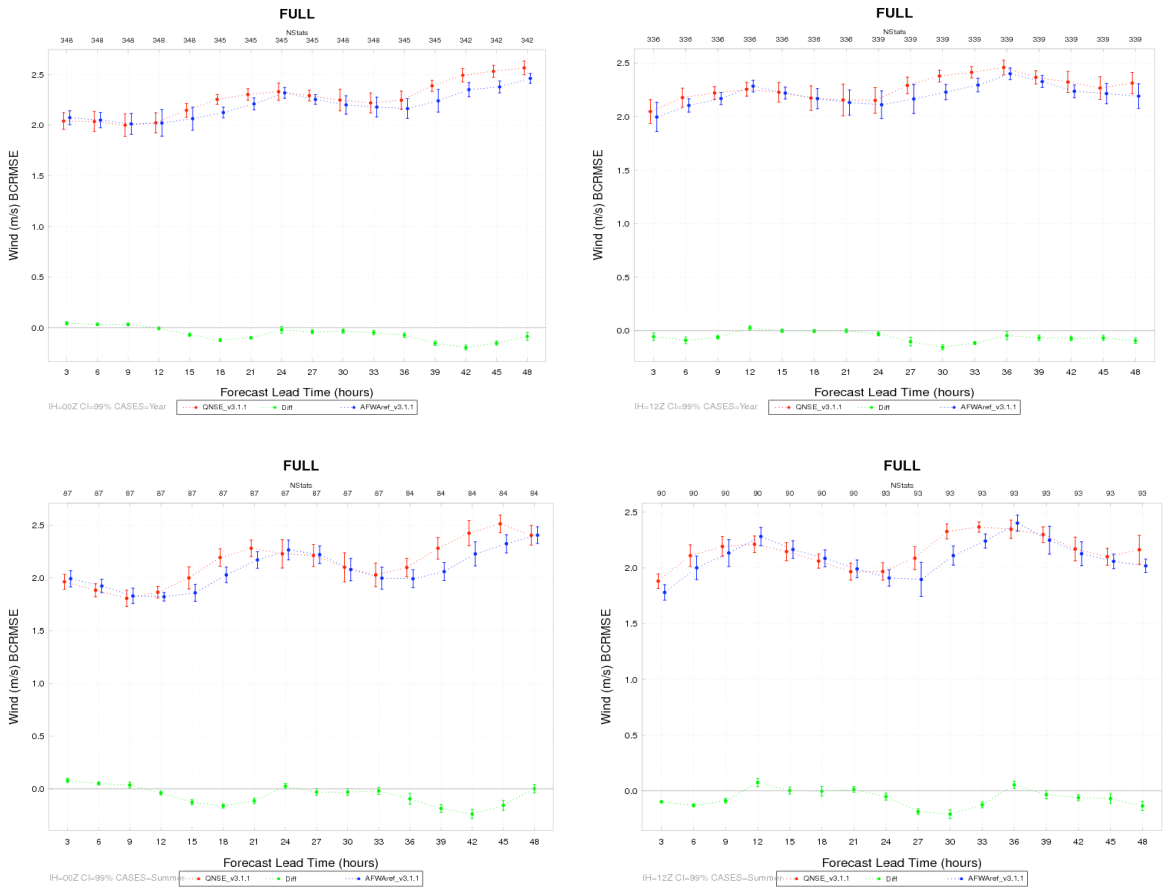

Figure 13. Time series plot of 10m AGL wind speed (m/s) for median BCRMSE for the 00 UTC (left column) and 12 UTC (right column) initializations aggregated across the entire year of cases (top row) and for the summer season (bottom row). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs.

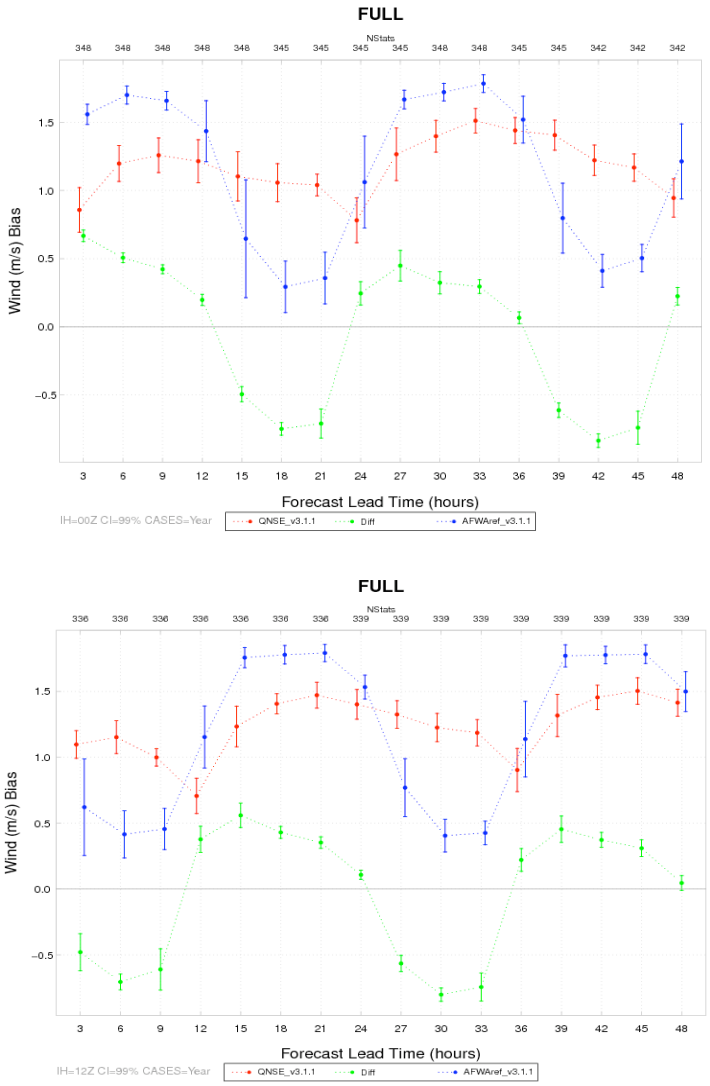

Figure 14. Time series plot of 10m AGL wind speed (m/s) for median bias for the 00 UTC (top) and 12 UTC (bottom) initializations aggregated across the entire year of cases. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs.

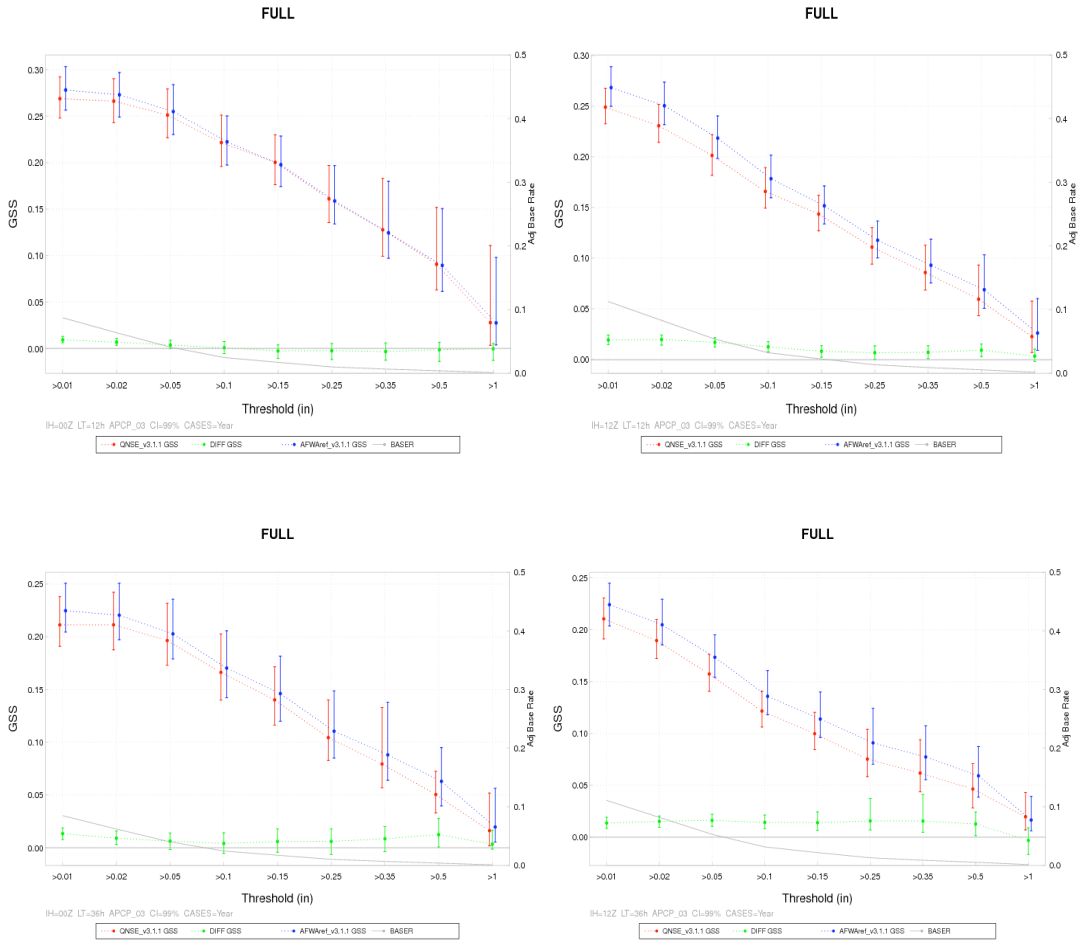

Figure 15. Threshold series plot of 3-hour accumulated precipitation (in) for median GSS for the 00 UTC (left column) and 12 UTC (right column) initializations aggregated across the entire year of cases for the 12-hour (top row) and 36-hour (bottom row) lead times. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs. Associated with the second y-axis, the light grey line is the adjusted base rate, or the ratio of observed grid box events to the total number of grid boxes in the domain, by threshold.

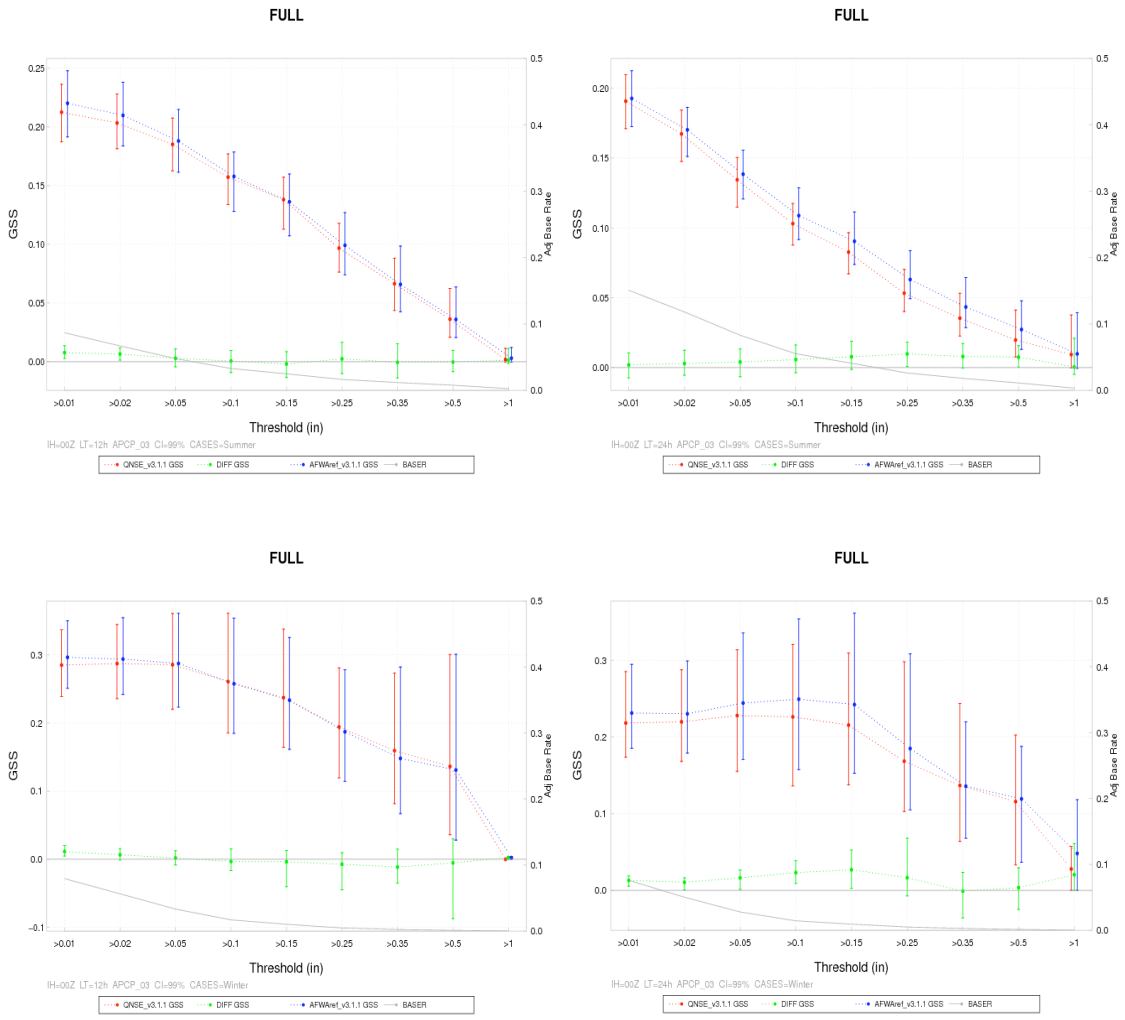

Figure 16. Threshold series plot of 3-hour accumulated precipitation (in) for median GSS for the 00 UTC initializations only across the summer season (top row) and the winter season (bottom row) for the 12-hour (left column) and 24-hour (right column) lead times. The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs. Associated with the second y-axis, the light grey line is the adjusted base rate, or the ratio of observed grid box events to the total number of grid boxes in the domain, by threshold.

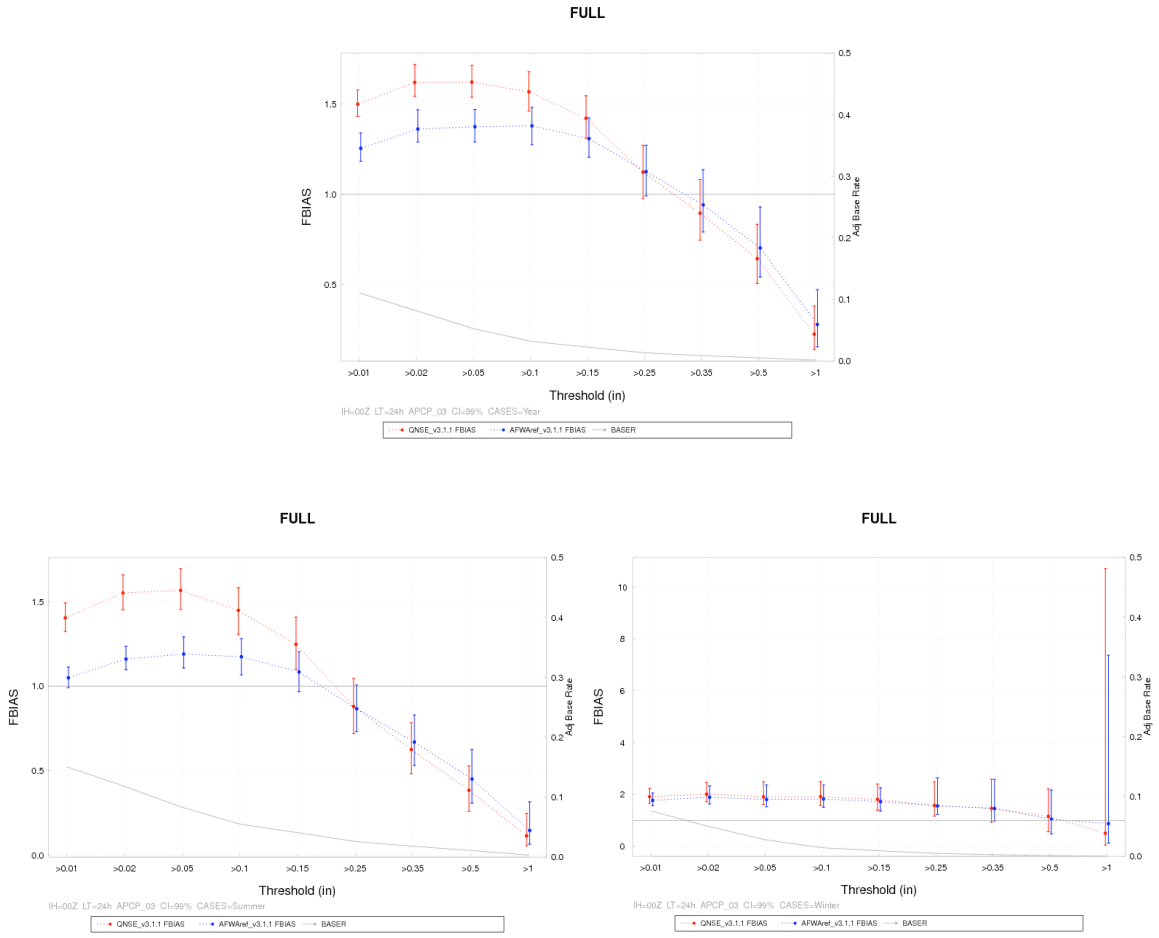

Figure 17. Threshold series plot of 3-hour precipitation accumulation (in) for median frequency bias for the 00 UTC initializations 24-hour lead time only aggregated across the entire year of cases (top), for the summer season (bottom left), and for the winter season (bottom right). The AFWA configuration is shown in blue and the QNSE configuration in red. The vertical bars represent the 99% CIs. Associated with the second y-axis, the light grey line is the adjusted base rate, or the ratio of observed grid box events to the total number of grid boxes in the domain, by threshold.

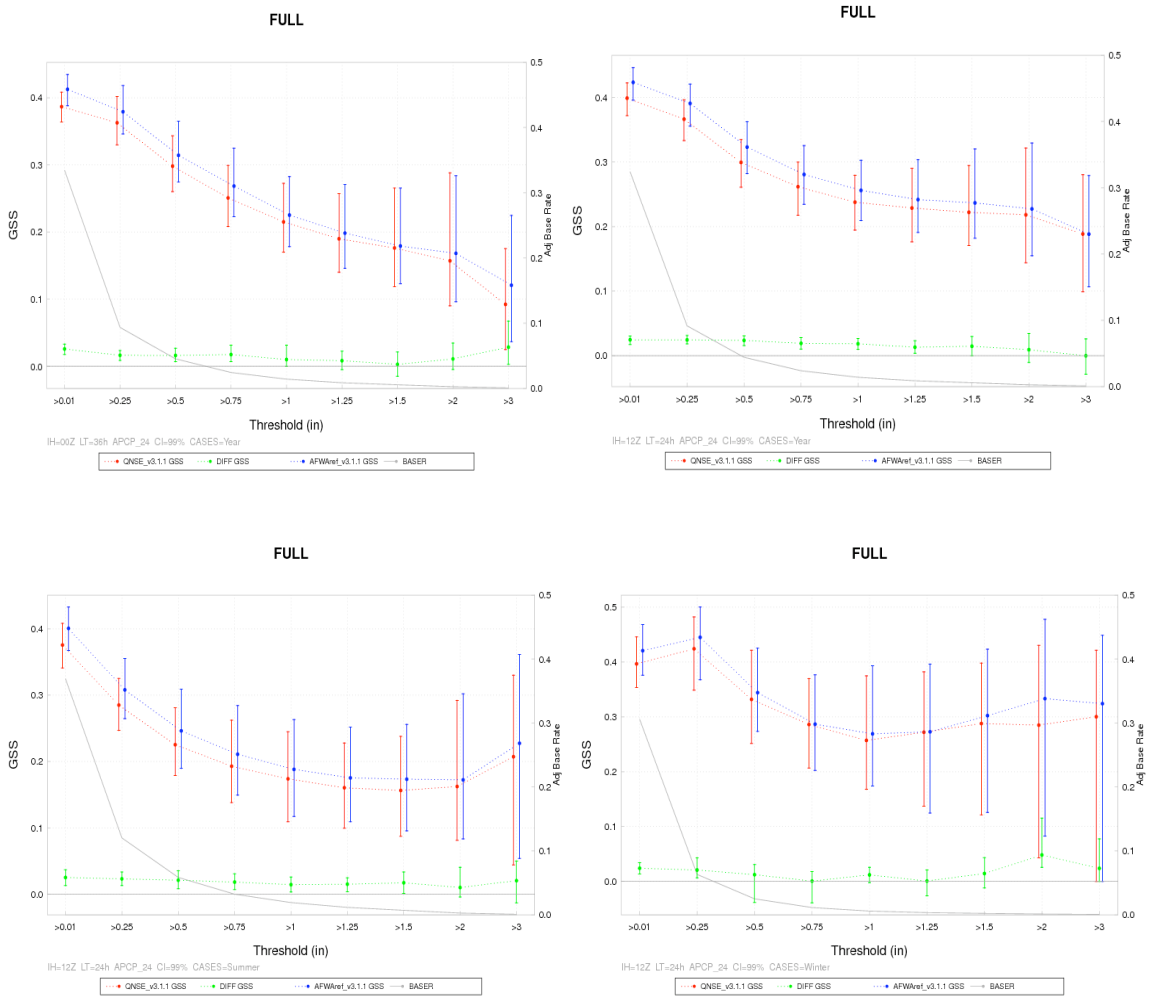

Figure 18. Threshold series plot of 24-hour precipitation accumulation (in) for median GSS for the 00 UTC 36-hour lead time (top left) and the 12 UTC 24-hour lead time (top right) aggregated across the entire year of cases, the summer season 12 UTC 24-hour lead time (bottom left), and the winter season 12 UTC 24-hour lead time (bottom right). The AFWA configuration is shown in blue, the QNSE configuration in red, and the differences (AFWA-QNSE) in green. The vertical bars represent the 99% CIs. Associated with the second y-axis, the light grey line is the adjusted base rate, or the ratio of observed grid box events to the total number of grid boxes in the domain, by threshold.

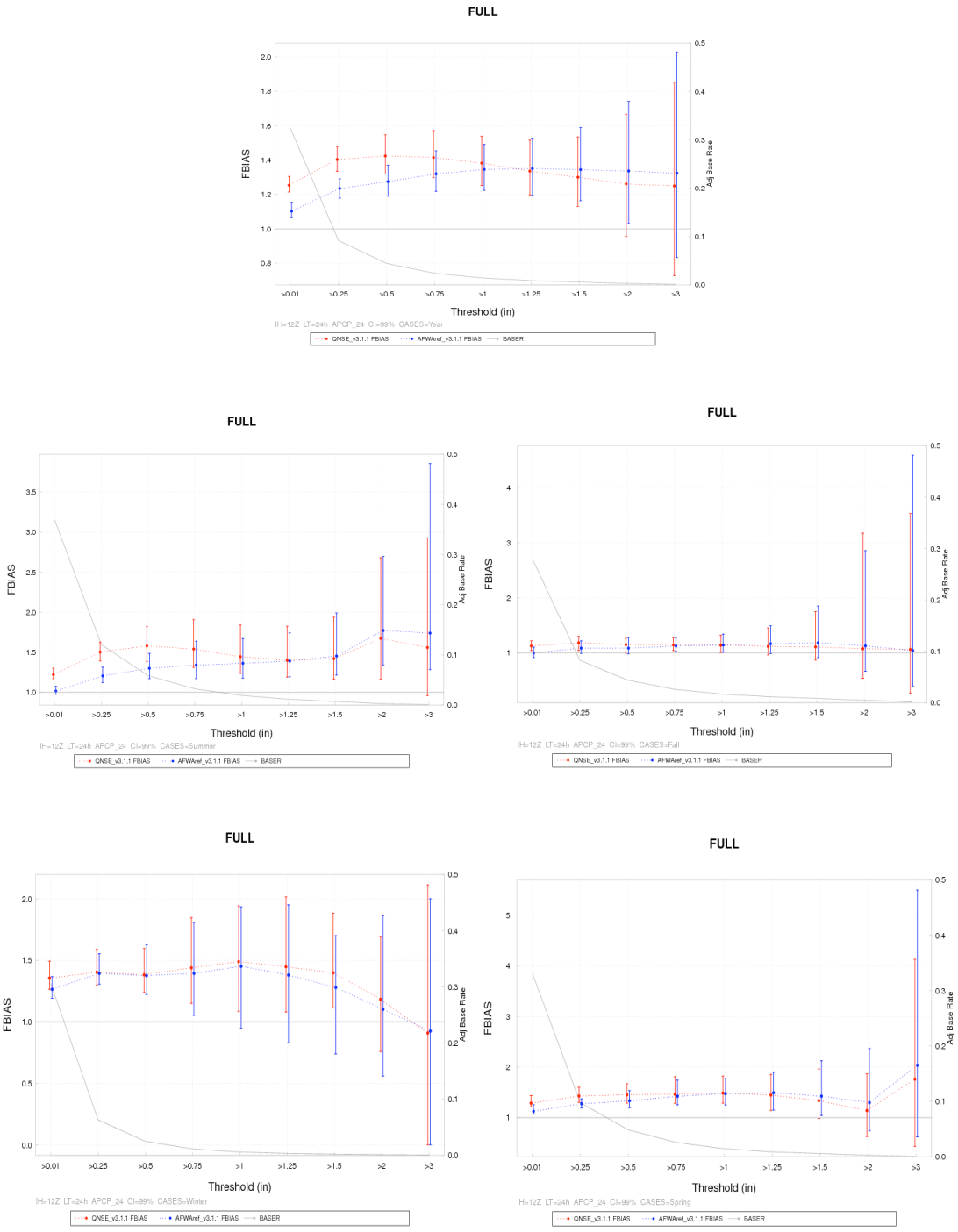

Figure 19. Threshold series plot of 24-hour precipitation accumulation (in) for median frequency bias for the 12 UTC initializations 24-hour lead time only aggregated across the entire year of cases (top), for the summer season (middle left), for the fall season (middle right), for the winter season (bottom left), and for the spring season (bottom right). The AFWA configuration is shown in blue and the QNSE configuration in red. The vertical bars represent the 99% CIs. Associated with the second y-axis, the light grey line is the adjusted base rate, or the ratio of observed grid box events to the total number of grid boxes in the domain, by threshold.

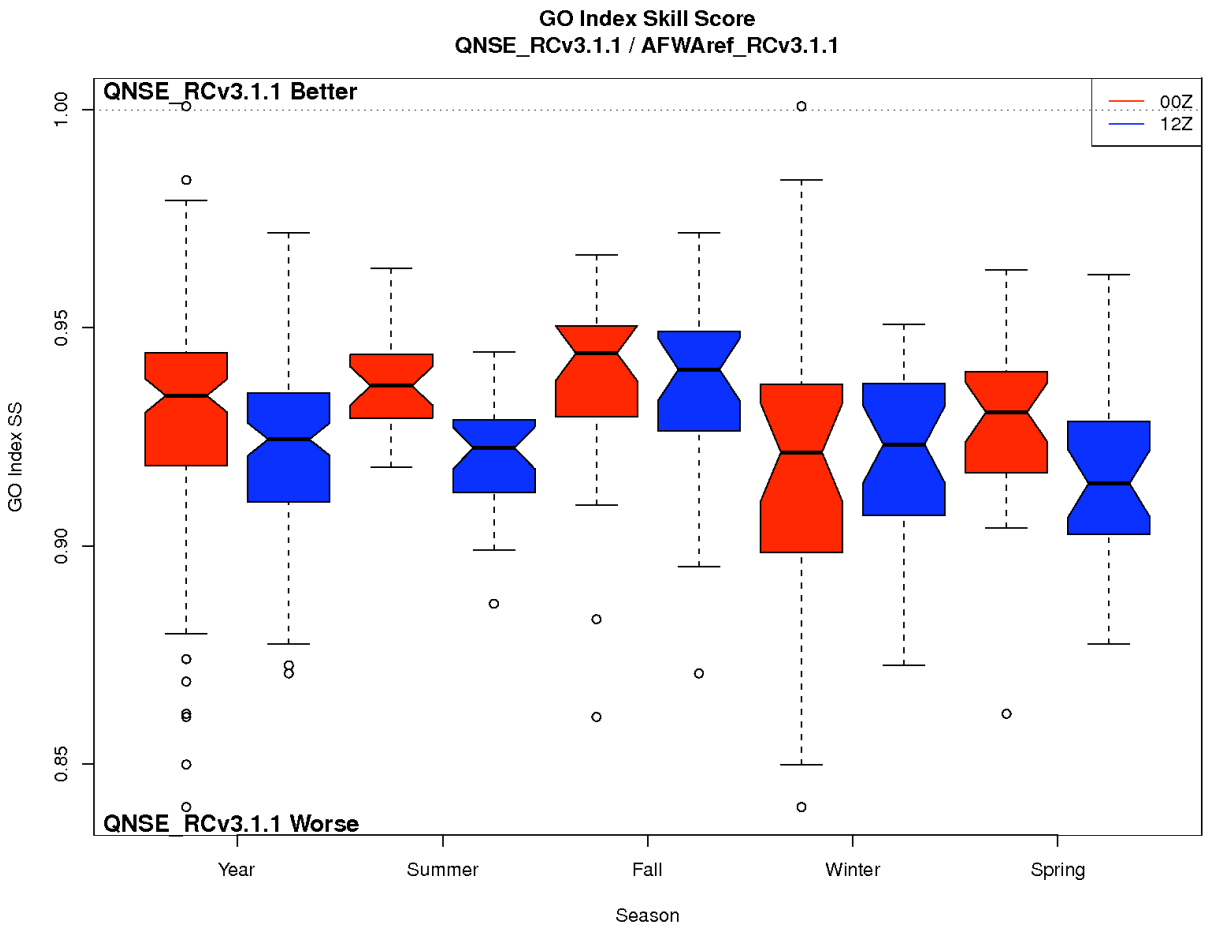

Figure 20. Boxplot of GO Index values aggregated across the entire year of cases and for each season stratified by initialization time (00 UTC – red, 12 UTC – blue). The "waist" of the boxplot indicates the median value, while the width of the notch around the waist is an approximation of 95% confidence about the median. The whiskers represent the largest values that are not outliers, while the circles are classified as outliers.

Appendix A: Case list

00 UTC Initialization	12 UTC Initialization
June 2008: 4, 7, 10, 13, 16, 19, 22, 25, 28	June 2008: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
July 2008: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	July 2008: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
August 2008: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	August 2008: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
September 2008: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	September 2008: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
October 2008: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	October 2008: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
November 2008: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28	November 2008: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
December 2008: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	December 2008: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
January 2009: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	January 2009: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
February 2009: 2, 5, 8, 11, 14, 17, 20, 23, 26	February 2009: 3, 6, 9, 12, 15, 18, 21, 24, 27
March 2009: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	March 2009: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
April 2009: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	April 2009: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28
May 2009: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	May 2009: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31

Appendix B: Subset of WRF ***namelist.input***

```
&time_control
run_hours                           = 48,
interval_seconds                    = 10800,
history_interval                    = 180,
frames_per_outfile                  = 1,
restart                             = .false.,
io_form_history                     = 2,
/
&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,
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                      = .true.,
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,
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                        = 0,
zdamp                           = 5000.,
dampcoef                        = 0.01
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,
pd_moist                        = .true.,
pd_scalar                       = .false.,
pd_chem                         = .false.,
pd_tke                          = .false.,
/
&bdy_control
spec_bdy_width                  = 5,
spec_zone                       = 1,
relax_zone                      = 4,
specified                       = .true.,
/
```