The Developmental Testbed Center's Final Report on the performance of the Noah multi-parameterization (Noah-MP) land surface model in the Air Force Weather Agency's weather forecast system

28 February 2014

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. A recently added surface scheme to WRF was the Noah land surface model with multi-parameterization options (Noah-MP; Niu et al. 2011), the performance of which the Air Force Weather Agency (AFWA) was interested in assessing for their operational configuration. To address this request, the Developmental Testbed Center (DTC) performed a rigorous test and evaluation which assessed forecast performance when substituting AFWA's current operational land surface scheme (Noah LSM) with the Noah-MP. The test was conducted in a functionally similar operational environment to AFWA; each configuration was initialized with a 6-hour "warm-start" spin up, including the WRF Data Assimilation (WRFDA) component. Version 3.5.1 of the WRF model with the Advanced Research WRF (ARW; Skamarock et al. 2008) dynamic core was used. The only difference between the two configurations extensively tested was the LSM used; this allowed for a direct assessment of how differing LSMs affect the performance of the model. For this testing, the two configurations will be referred to as AFWAOC (when using Noah LSM) and NoahMP, with AFWAOC used as the baseline. In addition to documenting the performance of the two configurations against each other, the AFWAOC was designated as a DTC Reference Configuration (RC) and the results have been made available to the WRF community.

2. Experiment Design

For this test, the end-to-end forecast system consisted of the WRF Preprocessing System (WPS), WRF Data Assimilation (WRFDA) system, WRF, Unified Postprocessor (UPP) and the NCAR Command Language (NCL) for graphics generation. Post-processed forecasts were verified using the Model Evaluation Tools (MET). In addition, the full data set was archived and is available for dissemination to the user community upon request. The codes utilized were based on the official released versions of WPS (v3.5), WRFDA (v3.5), WRF (v3.5.1), UPP (v2.1), and MET (v4.1). MET included relevant bug fixes that were checked into the code repository prior to testing.

2.1 Forecast Periods

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

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

Missing forecasts:

Affected Cycle	Missing data	Reason
2011080112	WRF output	Bad SST input data

2011082400	WRF output	Missing SST input data
2012050312	WRF output	Missing GFS input data
2012050612	WRF output	Bad obs_gts input data
2012060400	WRF output	Bad SST input data

Missing verification:

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

2.2 Initial and Boundary Conditions

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

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

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

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

2.3 Model Configuration Specifics

2.3.1 Domain Configuration

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

2.3.2 Model Configuration

The table below lists AFWA's current OC and the NoahMP replacement configuration that were used in this testing.

	Current AFWAOC	NoahMP Replacement Config
Microphysics	WRF Single-Moment 5 scheme	WRF Single-Moment 5 scheme
Radiation LW and SW	RRTM/Dudhia schemes	RRTM/Dudhia schemes
Surface Layer	Monin-Obukhov similarity theory	Monin-Obukhov similarity theory
Land-Surface Model	Noah	Noah-MP
Planetary Boundary Layer	Yonsei University scheme	Yonsei University scheme
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.

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

2.3.3 Noah-MP Configuration

Noah-MP uses multiple options for key land-atmosphere interaction processes. The settings used for this test included: dynamic vegetation turned off, Ball-Berry type stomatal resistance, soil moisture (β) factor for stomatal resistance set to CLM type, surface layer drag coefficient based on Monin-Obukhov, original surface and subsurface runoff (free drainage), no iteration for the supercooled liquid water option, linear effect for the soil permeability option (leading to higher permeability), two-stream radiative transfer option applied to vegetated fraction, Canadian Land Surface Scheme (CLASS) for ground surface albedo, precipitation partitioning between snow and rain based on Jordan (1991), soil temperature lower boundary condition at 8 m from the input file, and a semi-implicit snow/soil temperature time scheme.

2.4 Post-processing

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

3. Computational Efficiency

For the 239 initializations that ran to completion, the central processing unit (CPU) time required to run WRF for the two configurations was calculated to assess the increase in computational demands when using the two different LSMs. This testing effort was conducted on an IBM system, and each model initialization was run on 512 processors. The updated NoahMP LSM, being relatively more sophisticated than the Noah LSM, was expected to increase the CPU time for the NoahMP configuration. Overall, a

consistent difference in computational run time between the AFWAOC and the NoahMP configuration was noted indicating the NoahMP configuration, on average, takes 6.5% longer to run to completion.

4. Model Verification

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

Several domains were verified for the surface and upper air, as well as precipitation variables. Area-averaged results were computed for the CONUS domain, East and West regions, and 14 sub-regions (Fig. 3). While only a portion of the full results will be discussed in detail for this report, all East, West, and sub-domain results are available on the DTC webpage established for this particular testing and evaluation activity (http://www.dtcenter.org/eval/afwa_test/wrf_v3.5.1/). 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 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 difference statistics include zero, the differences are not statistically significant. Due to the nonlinear attributes of frequency bias, it is not amenable to a pair-wise difference comparison. Therefore, the more powerful method to establish SS could not be used and, thus, a more conservative estimate of SS was employed based solely on whether the aggregate statistics, with the accompanying CIs, overlapped between the two configurations. If no overlap was noted for a particular threshold, the differences between the two configurations were considered SS.

Due to the large number of cases used in this test, many SS pair-wise differences were anticipated. In many cases, the magnitude of the SS differences was quite small and did not yield practically meaningful results. Therefore, in addition to determining SS, the concept of establishing practical significance (PS) was also utilized 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://www.wmo.int/pages/prog/gcos/documents/gruanmanuals/CIMO/CIMO_Guide-7th_Edition-2008.pdf). To establish PS between the two configurations, the following criteria was applied: temperature and dew point temperature differences greater than 0.1 K and wind speed differences greater than 0.5 m s^{-1}. PS was not considered for metrics used in precipitation verification [i.e., Gilbert Skill Score (GSS) or frequency bias] because those metrics are calculated via a contingency table, which is based on counts of yes and no forecasts.

4.1 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.

4.2 Precipitation

For the QPF verification, a grid-to-grid comparison was made by first bilinearly interpolating the precipitation analyses to the 15-km model integration domain. This regridded analysis was then used to evaluate the forecast. Accumulation periods of 3 and 24 hours were examined. NCEP Stage II analysis was used as the observational dataset for both the 3- and 24-hour accumulations. 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 GSS and frequency bias. For the precipitation statistics, a bootstrapping CI method was applied.

4.3 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 AFWAOC and NoahMP 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 AFWAOC configuration has higher skill and values greater than one indicate the NoahMP configuration has higher skill.

5. Verification Results

This testing effort revealed many differences between the two configurations, with the largest differences seen at the surface and in the lowest vertical levels. Due to the nature of testing land-surface models, which have the largest impacts at and near the surface, this finding is not surprising. Due to this, the focus of the evaluation will be on surface variables, excluding precipitation, which did not show many differences; all verification plots (by type, metric, lead time, threshold, season, etc.), including vertical levels and threshold series, can be accessed on the project webpage provided in Section 3. The first part of the evaluation will dissect configuration performance over the CONUS for all temporal aggregations for the standard verification metrics. In addition to the time series plots provided,, further investigation of forecast performance for both configurations over diverse regions of the CONUS was 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 differences are apparent, both regionally and between configurations.. In addition, three sub-regions were chosen for a more in-depth analysis, and the results are provided in Appendices C – E; these regions were evaluated to highlight areas with larger differences between the two configurations and investigate how differing land use categories and topography impact the model simulations.

Differences between the two configurations are computed by subtracting NoahMP from AFWAOC. BCRMSE is always a positive quantity, and a perfect score is zero. This results in positive (negative) differences indicating the NoahMP (AFWAOC) 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 AFWAOC (NoahMP) 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, and forecast lead time aggregated over the CONUS domain for surface fields only is summarized in Tables 1-3, where the favored configuration is highlighted.

5.1 CONUS Surface Analysis

5.1.1 Temperature BCRMSE and Bias

For both configurations, a diurnal trend with a general increase in 2 m temperature BCRMSE is seen for all aggregations over the CONUS for both the 00 and 12 UTC initializations (Fig. 4). While both configurations have a diurnal signal, depending on temporal aggregation, the signal may be offset with the other configuration (i.e., peak in error is slightly shifted between configurations). This timing difference is most prominent in the summer aggregation, where peak BCRMSE values occur at times valid at 00 UTC for AFWAOC but occur three hours prior for NoahMP. The weakest diurnal signal is

observed in the annual aggregation for both configurations, while in the winter aggregation, AFWAOC has less diurnal variation that NoahMP. A number of SS and PS pair-wise differences are noted, with most being PS and occurring in the early-to-middle portion of the forecast period (Table 1). A majority of PS pair-wise differences favor AFWAOC, typically at and around times valid 15 – 18 UTC. During the summer aggregation at times valid 03 – 09 UTC, NoahMP is a PS better performer than AFWAOC.

When considering 2 m temperature bias, AFWAOC has a diurnal signal, regardless of initialization time or temporal aggregation, with the smallest signal in the summer aggregation (Fig. 5); maximum errors occur at times valid between 21 – 03 UTC, with exact timing dependent on temporal aggregation. A SS cold bias is noted for almost all forecast lead times and both initializations with exception to a few unbiased lead times in the fall and winter aggregations. A minimum in errors (i.e., typically the smallest cold bias) is seen at times valid around 12 – 15 UTC. NoahMP, on the other hand, displays a bimodal distribution, with peak bias values valid at 18 and 00 UTC over the CONUS domain for the annual aggregation. The temporal aggregations also have characteristics of a multiple-peak bias curve, with the fall aggregation having the weakest signal of this behavior. All forecast lead times for all temporal aggregations and both initializations have a SS cold bias, with exception to a single forecast lead time in the fall aggregation (00 UTC initialization). All differences for 2 m temperature bias are PS, with a weak diurnal signal in which configuration is favored (Table 4). For times valid between 18 – 21 UTC, if there are differences, NoahMP is the better performer; for times valid between 12 – 15 UTC, if there are differences, AFWAOC is the favored configuration. Valid times between 03 – 09 UTC have a dependence on seasonal and temporal aggregation when determining the better performer.

In the summer aggregation for the 30 h forecast lead time for all 00 UTC initializations (00f30; valid 06 UTC), a cold bias for both configurations is noted from the Midwest (MDW) extending eastward to the Atlantic Coast and southward into the Lower Mississippi Valley (LMV); cold biases are also observed along the Northwest Coast (NWC; Fig. 6a-b). In MDW, LMV, and NWC, NoahMP forecasts were typically closer to the observations than AFWAOC (i.e., less of a cold bias). For both configurations, warm biases are noted at a majority of stations in the western part of the Northern Plains (NPL) and the northern part of the Southern Plains (SPL). When examining 00f42 (valid 18 UTC), the AFWAOC has an overall cold bias (Fig. 6c). While NoahMP also exhibits a cold bias in several regions, a warm bias generally clustered in LMV and the Gulf of Mexico Coast (GMC) is noted (Fig. 6d). In addition there is a swath of warm bias values for NoahMP along the Appalachians (APL). While there is an indication of a warm bias for AFWAOC in these areas, the values are higher for NoahMP. Both configurations have large warm biases (>3° C) along the Southwest Coast (SWC) and southern portions of NWC. Generally, for western parts of MDW, the AFWAOC forecast bias is closer to 0 (smaller warm or cold bias) than NoahMP, making it a better performer. Transitioning to the fall season, focus was placed on 00f36 (valid 12 UTC), where the differences between the two configurations are small, but are noted in nearly all of the regions (Fig. 7). In the East, both configurations have an overall cold temperature bias, but with a few areas exhibiting a warm bias. Along the Southeastern Coast (SEC), a warm bias was noted, which was more widespread for NoahMP than AFWAOC; conversely, AFWAOC has a stronger warm bias in western MDW. This warm bias for AFWAOC spread westward into NPL and the Northern Rocky Mountains (NMT), as well. Two forecast lead times were chosen for the winter season to highlight the strong diurnal signal in temperature bias (00f30 and 00f48; valid at 06 and 00 UTC, respectively). At the 30 h forecast lead time, the CONUS has an overall cold bias, with exception to the NPL, SPL and SEC regions which have a near-zero or slight warm bias (Fig. 8a-b). In most regions, AFWAOC has forecasts that are closer to the observations than NoahMP, with some of the largest differences between the two configurations occurring in MDW. At the 48 h forecast lead time, there is a cold bias spread across the entire CONUS, with very few areas indicating a warm bias (Fig. 8c-d). In MDW, while both configurations have a cold bias, NoahMP has a stronger colder bias over much of the area than AFWAOC. In the West, however, the AFWAOC has a stronger, more widespread cold bias compared to NoahMP. For spring, 00f42 (valid at 18 UTC) was selected to show several areas of large differences between the two configurations (Fig. 9). With exception to MDW, where NoahMP is consistently too cold in all seasons, NoahMP has warmer 2 m temperature forecasts than AFWAOC over most of the CONUS. The most notable differences occur in GMC, LMV, APL, and the North East Coast (NEC), where a cold bias is noted for AFWAOC. Conversely, NoahMP, across the same regions, shows a mix of warm and cold bias values by

observation station, which translates to an overall near-zero bias in those areas. In the West, both configurations are generally cold, with AFWAOC having a deeper cold bias than NoahMP.

5.1.2 Dew Point Temperature BCRMSE and Bias

Similar to 2 m temperature, a diurnal signal superimposed within a gentle increase in errors with time is present in 2 m dew point temperature BCRMSE for all temporal aggregations, both initializations, and for both configurations (Fig. 10). A slightly amplified diurnal signal with larger errors around times valid 18 – 00 UTC is noted for NoahMP, except in the winter where median values of BCRMSE are consistently higher for NoahMP than AFWAOC. Pair-wise differences favoring AFWAOC are seen for valid times between 15 – 03 UTC (Table 3). Any PS pair-wise differences favoring NoahMP are seen in the summer and spring temporal aggregations, and generally between times valid 09 – 12 UTC.

For all temporal aggregations and both initializations, a prominent diurnal signal in 2 m dew point temperature bias is noted, with higher dew point temperature bias values around times valid 21 – 03 UTC and a minimum in bias values around 12 – 18 UTC depending on the configuration and temporal aggregation (Fig. 11). In the annual, summer, and spring aggregations, NoahMP has a wet bias at most forecast lead times, with exception to occasional lead times valid around 12 – 18 UTC, where there is an unbiased or dry forecast. The two configurations have the largest divergence in bias values in the overnight into morning hours (i.e., valid 03 – 12 UTC); this pattern is observed in all but the winter aggregation. Typically at these valid times NoahMP has higher magnitude values than AFWAOC with the favored configuration dependent on temporal aggregation (Table 3). In contrast, for the winter aggregation, during daytime hours (times valid around 12 – 00 UTC), NoahMP has lower median values than AFWAOC as well as near or below zero median biases (i.e., a dry bias). It should be noted, however, in the winter aggregation NoahMP has large CIs attached to the median bias values. When pair-wise differences are noted, all are PS, with the least amount of differences occurring in the winter temporal aggregation, which is likely due to the large CIs for NoahMP mentioned above. Most PS differences favor AFWAOC; however, there are PS pair-wise differences that favor NoahMP at some forecast lead times, depending on initialization and temporal aggregation.

In the summer aggregation, as was shown in the time series plots, NoahMP has a consistently higher dew point temperature bias than AFWAOC at 00f30 (i.e., valid at 06 UTC); this is valid over all regions, with exception to NWC (Fig. 12a-b). The largest differences between the two configurations are seen in MDW, where NoahMP has moist dew point temperature biases at most observation stations, while AFWAOC generally has dry dew point biases. Large differences are also noted in the West where a number of regions [NPL, SPL, NMT, the Southern Rocky Mountains (SMT), and Southwest Desert (SWD)] indicate NoahMP is more moist than AFWAOC, and in most cases more moist than the observations. Interestingly, at the 42 h forecast lead time, the dry bias noted for AFWAOC over MDW has transitioned to a moist bias and overall has a higher dew point temperature bias than NoahMP (Fig. 12c-d). Over NPL, SPL, SMT, SWD, and the Great Basin (GRB), NoahMP is still too moist as compared to the observations as well as higher than AFWAOC. Conversely, NoahMP over APL has dried out too much as compared to AFWAOC, which has a near-zero bias. For the fall aggregation, 00f36 (i.e., valid at 12 UTC) was chosen to display that while most regions did not have overly large high or low bias values in the aggregated line plots, the two configurations differed spatially in a variety of regions (Fig. 13). In portions of MDW (e.g., MN), NPL and NMT, AFWAOC has a moist bias, while NoahMP has a larger presence of a neutral or dry bias. However, over GMC, SEC, and APL, AFWAOC generally has a dry bias, while NoahMP typically a moist bias. The winter aggregation is where some of the largest regional differences in dew point temperature bias are seen. While variations in bias sign exist within MDW for AFWAOC, at 00f30 (i.e., valid at 06 UTC), AFWAOC has an overall small moist bias over MDW; NoahMP, however, exhibits a strong, consistent dry bias (Fig. 14a-b). Moving west, in NPL, NoahMP has forecasts closer to the observations than AFWAOC, which has a moist bias at most observation stations in the region. When examining 00f48 (i.e., valid at 00 UTC), the pattern observed at 00f30 for NoahMP over MDW intensifies, while interior areas of LMV, GMC, and APL display a large moist bias (Fig. 14c-d). For most regions in the East (excluding NEC), AFWAOC has a mix of dry and moist biases, leading to an overall near-zero aggregate bias. In NPL, NMT, and SMT, AFWAOC has a consistent moist dew point

bias as opposed to NoahMP which typically has a drier dew point bias for NMT and NPL moist bias for SMT. For the spring aggregation, 00f42 (i.e., valid at 18 UTC) was chosen to highlight areas of dew point temperature bias differences (Fig. 15). In MDW, while both configurations have a moist forecast bias, NoahMP has a larger area of bias values greater than or equal to 4° C. Similar findings are seen for SMT, SWD, and GRB with NoahMP. In GMC, SEC, and parts of APL, AFWAOC is typically moister than NoahMP, which has a dry bias at a number of observation stations.

5.1.3 Wind BCRMSE and Bias

For 10 m wind speed BCRMSE, both configurations display a weak diurnal signal with a general increase with forecast lead time for all temporal aggregations and both initialization times (Fig. 16); a weaker trend is present in the fall and winter aggregations. The errors are largest at times valid between 21 – 00 UTC, while the smallest errors are seen at times valid near 12 UTC. No PS pair-wise differences are seen (Table 3); however, a number of SS pair-wise differences are noted. A majority of the SS pair-wise differences favor the AFWAOC for valid times between 15 and 00 UTC and occur most during the first 24 hours of the forecast. Most pair-wise differences favoring NoahMP are seen in the summer aggregation in the overnight to early morning hours.

A prominent diurnal signal in bias is seen for all temporal aggregations and both initializations for 10 m wind speed bias, with highest errors seen at times valid 03 – 12 UTC and with lowest errors at times valid 15 – 00 UTC (Fig. 17). With only a few exceptions in the afternoon and early evening hours, a high wind speed bias is observed regardless of initialization and temporal aggregation; this high bias generally grows with forecast lead time. While a number of SS pair-wise differences are observed, none are PS.
A diurnal trend is noted in configuration performance (Table 3); in overnight hours NoahMP tends to have lower bias values (i.e., closer to 0), while in daytime hours AFWAOC is typically a better performer.

Due to the small differences between the two configurations, only AFWAOC, the baseline configuration, will be discussed regarding spatial distribution of wind biases over the CONUS. To highlight the diurnal signal in wind bias, 00f30 (i.e., valid at 06 UTC) and 00f42 (i.e., valid at 18 UTC) are shown (Fig. 18). A high bias is present for 00f30 across much of the CONUS, with the largest values in upper MDW and along the Atlantic Coast. SPL and NPL generally see the lowest biases, with a low bias seen at a number of observation stations in SPL. At 00f42, bias values over the entire CONUS decrease, with the highest biases still seen in upper MDW and along both the Atlantic and Pacific Coasts. SPL and NPL exhibit a low bias at a number of observation stations, which is reflected in the overall low bias of these regions in the aggregate line plots (not shown). Consistent with Figure 17, the wind speed biases are lower over the whole CONUS during the daytime hours compared to the overnight hours.

5.2 GO Index

For the annual, winter, and spring aggregations, both 00 and 12 UTC initializations have median GO Index values and associated CIs (estimated by the width of the notches about the median on the boxplot) less than one, indicating AFWAOC is the better performer; similar behavior is noted in the 12 UTC initializations for the fall aggregation. The 00 and 12 UTC initializations for the summer aggregation as well as the 00 UTC initializations for the fall aggregation have CIs encompassing one, indicating no SS difference in performance between AFWAOC and NoahMP. In addition, several outliers are observed below a value of one in the annual, fall, and winter aggregations, which signifies AFWAOC outperforms NoahMP in these particular cases.

6. Configuration Comparisons

Apart from the atmospheric temperature, dew point temperature, and wind fields, additional model variables which describe the simulated ground snow and soil conditions, and surface radiative and heat fluxes were examined and compared for the two configurations. Since observations of these variables are not readily available, no verification or evaluation against observations is attempted for these variables. Instead, our focus is on the differences in these fields caused by using the different LSMs in

WRF. Given that both the AFWAOC and NoahMP configurations are initialized with the same LIS datasets, the differences in the snow and soil fields between the two model configurations at a forecast time reflect the cumulative effect from the case cold-start time, through the warm-start processing, to the forecast time examined.

The configuration comparisons for selected variables were analyzed through examining the variables in individual cases as well as seasonal aggregations, and in point locations as well as CONUS aggregations. Two-dimensional maps of seasonal aggregations are plotted to illustrate the typical geographical distributions of the differences between the two configurations (for all lead time and both initializations). Temporal variability of the variables is shown by time series curves comprised of multiple cases and initializations, for the CONUS mean and at selected sites. Locations of the three sites (designated Sites 1-3) discussed in this section are marked in Figure 3. In this section, the overall characteristics of the configuration differences are presented first, followed by analysis for an individual case initialized at 00 UTC of January 27, 2012.

6.1 Snow Conditions

Two major variables that describe the snow conditions on the ground, the snow equivalent water (SWE) and accumulated melted snow mass (ACSNOM), were examined. Differences between the forecasts of SWE in the two configurations are noted in all seasonal aggregations except the summer, mainly in the mountainous regions of the west (Fig. 20). At the cold-start time, the two configurations start from the same SWE; however, the SWE values in NoahMP and AFWAOC quickly diverge during the 6-h warm-start and 48-h forecast period (45 h lead time shown in Fig. 20). While the overall CONUS-mean SWE in the NoahMP forecasts is lower than that in the AFWAOC forecasts, the regional SWE differences between the two configurations show correlations with the terrain height and land-use and vary with the season. In the fall and winter, the NoahMP configuration has significantly less SWE (warm colors in Fig. 20) in the southwest corner of GRB where the land-use types are barren and shrubland. This is also seen over the shrubland areas of SMT and grasslands of NPL during the winter. In contrast, more SWE is forecasted by NoahMP over the Northern Cascade (border of NMT and NWC; cool colors) where the land is mostly covered by evergreen forests. In the Yellowstone area of central NMT, NoahMP has more SWE in the winter and less SWE in the spring. In the spring melting season, it is found that NoahMP generally has less SWE (mainly in SMT and NMT) which is, as further investigation suggests and later results illustrate, a result of stronger melting in NoahMP.

Time Series of SWE at the selected sites show that, during the winter months, Site 1 (in the Yellowstone evergreen forest area) is covered by deep snow, where the SWE value peaks in mid-March and is non-zero through May (Fig. 21a-b). Site 2 (north-central Arizona, shrubland) is basically snow-free in all months, and Site 3 (southwest Wisconsin, crop/grassland) has shallow snow in January-February and is snow-free by March (not shown). The day-to-day forecasts of SWE at Site 1 show that the NoahMP configuration consistently produces more snow in March and less snow in April and May. The model forecasts of accumulated melted snow mass (ACSNOM) at Site 1 (Figure 21c-d) show that, compared to AFWAOC, the NoahMP configuration has less melting in March and more melting in April and May. At Site 3, the NoahMP configuration generally has more melting and less SWE than AFWAOC for the entire winter season from November through February (not shown).

6.2 Soil Properties

6.2.1 Soil Temperature (SOILT)

For both configurations, the soil temperature (SOILT) in the snow-covered regions has very different characteristics from that in the snow-free regions. In general, the top-layer (0-10 cm) soil temperature (SOILT1) for the snow-covered areas experiences little diurnal trend while SOILT1 for the snow-free regions has diurnal cycles similar to those in the surface air temperature. It was found that, for valid times in the afternoon, the SOILT1 values in the NoahMP configuration are generally warmer than those in the

AFWAOC configuration for most snow-covered and snow-free regions in all seasonal aggregations (Fig. 22; valid at 21 UTC). The largest differences (cool colors in Fig. 22) between the two configurations are noted in portions of SPL, LMV, and GMC, as well as over much of SWC and NMT in the spring and summer and can reach more than 4 ^{o}C. The exceptions are SWD and the southwestern corner of MDW where 1-2 ^{o}C colder daytime soil temperature forecasts are seen in the NoahMP configuration in the spring and summer. During the nighttime, in all snow-free regions regardless of season except SEC, the opposite is true: the top-layer soil temperature values in NoahMP are generally 1-2 ^{o}C colder than in AFWAOC (Fig. 23). As a result, for most snow-free regions, the diurnal variation in SOILT1 is more pronounced in the NoahMP configuration. For SEC, the nighttime soil temperature in the NoahMP configuration is warmer in all seasons. For the snow-covered regions (SMT, NMT, GRB and MDW in the spring and winter, and the northern portions of APL and NEC in the winter), the forecast nighttime soil temperature in the NoahMP configuration is 1-2 ^{o}C warmer than the AFWAOC configuration.

The collective time series of SOILT1 at Site 1 (Fig. 24) show that, during late fall when snow starts to occur but the accumulation on the ground is low, the soil temperature is lower in the NoahMP forecasts than in AFWAOC; then during mid-January through March, SOILT1 in NoahMP is higher than in AFWAOC. The winter SOILT1 in both configurations is unresponsive to the solar diurnal cycles and uncorrelated with the surface air temperature. However, during late spring after much of the snow cover has melted, SOILT1 at Site 1 exhibits diurnal cycle modulations, and the NoahMP forecasts have larger variations than the AFWAOC forecasts. At Site 3 which is covered with shallow snow during January - February, the NoahMP configuration has warmer SOILT1 in the winter and is even less correlated with the surface air temperature than the AFWAOC configuration. NoahMP has consistently warmer daytime SOILT1 values in the summer and colder nighttime temperatures in the late spring and early summer compared to AFWAOC (not shown).

6.2.2 Soil Moisture (SOILM)

During the winter and fall, for all forecast valid hours, the top-layer (0-10cm) soil moisture (SOILM1) in the NoahMP configuration is lower (drier) than that in the AFWAOC (Fig. 25; valid at 21 UTC). Significant differences are mainly noted during seasons and across regions with snow-cover. Conversely, during the spring season over areas with active snow melt, NoahMP has a wetter SOILM value for the top layer. Examinations of individual cases reveal that, for each case, the two configurations start from the same soil moisture at the cold-start time, then the SOILM1 values in NoahMP begin a sharp decrease immediately after model integration begins. The largest rates of decrease are noted during the 6-h pre-forecast and the first 12hours of the forecast integration. The variation of SOILM1 in AFWAOC during the pre-forecast period and the 48-h forecast time is relatively small. The drying of SOILM1 directly after the start of the model integration in NoahMP is not noted in the spring and summer. While values during the summer and fall seasons are similar between the two configurations (Fig 26c-d), at Site 1, the drying of SOILM1 in NoahMP occurs throughout the winter season and is most significant in the March timeframe (Fig. 26a); it continues through the melting month of April and is reversed around mid-May when NoahMP produces wetter SOILM1 (Fig. 26b). At Site 3 (in MDW), the drying of SOILM1 in NoahMP is present in all forecasts in January and February, and the timing of the trend reversal is closer to early March (not shown), corresponding to the time of depletion of ground snow cover at that particular site.

It was also found that, for the NoahMP configuration, the soil moisture at deeper layers (SOILM2 for 10-40 cm, SOILM3 for 40-100 cm, and SOILM4 for 100-200 cm) may exhibit different trends than that of the top-layer soil moisture. For example, at Site 3 during January and February, the soil moisture at the two top-layers have a drying trend with model integration time while the soil moisture at the two deeper layers have a moistening trend (Fig. 27). As a result, for SOILM1 and SOILM2, the NoahMP forecasts are mostly drier than the AFWAOC forecasts, while for SOILM3 and SOILM4, the NoahMP forecasts are mostly wetter than the AFWAOC forecasts. The column-total soil moisture (sum of SOILM1-4) at Site 3 in NoahMP remains lower than in AFWAOC for January and February (not shown).

Further investigations revealed that the drying trend in the top soil layers in the NoahMP configuration is related to the soil permeability options in the NoahMP LSM in WRF (*opt_inf*). In the year-long test,

opt_inf = 1 is used (Appendix B) which assumes a linear effect of soil ice on infiltration and more permeable soil properties. In a re-run of one test case using *opt_inf* = 2 (which assumes non-linear effect of soil ice and less permeable soil), the drying trend in the top-layer soil moisture almost completely disappeared. However, this is not to say that less permeable soil setting is preferred in the NoahMP configuration. A frozen soil scheme that allows a greater soil permeability was purposely included for the NoahMP LSM to overcome the impeding effect of frozen soil on infiltration in the Noah LSM (Niu et al. 2011). This result does point to possible spin-up problems in the more permeable frozen soil scheme in NoahMP LSM when the model is initialized with datasets from a Noah-based LIS system. Personal communications with Dr. Fei Chen (NCAR) suggested that NoahMP spun-up initial conditions are preferred when the WRF model is run with the NoahMP LSM option.

6.3 Surface Albedo

The overall surface albedo in the NoahMP configuration was found to be significantly lower than the AFWAOC configuration, especially in the West, over snow-cover areas and shrubland land-use categories. During daytime in the winter, the CONUS mean albedo in the NoahMP configuration is about 0.03-0.2 lower than in the AFWAOC configuration (Fig. 28a). At the snow-covered sites, the difference between the two configurations can be as big as 0.3. At Site 1 (Fig. 28c), with snow accumulating from November through March, the daytime albedo in AFWAOC remains close to 0.6, while in NoahMP the value is closer to 0.3 or 0.4 (the fall values not shown). As the snow melts in April-May, the albedo in both configurations decreases and the NoahMP values remain lower than the AFWAOC values (Fig. 28d). In June through the summer (not shown), the albedo at Site 1 (evergreen forest) in both configurations is about 0.18. At Site 2 (shrubland), the values for both configurations change very slowly with the season (Fig. 28e-f). The spring values are higher than the winter values for NoahMP while the opposite is true for AFWAOC leading to the two configurations have closer values to each other in the spring (difference of about 0.05) and diverging to a degree in the winter (with differences of about 0.15). At Site 3 (in MDW, crop/grassland), in January and February when shallow snow is on the ground, the daytime albedo values in both configurations vary from case to case; more variation is seen in the NoahMP configuration with some extreme values at dusk hours (Fig. 28g-h).

Unlike the ground snow and soil variables, the surface albedo is not a prognostic variable and is not initialized from the LIS datasets. In Noah LSM (AFWAOC), the albedo values are prescribed based on the surface conditions. In NoahMP LSM, the albedo values are diagnosed from the surface solar radiation budget and is affected by the physical processes at the surface and canopy. In general, apart from the snow accumulating or melting effect, there should be little change in albedo values in both configurations during the 48-h forecast period. The diurnal oscillations of albedo in the NoahMP configuration is an artifact of the diagnostic procedures. Since there is no incoming solar radiation at night, the nighttime albedo values in NoahMP are not meaningful and should be ignored. In addition, in the NoahMP configuration, some extreme values of albedo also appear at dawn/dusk when the incoming solar radiation is very low (e.g. for Site 3 in winter). They do not imply an issue with the solar radiation budget. The net absorbed solar radiation in the two configurations will be further discussed in the case study in Section 6.5.

6.4 Surface Heat Fluxes and Energy Budget

6.4.1 Latent Heat Flux (LH)

Overall, for CONUS, the latent heat flux (LH) at the surface is lower in the NoahMP configuration than in the AFWAOC configuration, for both January and July (Fig. 29a-b), where the daily peak values of the CONUS-mean LH at noontime are consistently lower in NoahMP than those in AFWAOC. From January to July, there is a 6 - 8 fold increase in LH for both configurations. On the average, the NoahMP noontime LH values are approximately 10 W m^{-2} or 20% lower than the AFWAOC values in January, and about 25 W m^{-2} or 10% lower than the AFWAOC values in July.

At Site 1, the differences between the two configurations are larger than the muted CONUS-mean (Fig. 29c-d). During January when the site is covered with snow, there is virtually no latent heat flux in the NoahMP configuration, while in the AFWAOC configuration the LH values oscillate between -20 to 50 W m^{-2}. The daily peak values at Site 1 in July are about 100-250 W m^{-2} for NoahMP and 250-400 W m^{-2} for AFWAOC. The differences in latent heat flux between the two configurations at Site 2 are quite different from those at Site 1 (Fig. 29e-f). In both January and July, the NoahMP values are higher than the AFWAOC values. The overall July values at this site for both configurations are lower than the CONUS mean. The LH behavior at Site 3 in January is similar to that at Site 1, while in July the LH values for Site 3 are higher than those at Site 1 for both configurations (Fig. 29g-h). The NoahMP LH values at Site 3 in July are approximately 50-150 W m^{-2} or 10-25% lower than the AFWAOC LH values.

6.4.2 Sensible Heat Flux (HFX)

In general, the surface sensible heat flux (HFX) in both configurations exhibits a strong diurnal signal, with larger positive values around noontime and smaller, negative values at night (Fig. 30). For the CONUS mean during both January and July, the NoahMP configuration forecasts higher daytime heat flux values and lower (more negative) nighttime values. A similar trend is generally true at the three selected sites. The configuration differences in January are the largest at Site 1 (covered with deeper snow). In terms of relative magnitudes of LH and HFX for both configurations, Site 2 has smaller LH but larger HFX while Sites 1 and 3 have larger LH (especially in July) but smaller HFX.

6.4.3 Ground Heat Flux (GRDFLX)

When the analysis first began, it was found that the ground heat flux (GRDFLX) values in the two configurations had opposite signs. Consultation with the WRF NoahMP LSM development group (Barlage, personal communications) confirmed that the variable in the two LSMs are defined differently. In Noah LSM, GRDFLX is positive when heat is going toward the surface, while in NoahMP LSM, GRDFLX is defined as positive when heat is leaving the surface. It is important to align the definitions when comparing the variable fields in the two configurations. In the discussions to follow, the definition in the NoahMP LSM is adopted and GRDFLX is positive if heat is leaving the surface.

The ground heat flux also exhibits strong diurnal cycles, having larger positive values around noontime and smaller, negative values at night (Fig. 31), with exception to the snowy sites. The magnitudes of the negative flux at night are generally larger (more negative) than the magnitudes of nightly HFX. For the CONUS mean, the NoahMP configuration has higher daytime ground heat flux values and lower (more negative) nighttime values than AFWAOC, similar to that for HFX. The same is true for Site 1 and 3 in July. During January, the ground heat flux values at Site 1 are generally small for both configurations with more variation noted for NoahMP. More negative flux values (heat from soil to surface) are noted at Site 3 in January for the AFWAOC configuration. The configuration differences appear smaller for Site 2; however, the magnitude of the GRDFLX values at Site 2 for both configurations are significantly higher than those at Site 1 and 3.

6.5 Case Study – 27 January 2012, 00 UTC Initialization

To further investigate results that have been presented, a single initialization was chosen to focus on how surface energy components differ between the two configurations and how their surface energy budgets contribute to and influence the resulting forecasts. Because increased differences were noted during time periods of snow cover, a winter case (27 January 2012 00 UTC initialization) was selected to highlight some of the largest differences between the two configurations (e.g., differences in 2 m temperature and dew point temperature). Individual components of the surface energy budget are shown for the 42-h forecast lead time (valid at 21 UTC); in addition, a time series of the 48-h forecast with each of the components as well as the difference between the two configurations is presented.

For context, the meteorological setup for the chosen case study was particularly active, with a frontal system moving through the area east of the Mississippi River. Behind the first system was a second

system building in the Plains, with yet another system moving onshore into the Pacific Northwest. Throughout the forecast period, most regions experienced affects from this active pattern (e.g., snow, periods of cloudy and clear skies).

6.5.1 Radiation Flux Component

Due to the differences between the configurations in the winter (e.g., SWE, soil moisture, 2 m temperature and dew point temperature), a concentrated look at the surface radiation fluxes is provided in order to help determine the potential source of differences. Prior to this investigation, AFWAOC was anticipated to have higher albedos over snow-covered areas, due to modifications that have been made to Noah LSM over time to adjust for snow melting too fast. In general, this was seen, especially over the West; however, in areas over the Midwest and Northern Plains, NoahMP has larger albedo values than AFWAOC (Fig. 32). These areas are typically snow-covered cropland or grassland, where theland types generally consist of short vegetation, which do not require much snowfall to be completely and uniformly covered. The flat, smooth snow-covered land is likely leading to the higher albedos seen in NoahMP. It is possible that smaller amounts of vertical mixing in this type of environment may also be contributing to the strong cold bias seen in NoahMP. Similar spatial distributions are also seen in the net short-wave flux at the surface, due to the relationship with albedo (Fig. 33). The difference in net short-wave radiation flux is strongly driven by the differences in albedo, with the largest differences occurring over snow-covered areas or in areas where the clouds from the frontal system are impacting the radiation budget. Overall, when looking at the net long-wave radiation fluxes at the surface, AFWAOC has higher values CONUS-wide at the 42-h forecast lead time, with the largest differences near the comma head associated with the first frontal system (Fig. 34).

6.5.2 Heat Flux Component

A curious behavior of NoahMP is the small, near-zero latent heat flux values over the snow-covered areas of the West and Midwest; this behavior is not seen in AFWAOC (Fig. 35). The highest latent heat fluxes for both configurations are in the Southeast US, where in areas of LMV, NoahMP has higher latent heat flux values than AFWAOC, while AFWAOC has higher values in the SEC region. These differences described above may potentially explain some of the large differences in 2 m dew point temperature biases seen in the winter between the two configurations (particularly, Fig. 14b,d).

Both configurations have a neutral to positive sensible heat flux at the surface over a majority of the CONUS with exception to the NW CONUS where some areas have a negative sensible heat flux, especially for AFWAOC (Fig. 36). In general, NoahMP has higher values than AFWAOC; there are some areas in MDW and LMV where it is opposite. The largest differences, sensible of 200 W/m^2, are observed in the Rocky Mountains as well as in Iowa, where AFWAOC had more snow cover than NoahMP at this particular time; in these areas NoahMP has higher sensible heat fluxes than AFWAOC. Overall, these results would contribute to NoahMP warming more than AFWAOC, which is seen in the overall 2 m temperature bias statistics (particularly, Fig. 5d).

As described in Section 6.4.3, to keep the analysis consistent and meaningful, the ground heat flux from AFWAOC was adjusted to match the definition of NoahMP. Both configurations have near-zero values of ground heat flux over snow-covered areas (Fig. 37). Over the West, areas of SPL and LMV, and within the SEC region, the NoahMP configuration typically has higher values than AFWAOC.

6.5.3 Energy Budget Time Series

6.5.3.1 Site 1 – Snow-covered Needleleaf Forest

The largest differences in radiation budget occur during the daytime hours between valid times of 18 – 00 UTC (Fig. 38). NoahMP has overall larger net and downward short-wave radiation; the upward component of short-wave radiation is larger for AFWAOC, which can be attributed to the higher albedo values discussed above. In conjunction with larger sensible heat flux values for NoahMP, this could lead to stronger warming and higher temperatures for NoahMP during the peak heating. NoahMP has very

little fluctuation in latent heat flux, while AFWAOC has a weak diurnal undulation, with peak, positive latent heat fluxes during the daytime hours and negative values overnight. This is indicative of increased evaporation taking place in AFWAOC, which would be a cooling process. Also, heading into day two of the forecast period, as the system from the Pacific Northwest moves inland and Site 1 becomes cloudy, AFWAOC has a higher downward long-wave radiation flux, which could contribute to cooler temperatures with a lack of solar impacts during the daytime but warmer temperatures during the overnight hours. The result of AFWAOC being warmer than NoahMP in the overnight hours and cooler during the daytime is a consistent result seen in earlier presented verification for this particular location.

6.5.3.2 Site 2 – Not Snow-covered Shrubland

Similar to Site 1, Site 2 also shows larger having larger sensible heat flux for NoahMP, which could contribute to the warmer temperatures seen during the daytime hours for NoahMP; conversely, AFWAOC has larger sensible heat fluxes during the overnight hours, which, along with the consistent 20 W/m^2 greater net long-wave radiation values, could act to warm the temperature (Fig. 39). Both configurations have similar curves for downward short-wave radiation flux at the surface; however, differences are again noted in upward short-wave radiation, where NoahMP typically has lower peak values in the daytime hours. These differences affect the overall net short-wave radiation flux at the surface and are attributed to the albedo differences noted previously (i.e., AFWAOC has larger albedo values than NoahMP).

7. Summary

An end-to-end sensitivity test was conducted to test and evaluate the resulting forecast performance when substituting AFWA's current operational land surface scheme (Noah LSM) with the Noah-MP LSM. Each configuration included a 6-hr warm-start data assimilation procedure and was run over the same set of cases, spanning one full year. The goal of this testing effort was to assess the current state of development within the Noah-MP LSM and to inform developers of future tuning that may need to occur prior to consideration for operational implementation.

The employed testing methodology allowed for pair-wise differences to be computed for several verification metrics, with an assessment of SS and PS pair-wise difference. Overall, a large number of SS and PS pair-wise differences were observed; however, a sensitivity in which configuration was favored is dependent on verification metric, temporal aggregation, initialization time, vertical level, lead time, and threshold. The focus of this report was concentrated on the surface variables, where the largest differences between the two configurations were noted (not unexpected given the focus of the test). A large number of PS pair-wise differences were observed for the 2 m temperature and dew point temperature variables, while none of the pair-wise differences in 10 m wind speed were PS. For both 2 m temperature and dew point temperature, the largest differences between the two configurations occurred overnight. For 2 m temperature, the BCRMSE results tended to favor the AFWAOC, while the bias results were a bit more mixed. For 2 m surface dew point temperature, a majority of the PS pair-wise differences favored the AFWAOC for both BCRMSE and bias. A preference for the AFWAOC was also found in the GO Index calculation for all seasons except summer, which had no SS difference in performance between the two configurations.

Spatial distribution plots helped diagnose regional patterns that were not evident in the time series plots, with major findings summarized below. When examining 2 m temperature for winter, while NoahMP displays a stronger cold bias across much of the East during the overnight hours as compared to AFWAOC, the opposite is true over the West during the daytime, where AFWAOC has a colder bias. For the summer season, the largest difference between the two configurations occurs across the southern states within the LMV and GMC regions, where NoahMP has a warm bias during the day. When examining 2 m dew point temperature bias, a consistent wet bias is seen during the summer season across the southwest CONUS for NoahMP. This trend is also seen within the winter season during the daytime hours. In the East, during the winter season a significant dry bias is noted across MDW regardless of time of day, while the southern portion of the CONUS exhibits a significant wet bias during the daytime hours.

In general, areas with snow cover produced significantly different forecast results for 2 m temperature and dew point temperature among the two configurations, especially over the MDW region. When looking at the differences over snow-covered areas, the NoahMP configuration frequently exhibited lower SWE than the AFWAOC with a distinct exception to land-use types classified as evergreen needleleaf. However, overall NoahMP produced a greater snow melt value during the spring season. Interestingly, the soil moisture near the top-most layers for NoahMP went through a drying process throughout the model integration time during the winter and early spring season; the moisture may have drained too efficiently into the bottom two soil layers. This may be an indication of an imbalance at the model initialization time when using the AFWA LIS with Noah LSM to initialize the NoahMP configuration for this test.

The Noah LSM has been extensively used in the research and operational community for a number of years and has benefited from a large number of testing and evaluations over time. The Noah-MP LSM is an advancement in the representation of physical processes; however, because it is relatively new to the research community, additional modifications and tuning may be necessary prior to operational implementation.

8. References

Niu, G.-Y., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res., 116, D12109, doi: 10.1029/2010JD015139.

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

Table 1. SS (light shading) and PS (dark shading) pair-wise differences for the AFWAOC and NoahMP configurations run with WRFv3.5.1 (where the highlighted configuration is favored) for surface temperature BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

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

Table 2. SS (light shading) and PS (dark shading) pair-wise differences for the AFWAOC and NoahMP configurations run with WRFv3.5.1 (where the highlighted configuration is favored) for surface dew point temperature BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

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

Table 3. SS (light shading) and PS (dark shading) pair-wise differences for the AFWAOC and NoahMP configurations run with WRFv3.5.1 (where the highlighted configuration is favored) for surface wind BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

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

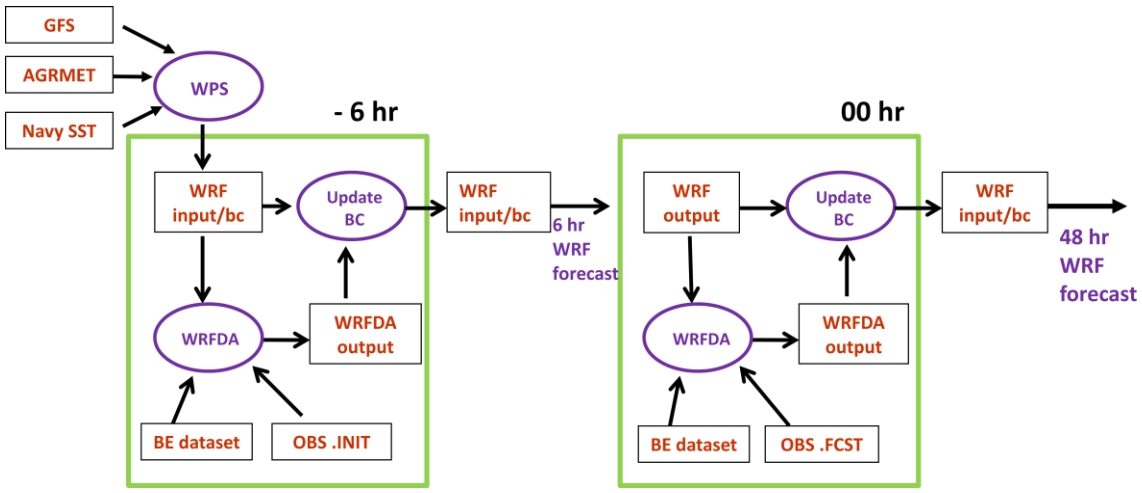


Figure 1. Overview of 6-hr "warm start" spin-up.

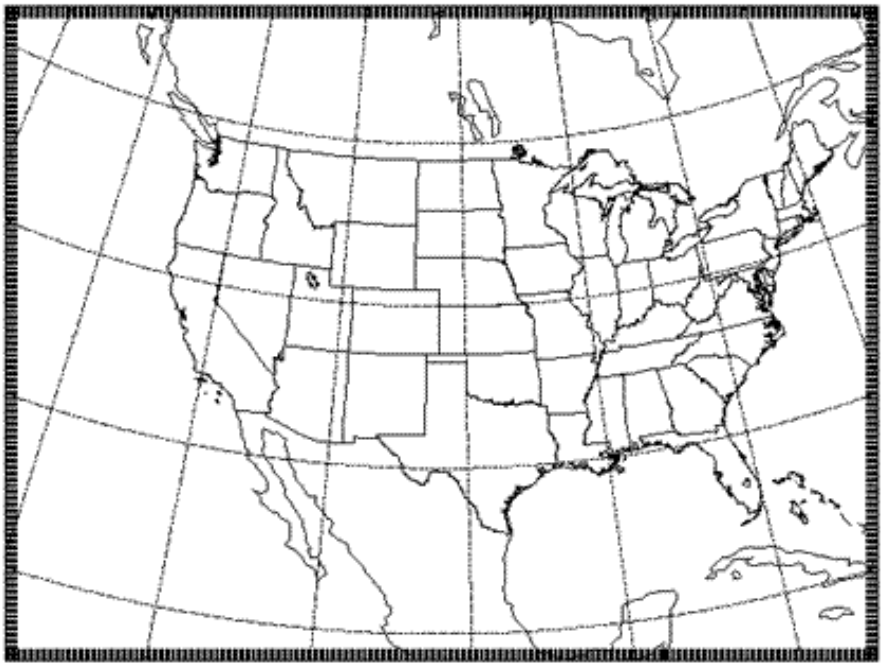


Figure 2. WRF-ARW computational domain.

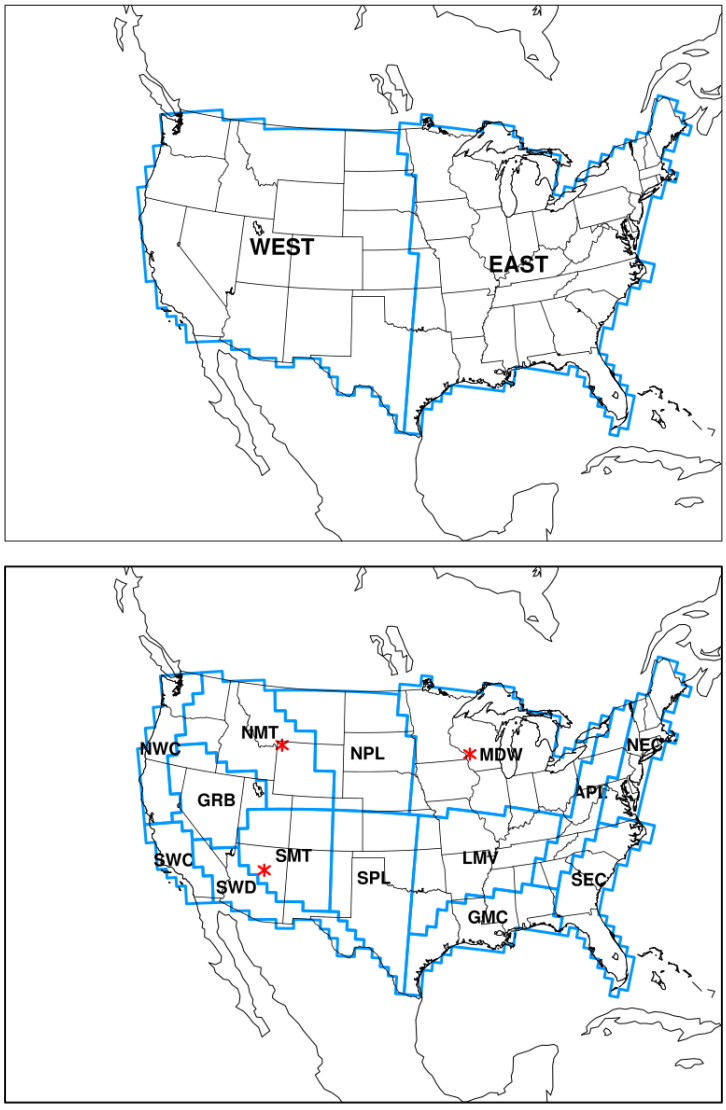
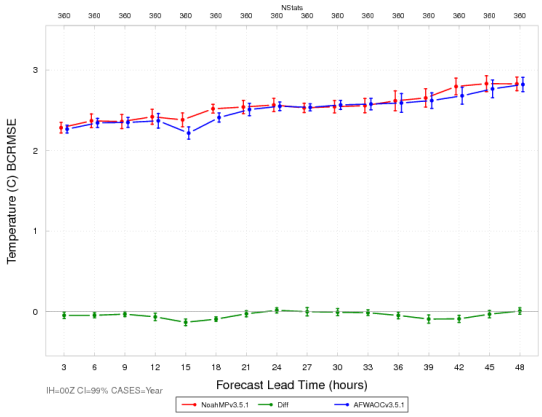
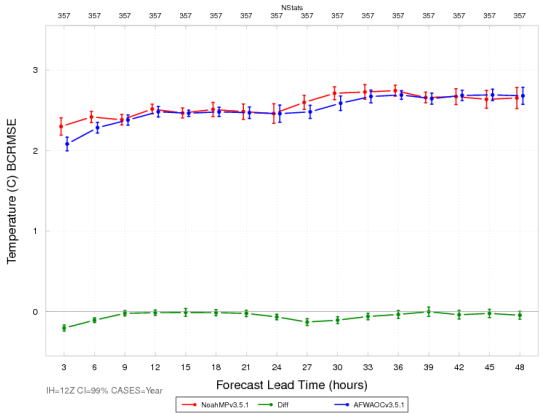


Figure 3. Map showing the locations of the CONUS-West, CONUS-East (top) and 14 verification regions (bottom). The outermost outline of the verification regions depicts the CONUS verification region.

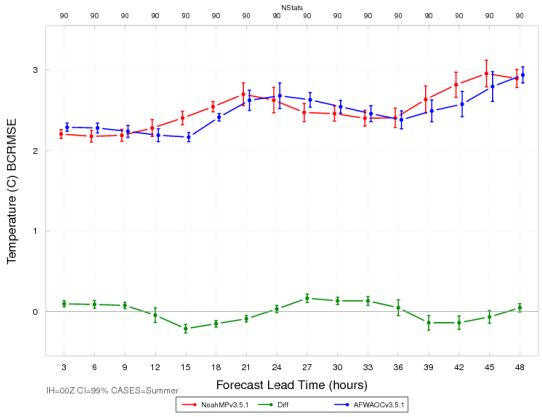
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=00 UTC



(d) Winter IH=00 UTC

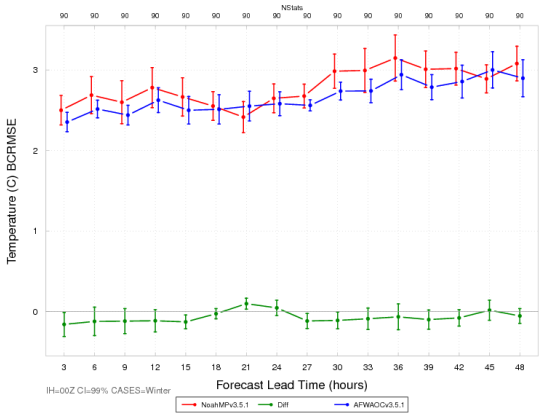
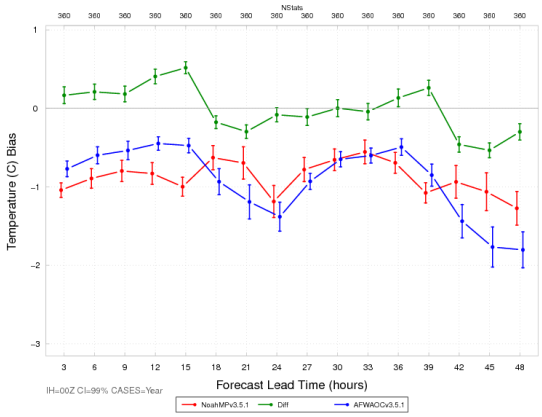
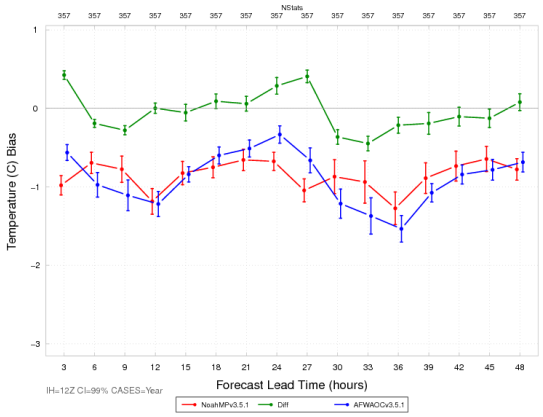


Figure 4. Time series plot of 2 m AGL temperature (°C) for median BCRMSE for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. AFWAOC is in blue, NoahMP in red, and the differences (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

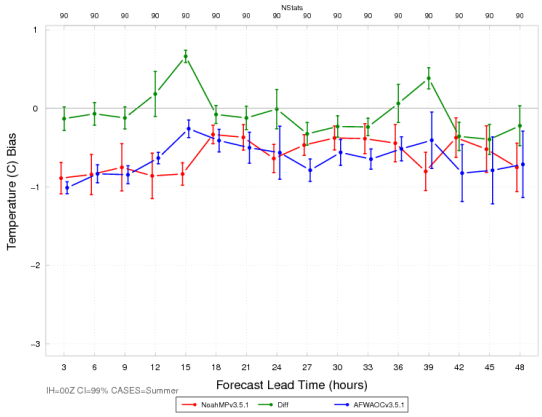
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=00 UTC



(d) Winter IH=00 UTC

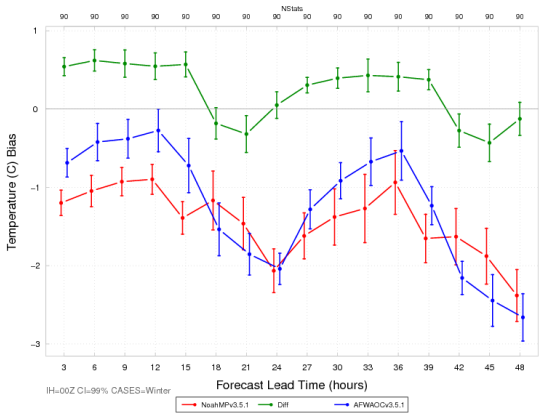
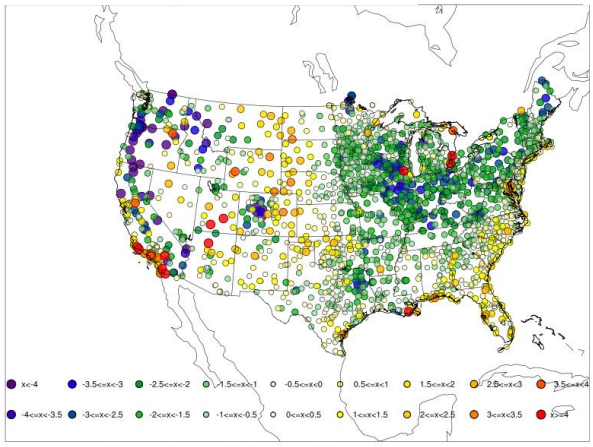


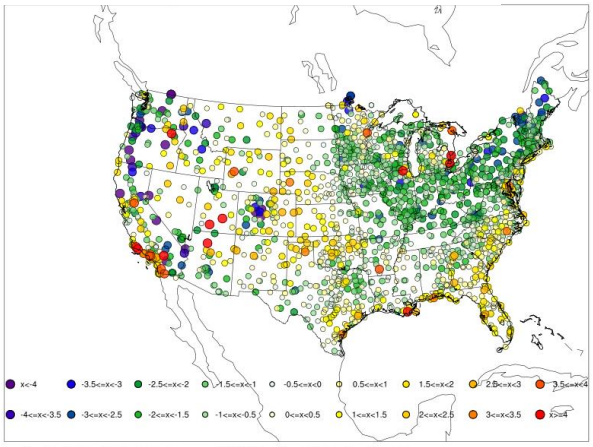
Figure 5. Time series plot of 2 m AGL temperature (°C) for median bias for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. AFWAOC is in blue, NoahMP in red, and the differences (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

(a) AFWAOC 00 UTC 30 h



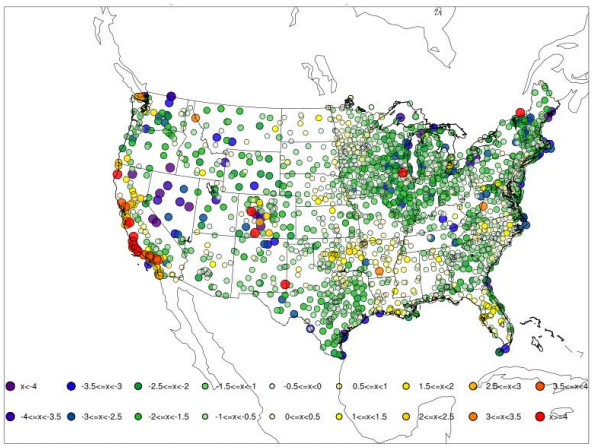
Config=AFWAOC_WRFv3.5.1 Season=SUMMER Init=00UTC Fcst Hr=30h

(b) NoahMP 00 UTC 30 h



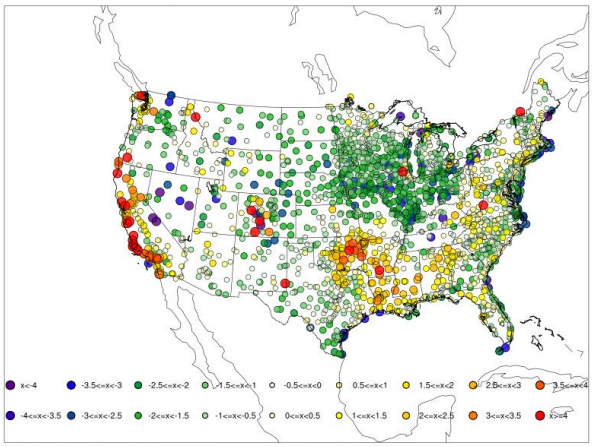
Config=NoahMP_WRFv3.5.1 Season=SUMMER Init=00UTC Fcst Hr=30h

(c) AFWAOC 00 UTC 42 h



Config=AFWAOC_WRFv3.5.1 Season=SUMMER Init=00UTC Fcst Hr=42h

(d) NoahMP 00 UTC 42 h



Config=NoahMP_WRFv3.5.1 Season=SUMMER Init=00UTC Fcst Hr=42h

Figure 6. Spatial plot of 2 m AGL temperature (°C) bias by observation station for all 00 UTC initializations in the summer aggregation for (a) AFWAOC at the 30 h forecast lead time (b) NoahMP at the 30 h forecast lead time (c) AFWAOC at the 42 h forecast lead time (d) NoahMP at the 42 h forecast lead time.

(a) AFWAOC 00 UTC 36 h

(b) NoahMP 00 UTC 36 h

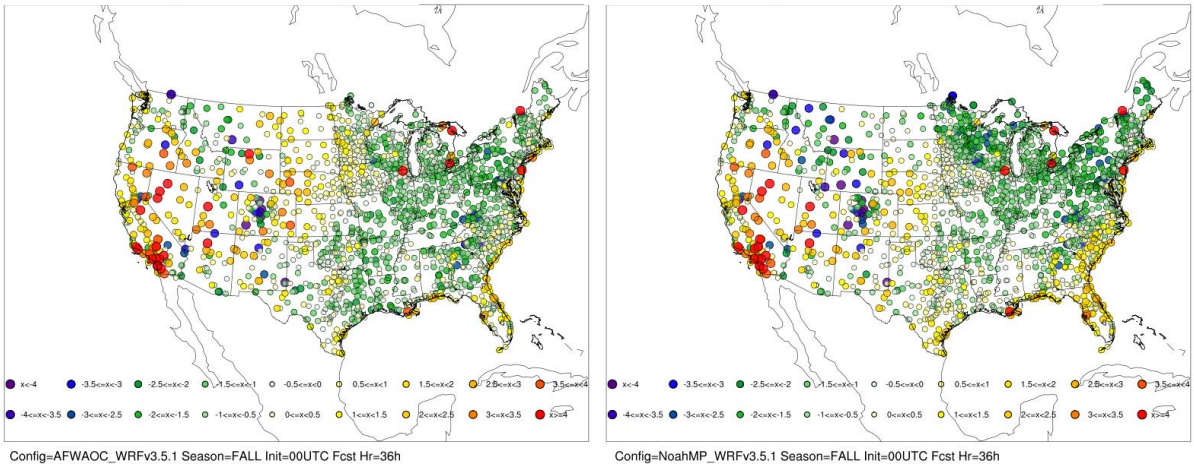
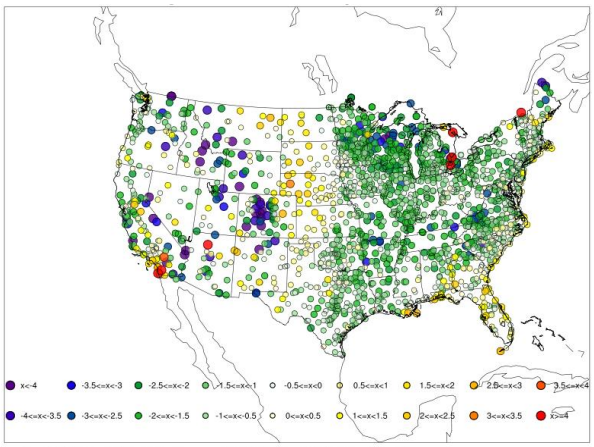


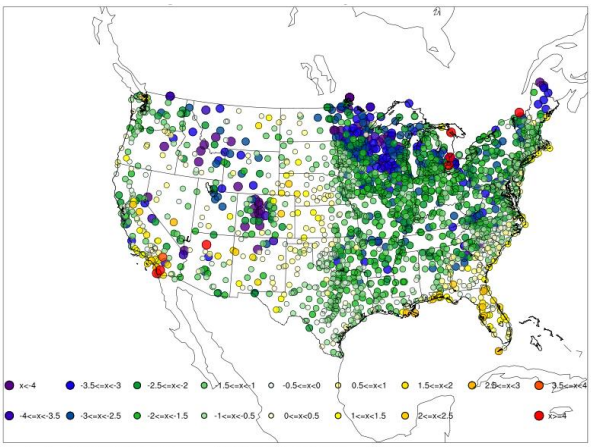
Figure 7. Spatial plot of 2 m AGL temperature (°C) bias by observation station for all 00 UTC initializations in the fall aggregation for (a) AFWAOC at the 36 h forecast lead time (b) NoahMP at the 36 h forecast lead time.

(a) AFWAOC 00 UTC 30 h



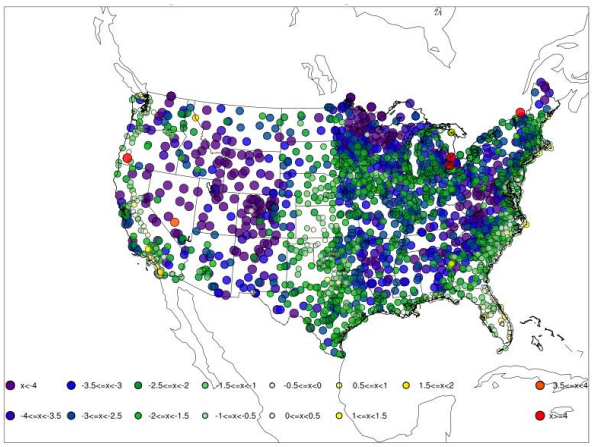
Config=AFWAOC_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=30h

(b) NoahMP 00 UTC 30 h



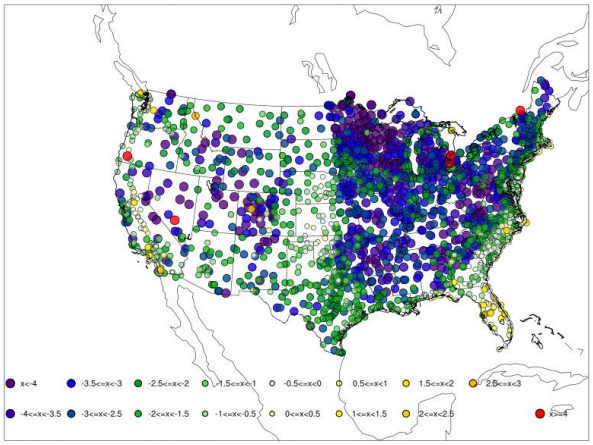
Config=NoahMP_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=30h

(c) AFWAOC 00 UTC 48 h



Config=AFWAOC_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=48h

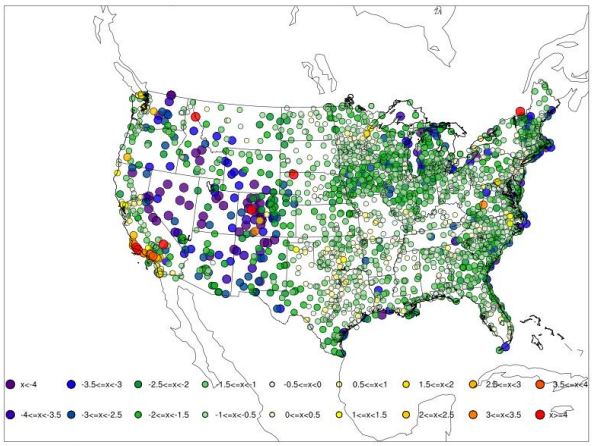
(d) NoahMP 00 UTC 48 h



Config=NoahMP_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=48h

Figure 8. Spatial plot of 2 m AGL temperature (°C) bias by observation station for all 00 UTC initializations in the winter aggregation for (a) AFWAOC at the 30 h forecast lead time (b) NoahMP at the 30 h forecast lead time (c) AFWAOC at the 48 h forecast lead time (d) NoahMP at the 48 h forecast lead time.

(a) AFWAOC 00 UTC 42 h



(b) NoahMP 00 UTC 42 h

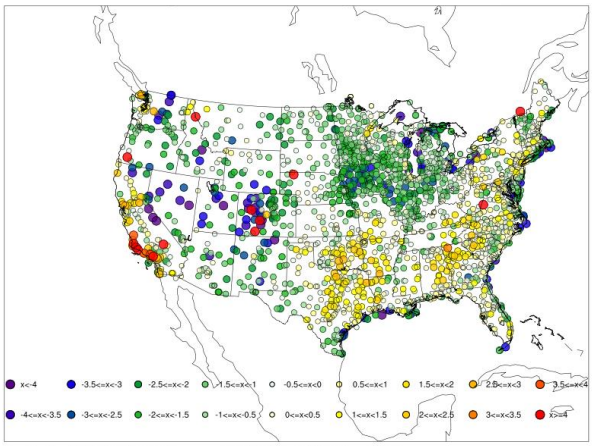
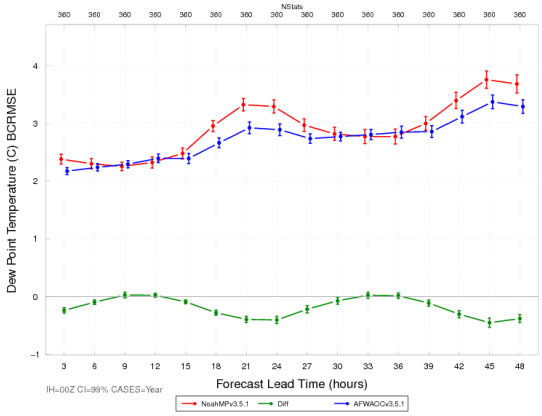
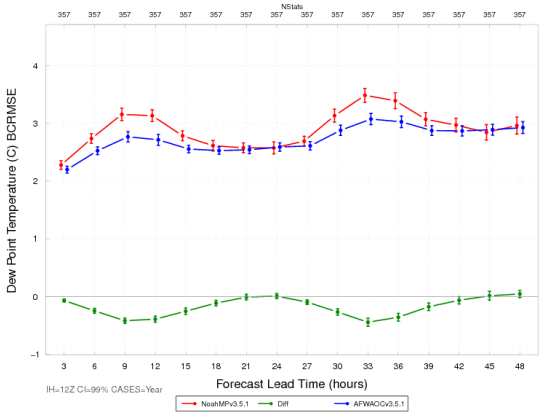


Figure 9. Spatial plot of 2 m AGL temperature (°C) bias by observation station for all 00 UTC initializations in the spring aggregation for (a) AFWAOC at the 42 h forecast lead time (b) NoahMP at the 42 h forecast lead time.

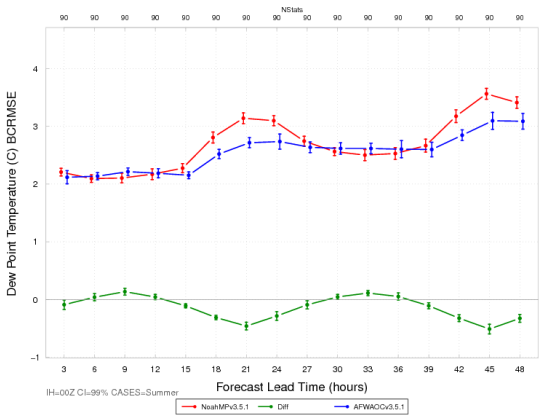
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=00 UTC



(d) Winter IH=00 UTC

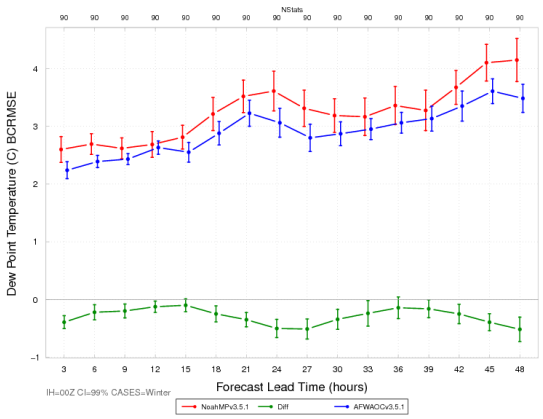
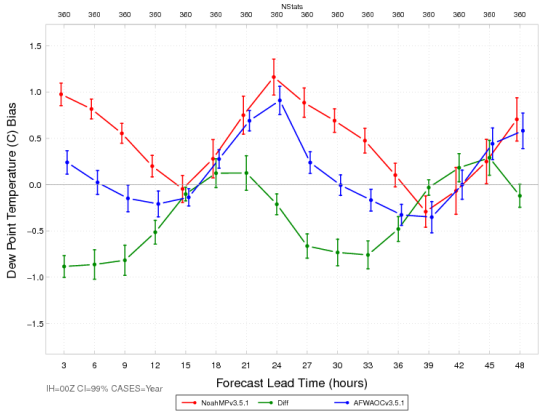
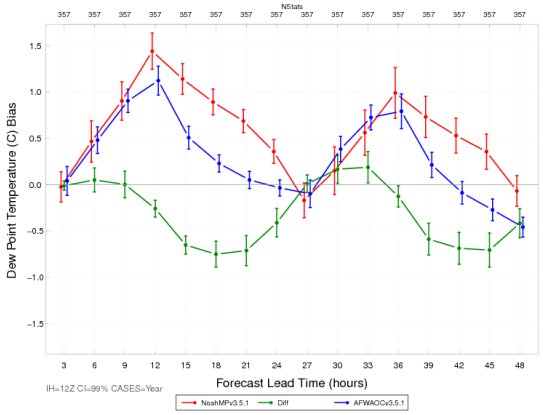


Figure 10. Time series plot of 2 m AGL dew point temperature (°C) for median BCRMSE for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. AFWAOC is in blue, NoahMP in red, and the differences (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

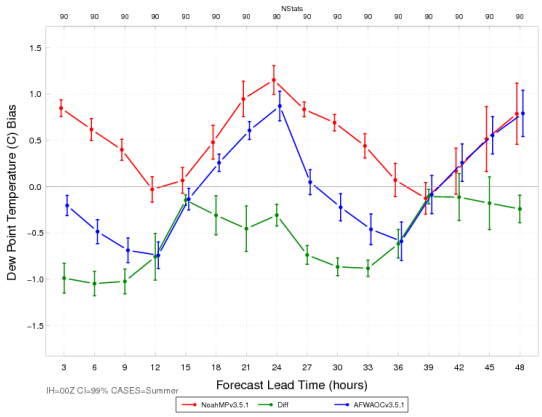
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=00 UTC



(d) Winter IH=00 UTC

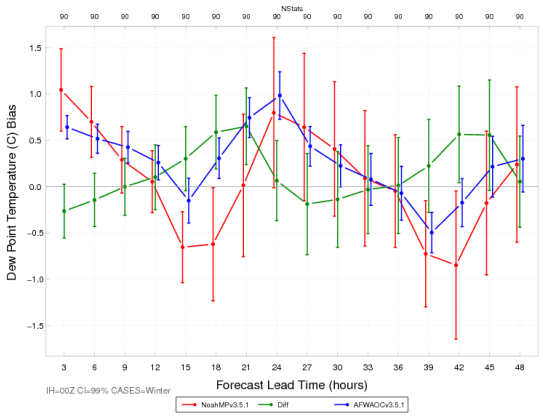
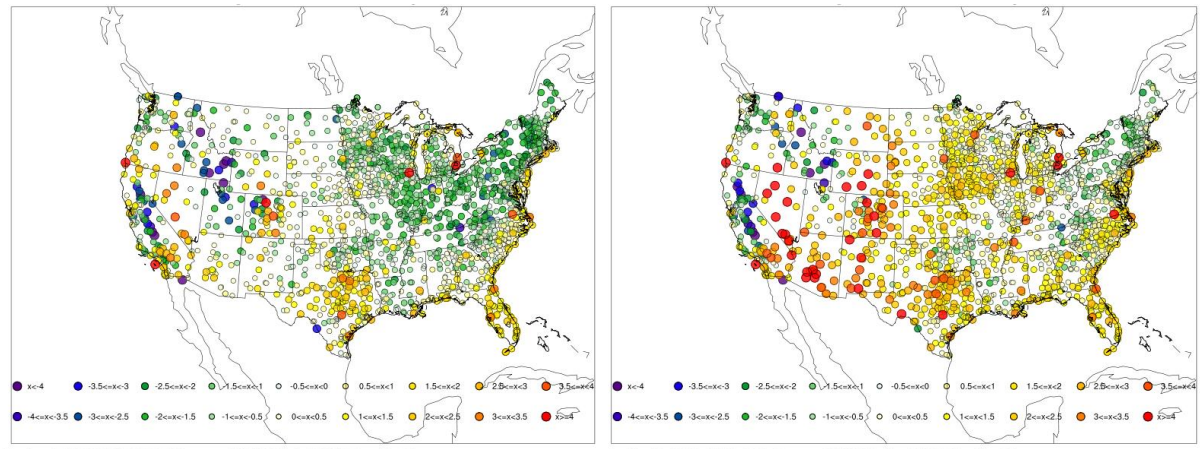


Figure 11. Time series plot of 2 m AGL dew point temperature (°C) for median bias for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter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

(a) AFWAOC 00 UTC 30 h

(b) NoahMP 00 UTC 30 h



(c) AFWAOC 00 UTC 42 h

(d) NoahMP 00 UTC 42 h

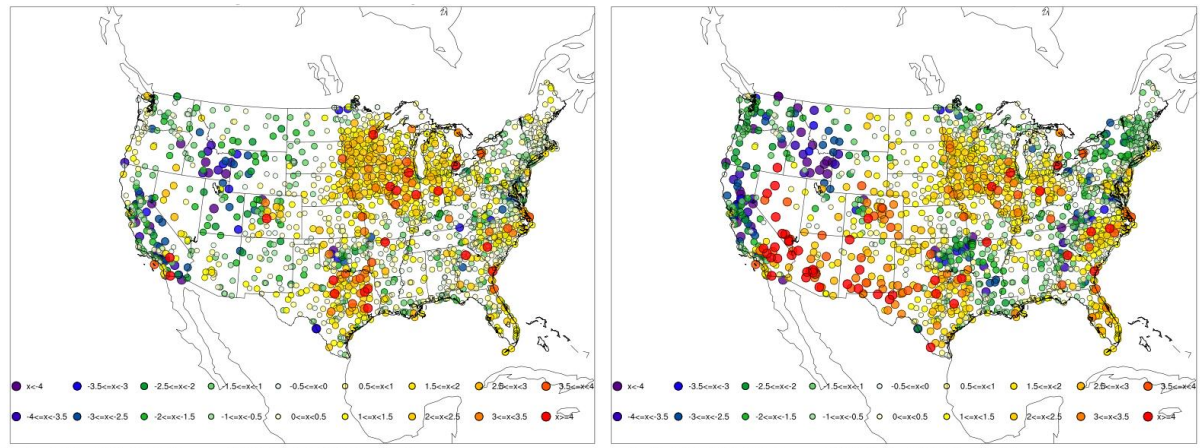
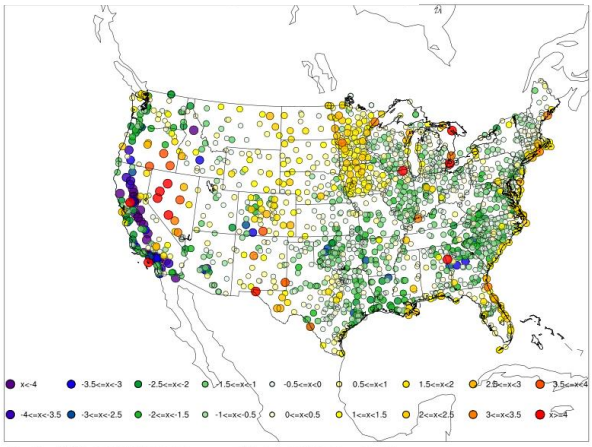


Figure 12. Spatial plot of 2 m AGL dew point temperature (°C) bias by observation station for all 00 UTC initializations in the summer aggregation for (a) AFWAOC at the 30 h forecast lead time (b) NoahMP at the 30 h forecast lead time (c) AFWAOC at the 42 h forecast lead time (d) NoahMP at the 42 h forecast lead time.

(a) AFWAOC 00 UTC 36 h



(b) NoahMP 00 UTC 36 h

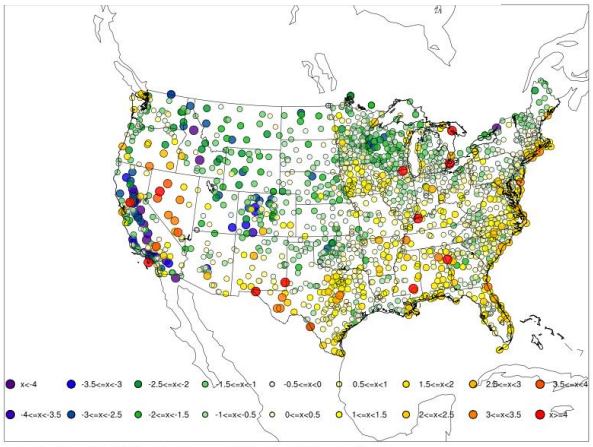
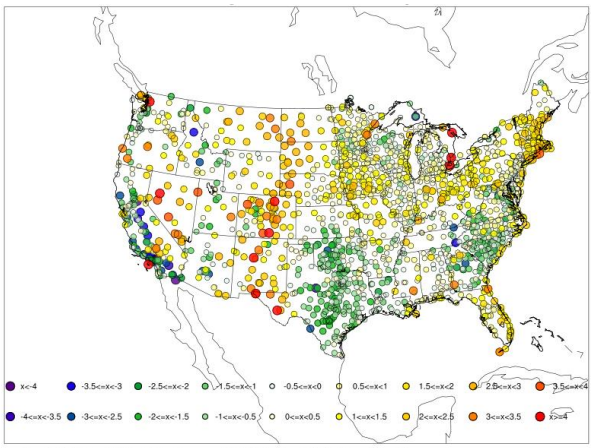


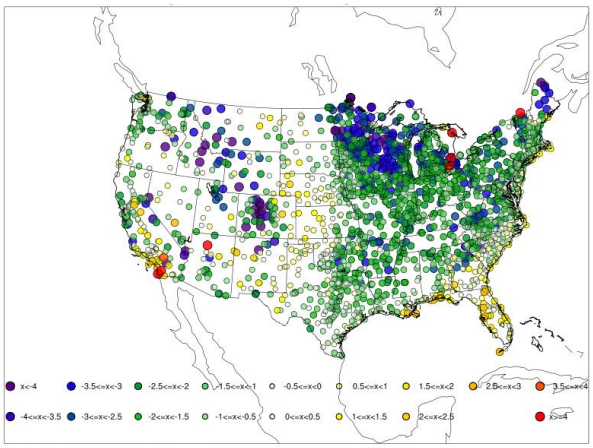
Figure 13. Spatial plot of 2 m AGL dew point temperature (°C) bias by observation station for all 00 UTC initializations in the fall aggregation for (a) AFWAOC at the 36 h forecast lead time (b) NoahMP at the 36 h forecast lead time.

(a) AFWAOC 00 UTC 30 h



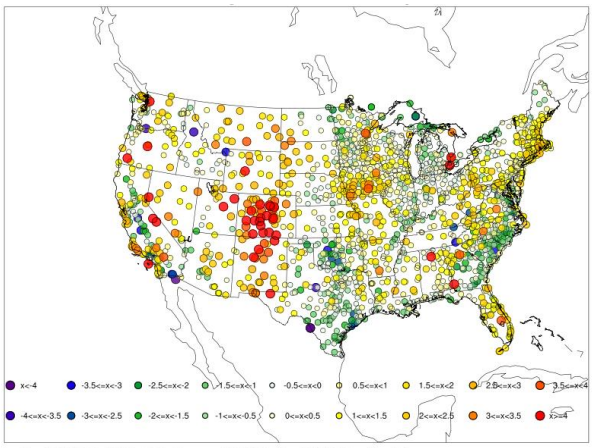
Config=AFWAOC_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=30h

(b) NoahMP 00 UTC 30 h



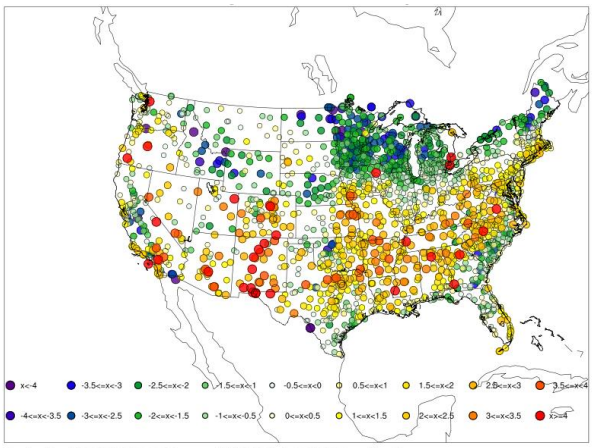
Config=NoahMP_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=30h

(c) AFWAOC 00 UTC 48 h



Config=AFWAOC_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=48h

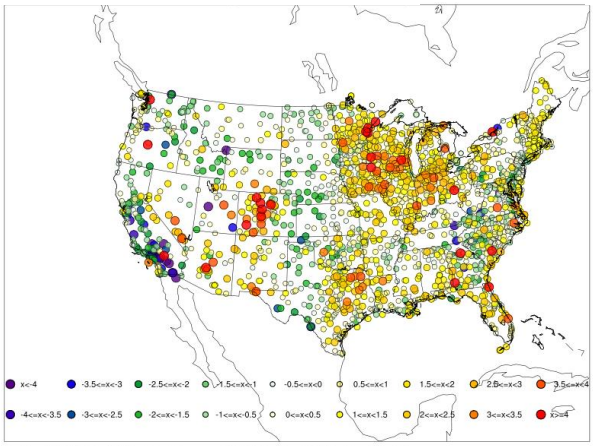
(d) NoahMP 00 UTC 48 h



Config=NoahMP_WRFv3.5.1 Season=WINTER Init=00UTC Fcst Hr=48h

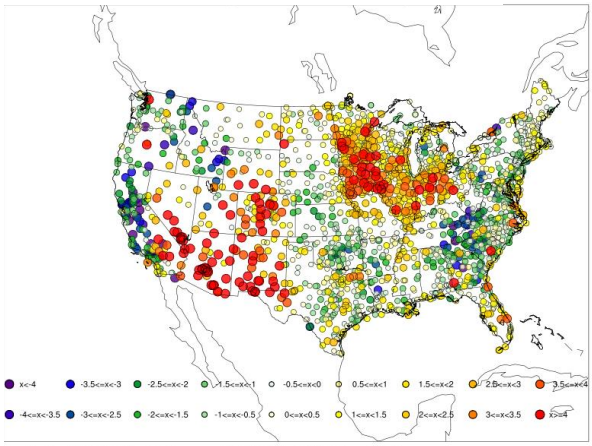
Figure 14. Spatial plot of 2 m AGL dew point temperature (°C) bias by observation station for all 00 UTC initializations in the winter aggregation for (a) AFWAOC at the 30 h forecast lead time (b) NoahMP at the 30 h forecast lead time (c) AFWAOC at the 48 h forecast lead time (d) NoahMP at the 48 h forecast lead time.

(a) AFWAOC 00 UTC 42 h



Config=AFWAOC_WRFv3.5.1 Season=SPRING Init=00UTC Fcst Hr=42h

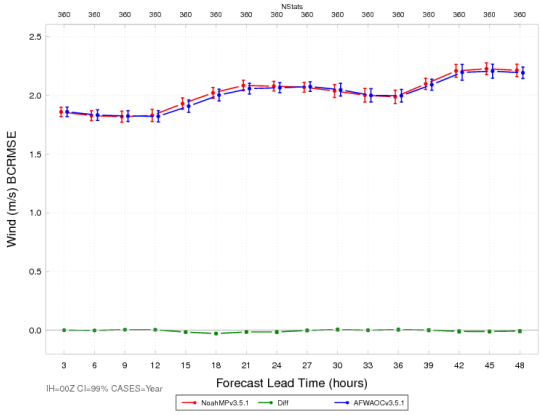
(b) NoahMP 00 UTC 42 h



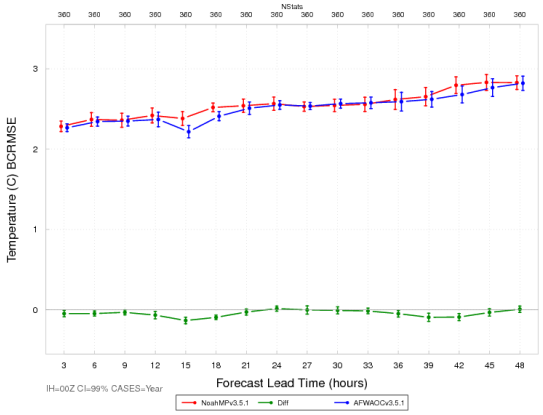
Config=NoahMP_WRFv3.5.1 Season=SPRING Init=00UTC Fcst Hr=42h

Figure 15. Spatial plot of 2 m AGL dew point temperature (°C) bias by observation station for all 00 UTC initializations in the spring aggregation for (a) AFWAOC at the 42 h forecast lead time (b) NoahMP at the 42 h forecast lead time.

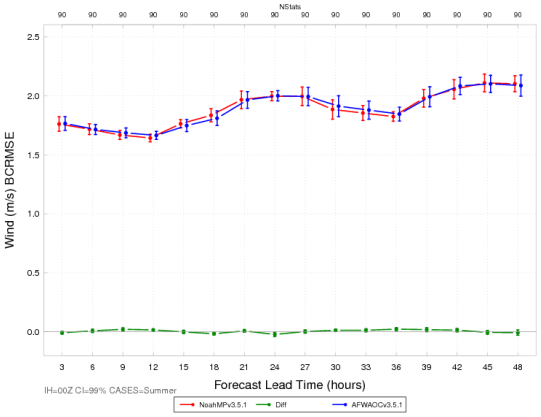
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=00 UTC



(d) Winter IH=00 UTC

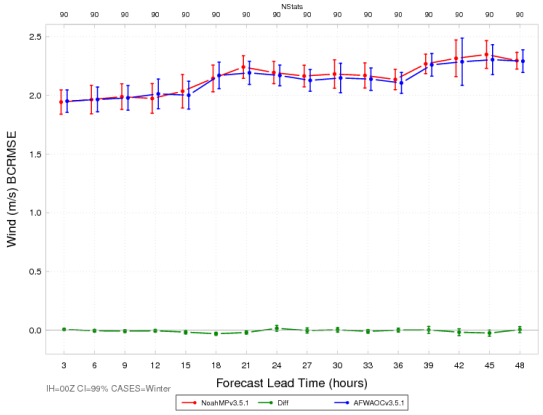


Figure 16. Time series plot of 2 m AGL wind speed (m s^{-1}) for median BCRMSE for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. AFWAOC is in blue, NoahMP in red, and the differences (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

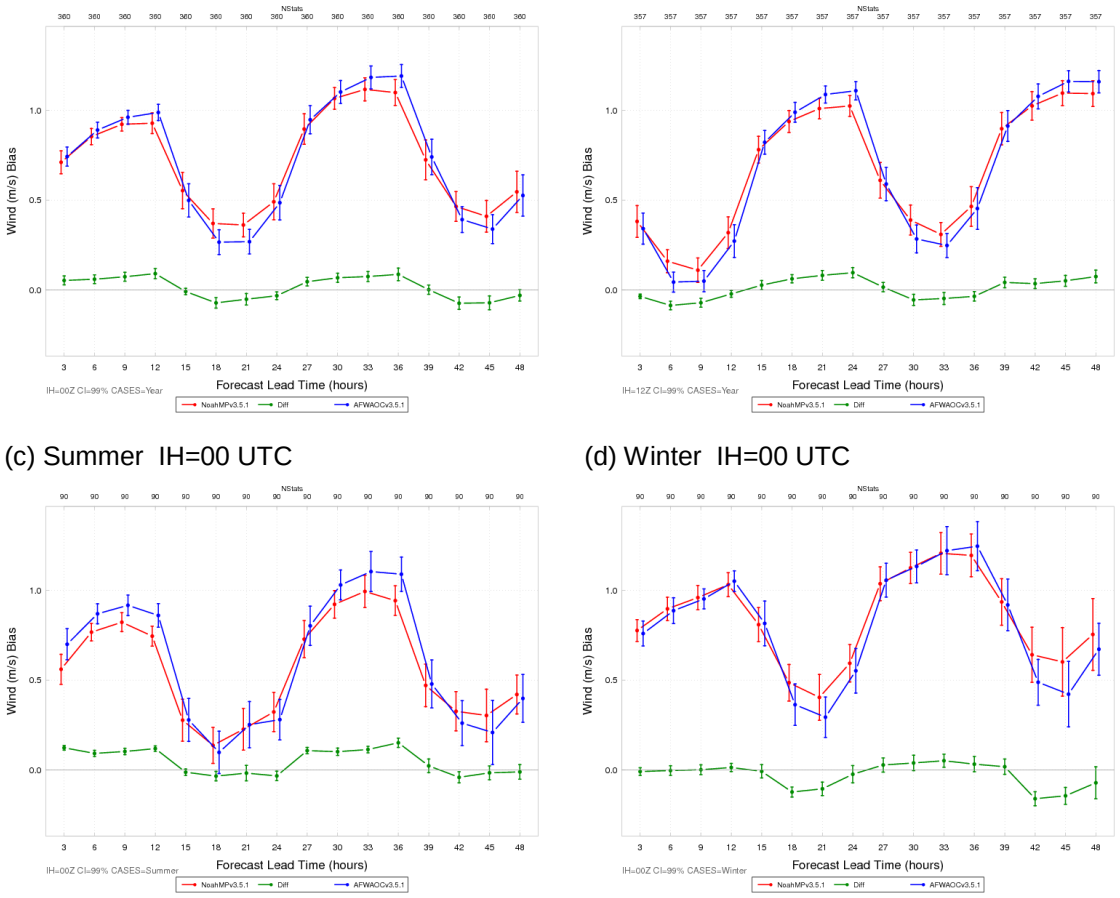
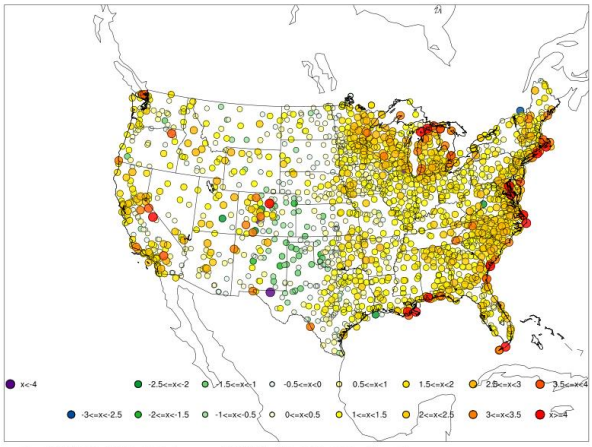


Figure 17. Time series plot of 2 m AGL wind speed (m s^{-1}) for median bias for the full integration domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. AFWAOC is in blue, NoahMP in red, and the differences (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

(a) AFWAOC 00 UTC 30 h



(b) AFWAOC 00 UTC 42 h

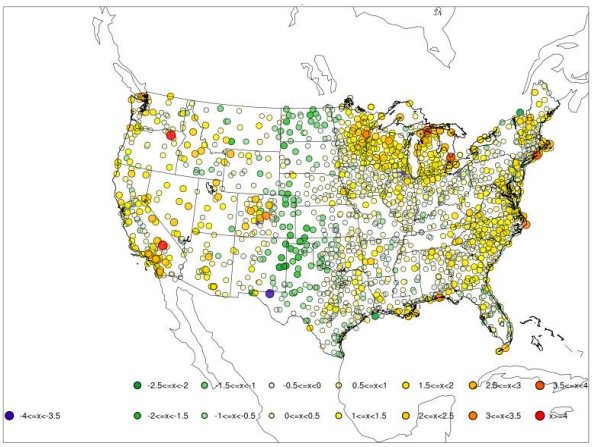


Figure 18. Spatial plot of 10 m AGL wind speed (m s^{-1}) bias by observation station for all 00 UTC initializations in the annual aggregation for (a) AFWAOC at the 30 h forecast lead time (b) AFWAOC at the 42 h forecast lead time.

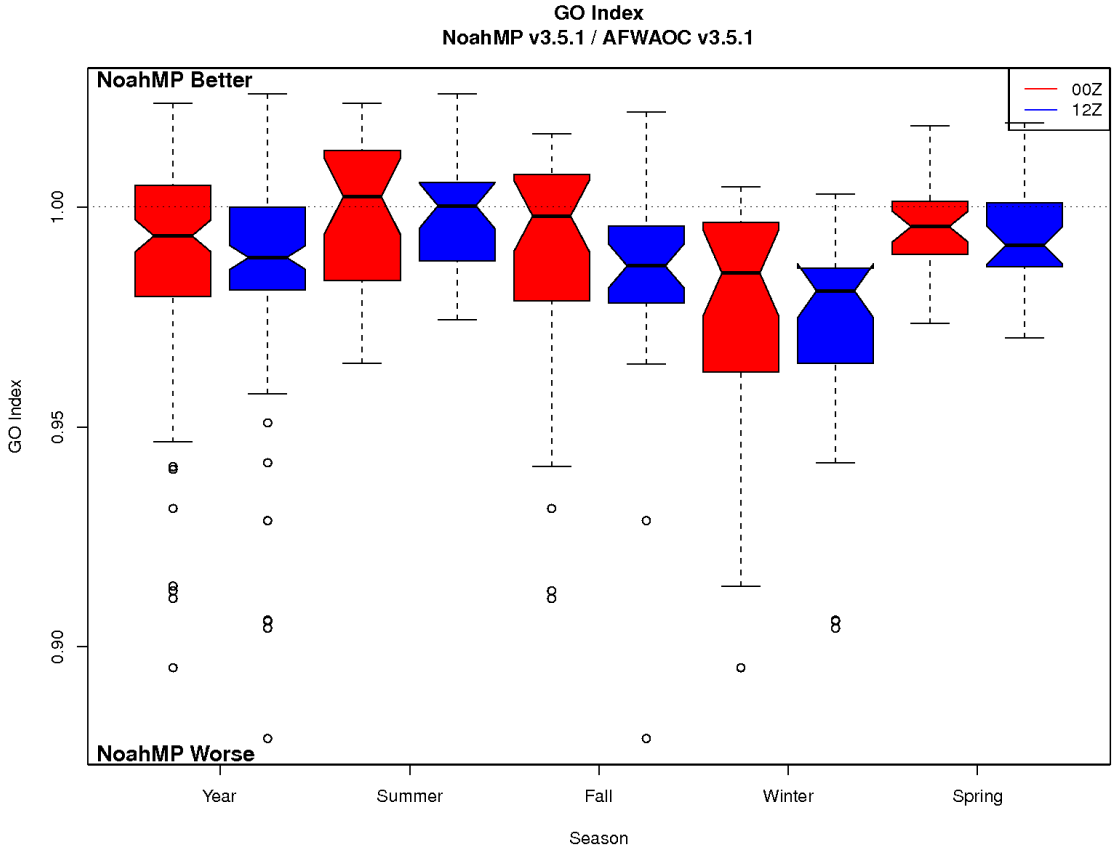


Figure 19. Boxplot of GO Index values aggregated across the entire year of cases and for all seasons, stratified by initialization time where 00 UTC is in red and 12 UTC is in blue. The median value is the thick black line located at the vertex of the notches, the notches around the median is an approximation of the 95% confidence about the median, the whiskers, denoted by the black, dashed lines, denote the largest values that are not outliers, and the circles represent the outliers.

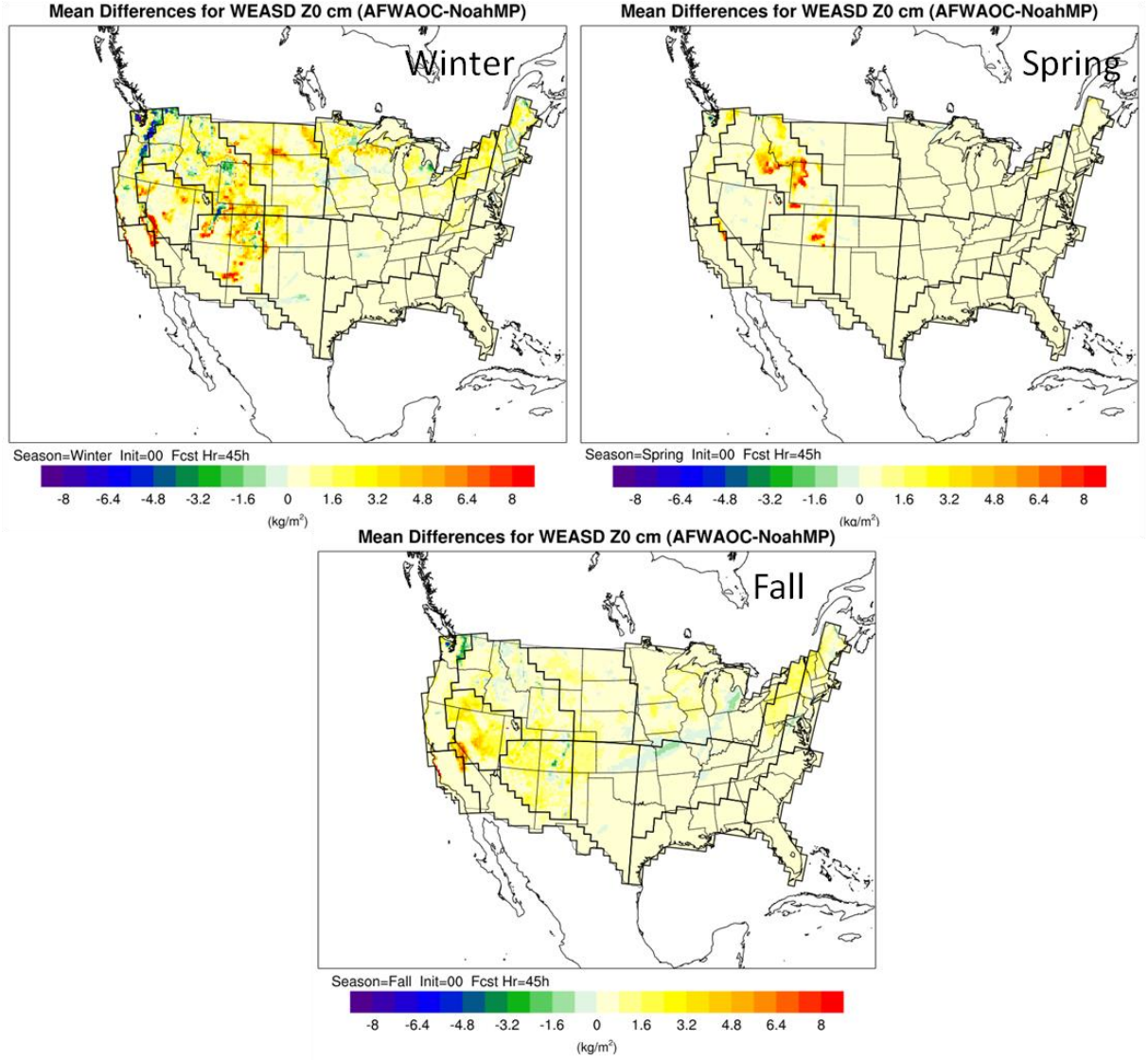


Figure 20. Configuration differences (AFWAOC-NoahMP) of snow equivalent water (SWE) aggregated across the season of (a) winter, (b) spring, and (c) fall. The plots are for 00 initialization and 45-h lead time (21 UTC valid time).

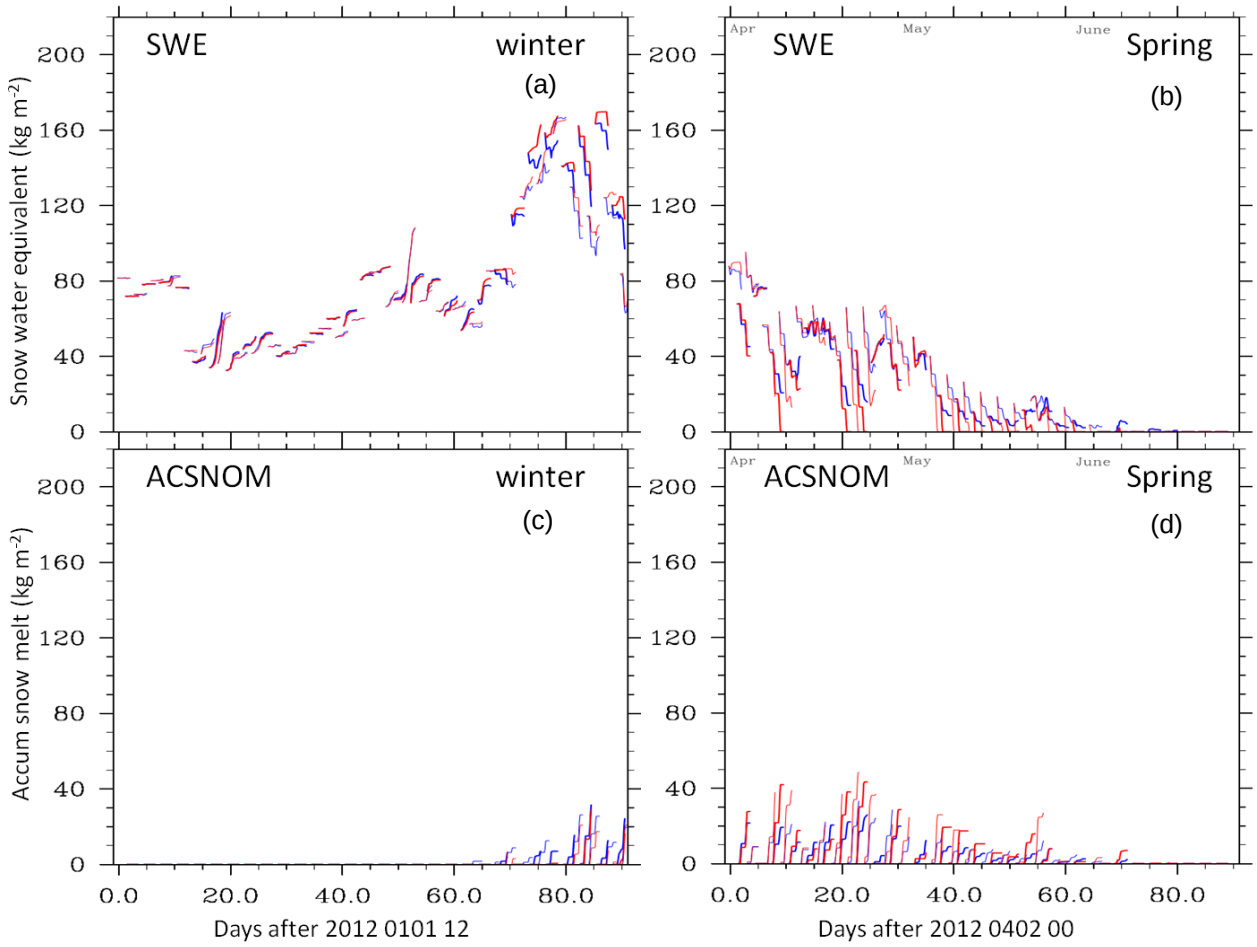


Figure 21. Collective time series of ground snow equivalent water (SWE; top row) and accumulated melted snow (ACSNOM; bottom row) at Site 1 in the winter and spring. Each individual curve is for one continuous simulation including values at the cold-start, warm-start and 3 - 48 h forecasts. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

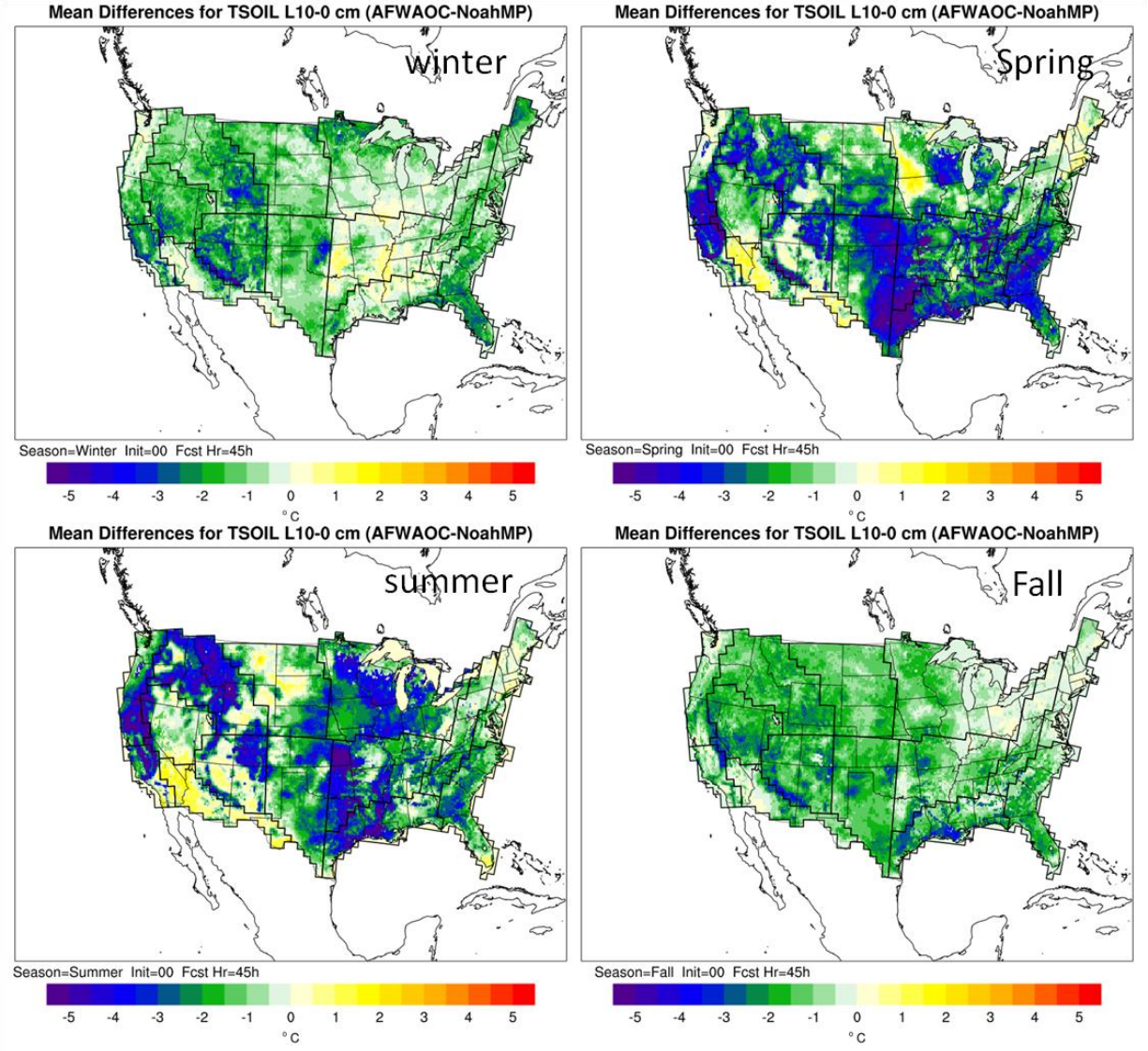


Figure 22. Configuration differences (AFWAOC-NoahMP) of the top-layer (0-10 cm) soil temperature (SOILT) aggregated across the season of (a) winter, (b) spring, (c) summer, and (d) fall. The plots are for 00 initialization and 45-h lead time (21 UTC valid time).

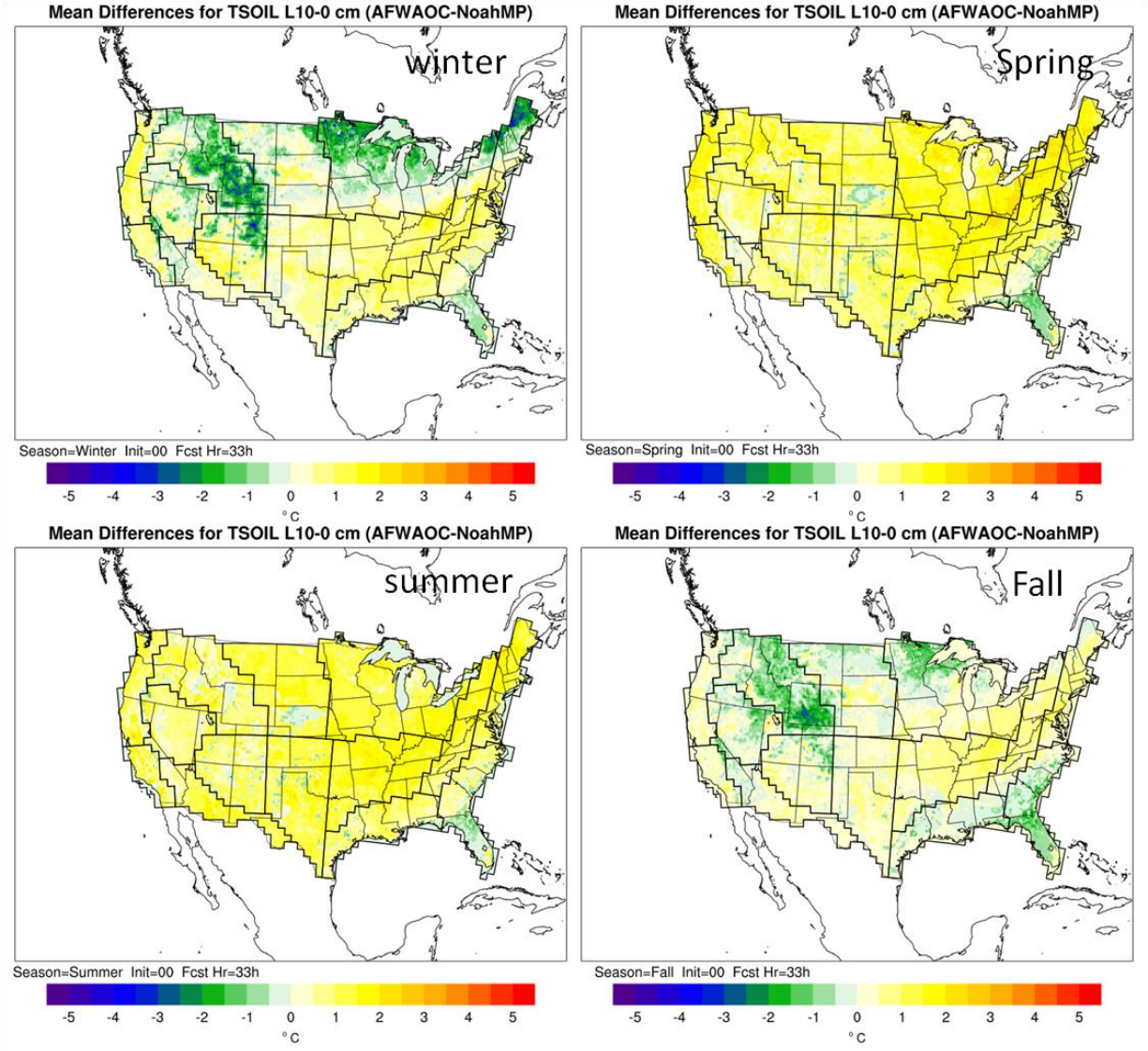


Figure 23. The same as Figure 22 but for 33-h lead time (09 UTC valid time).

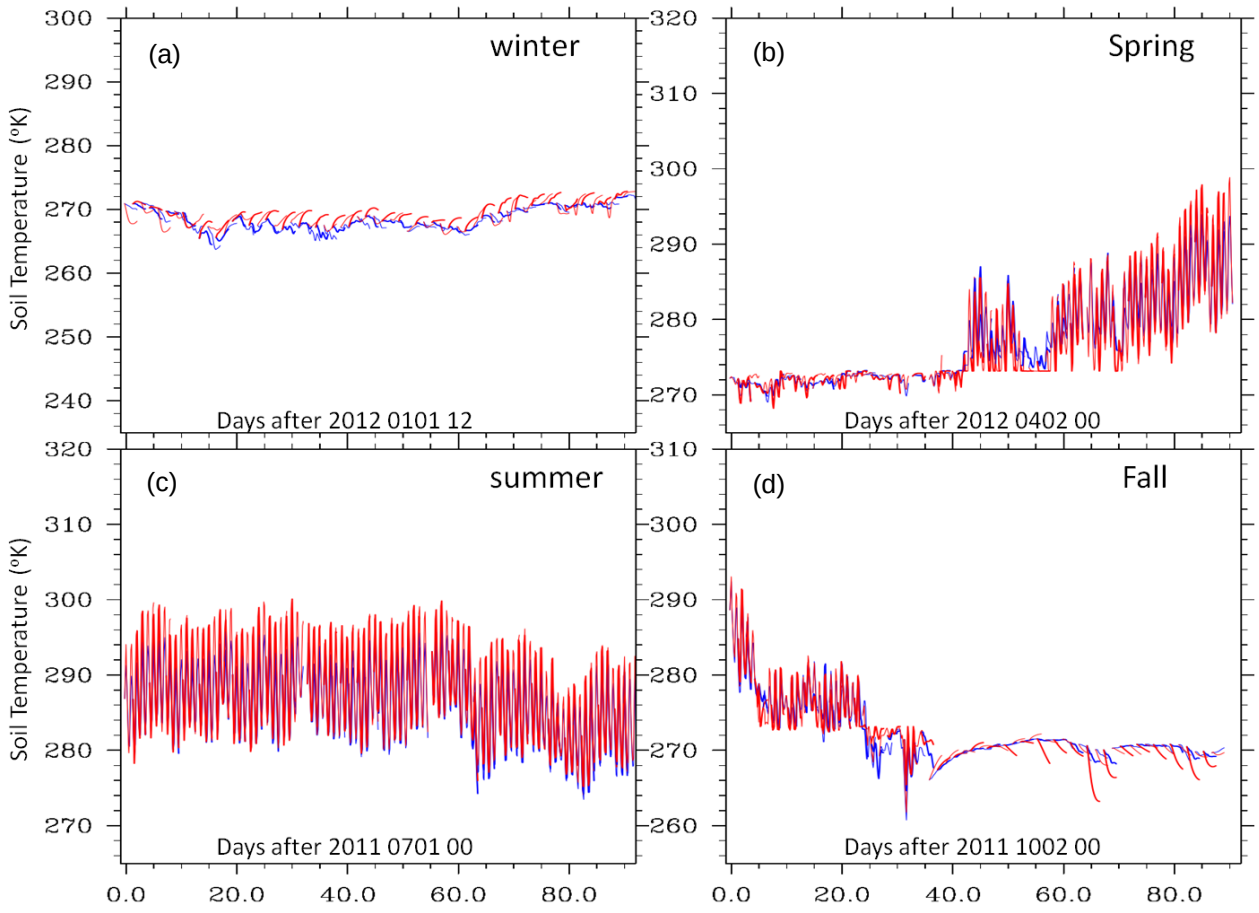


Figure 24. Collective time series of the top-layer (0-10 cm) soil temperature (SOILT) at Site1 for the season of (a) winter, (b) spring, (c) summer, and (d) fall. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

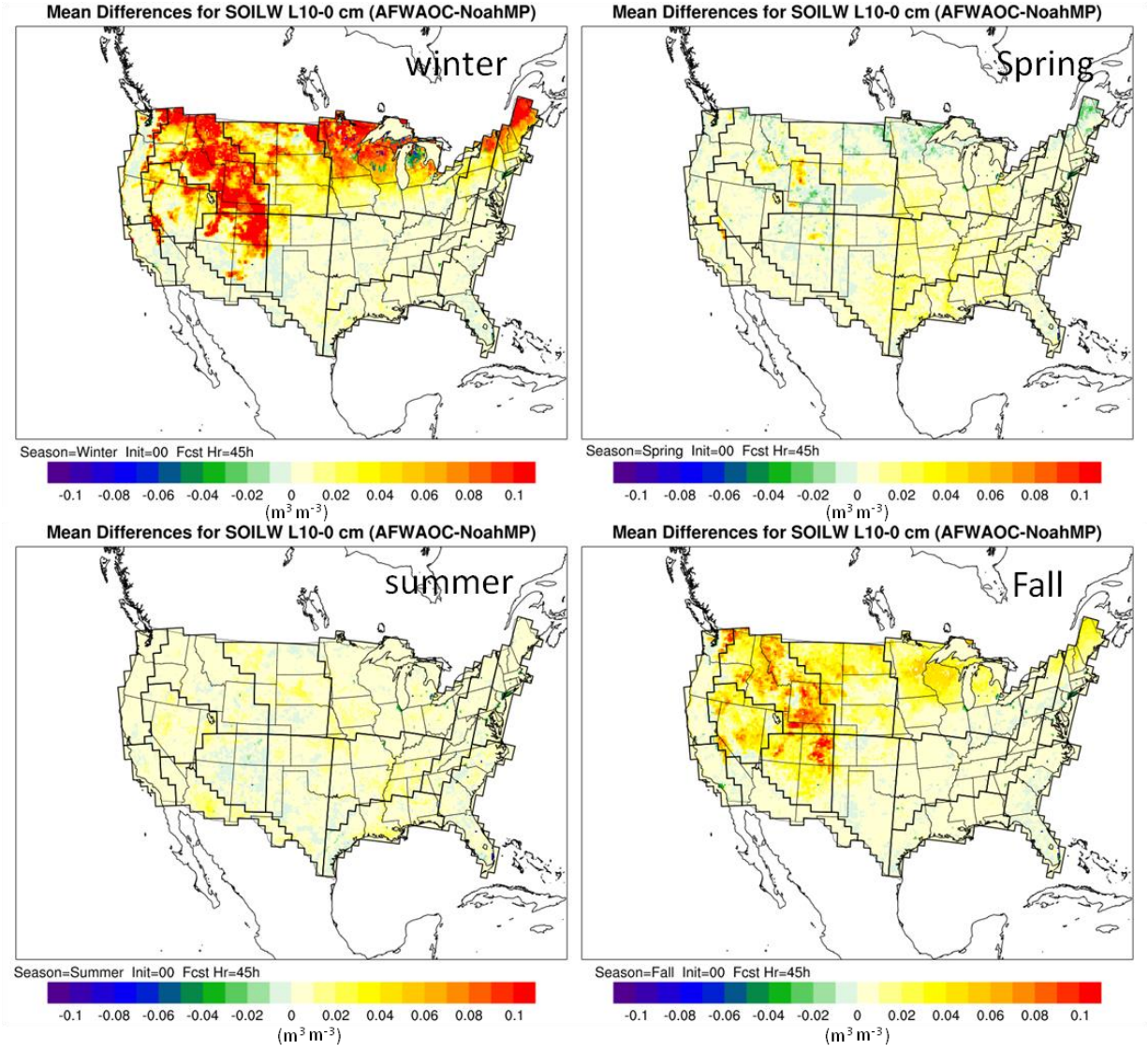


Figure 25. Configuration differences (AFWAOC-NoahMP) of the top-layer (0-10 cm) soil moisture (SOILM) aggregated across the season of (a) winter, (b) spring, (c) summer, and (d) fall. The plots are for 00 initialization and 45-h lead time (21 UTC valid time).

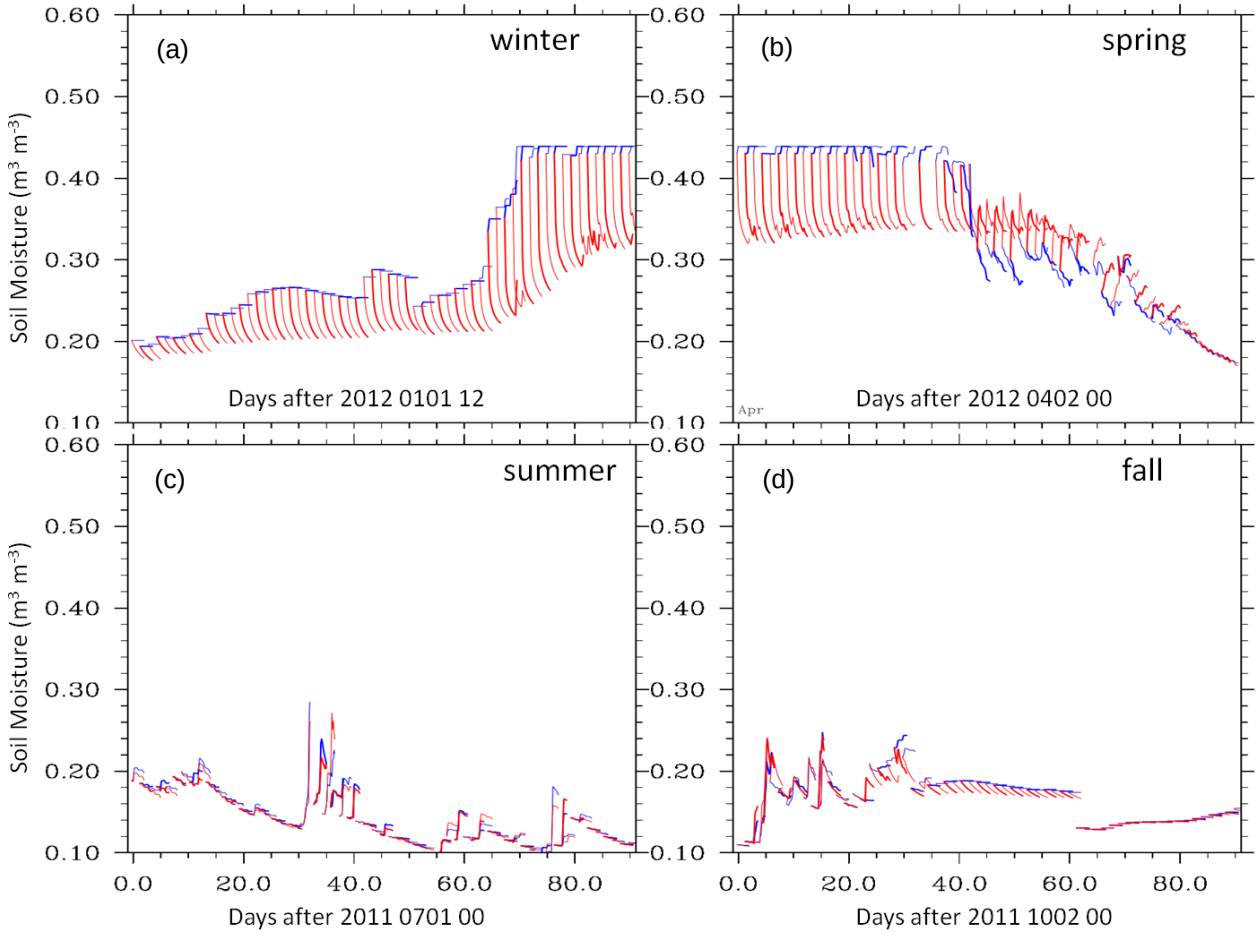


Figure 26. Collective time series of the top-layer (0-10 cm) soil moisture (SOILM) at Site1 for the season of (a) winter, (b) spring, (c) summer, and (d) fall. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

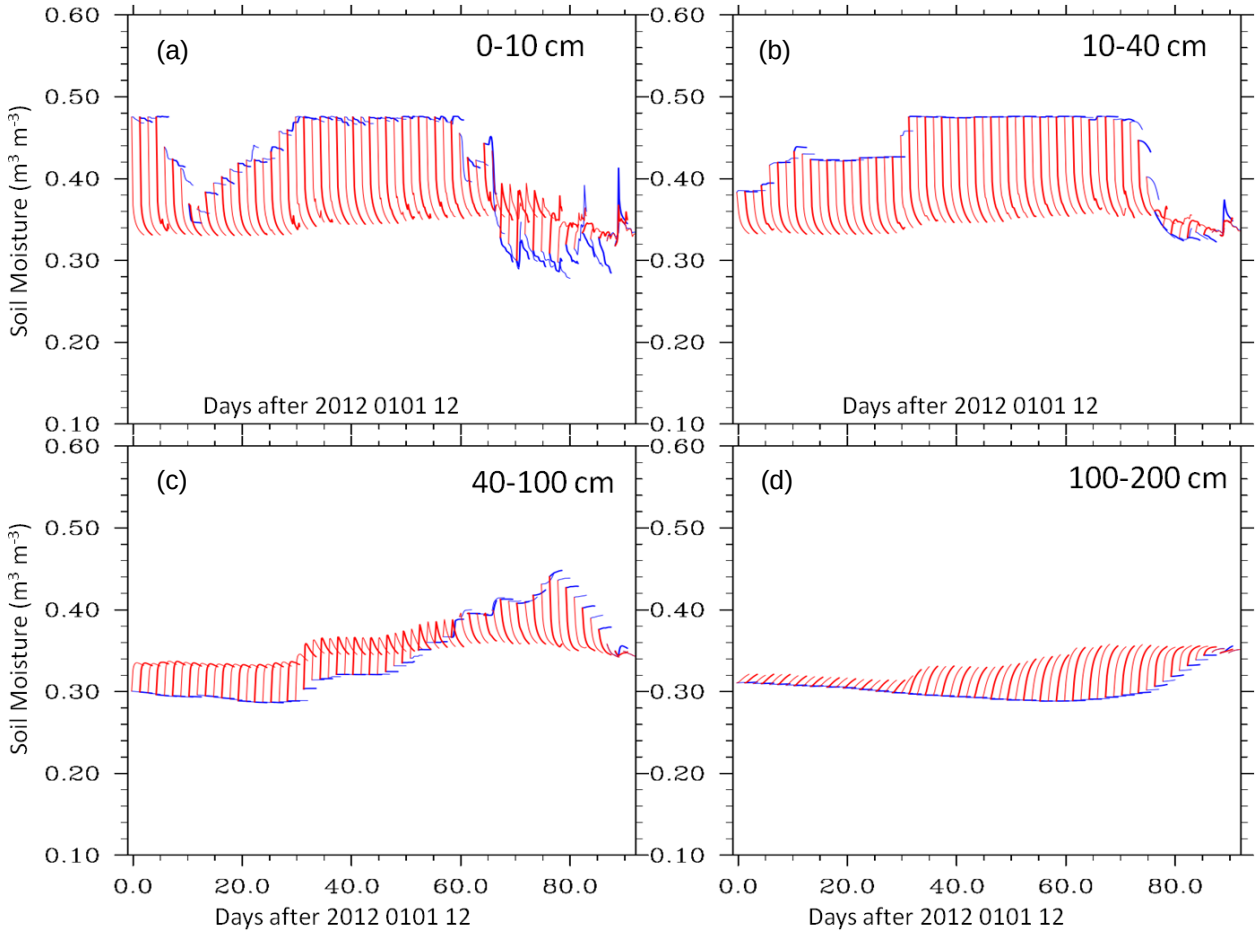


Figure 27. Collective time series of the soil moisture (SOILM) at Site1 in the winter at the soil layer of (a) 0-10 cm, (b) 10-40 cm, (c) 40-100 cm, and (d) 100-200 cm. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

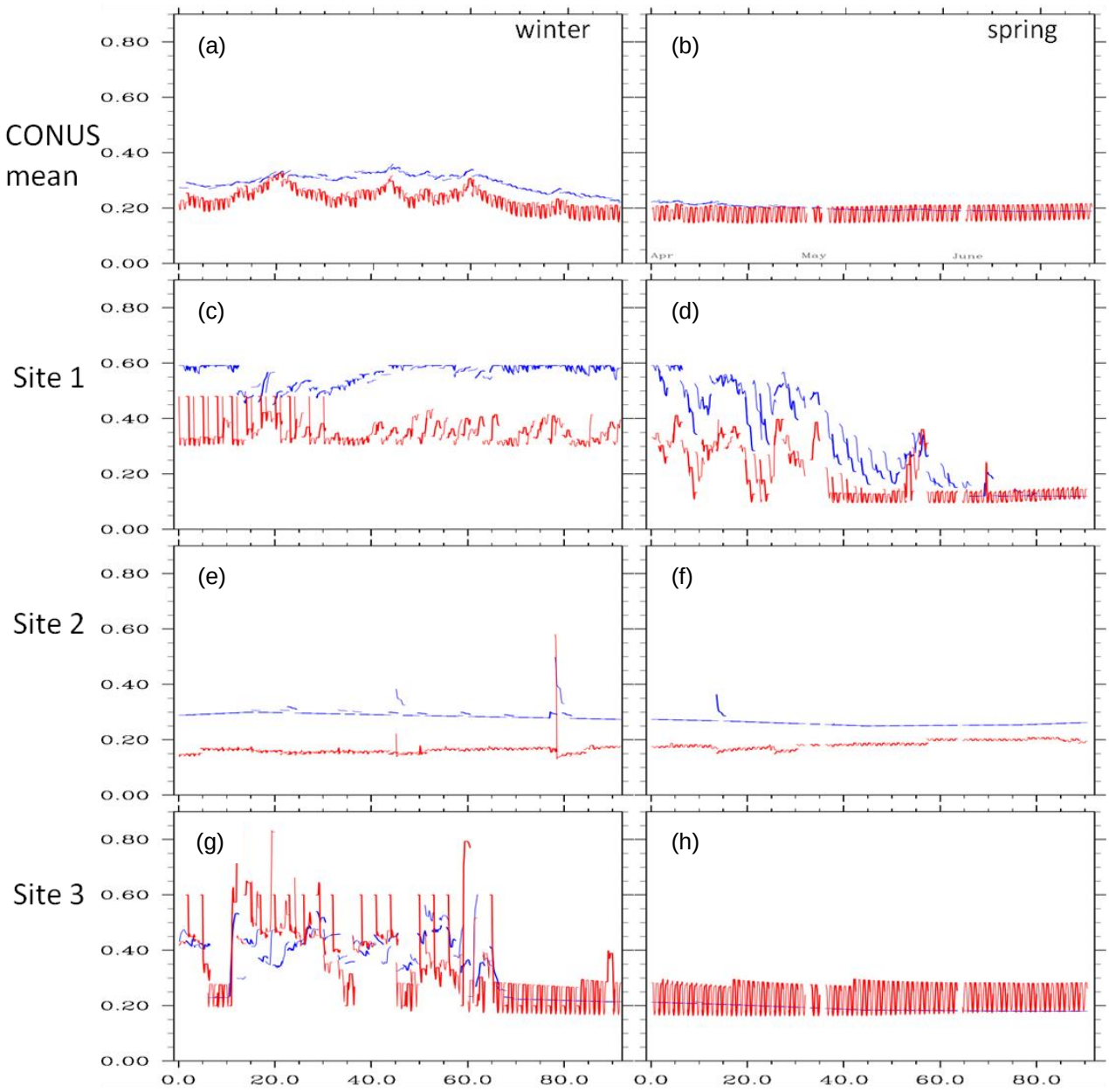


Figure 28. Collective time series of the surface albedo in the winter (left panels) and spring (right panels). Each individual curve is for one continuous simulation including 03 – 48 h forecasts. From top to bottom panels are for (a-b) CONUS mean, (c-d) Site 1, (e-f) Site 2, and (g-h) Site 3, respectively. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

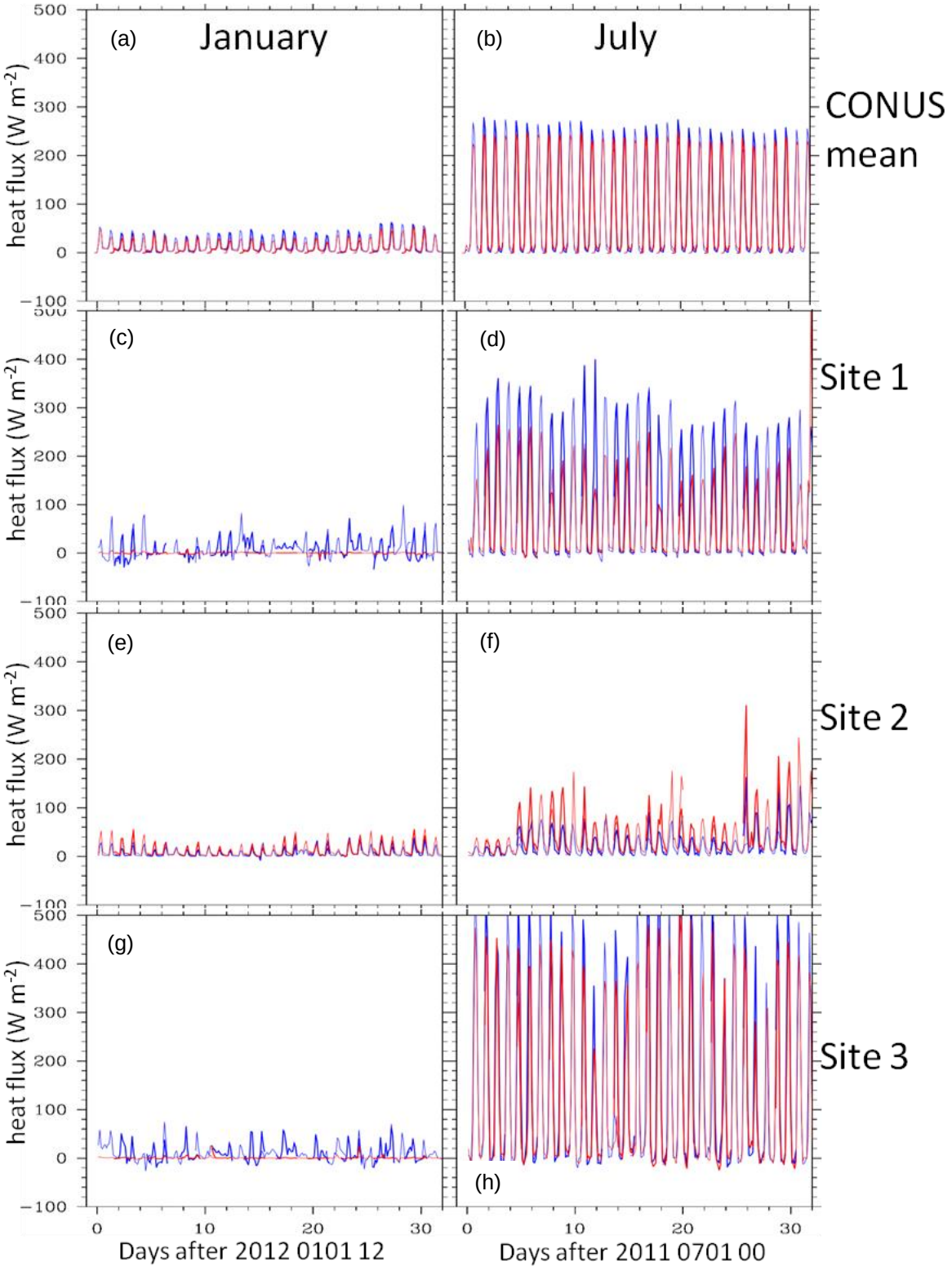


Figure 29. Collective time series of surface latent heat flux (LH) for the month of January (left panels) and July (right panels). Each individual curve is for one continuous simulation including 03- 48-h forecasts. From top to bottom panels are for (a-b) CONUS mean, (c-d) Site 1, (e-f) Site 2, and (g-h) Site 3, respectively. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

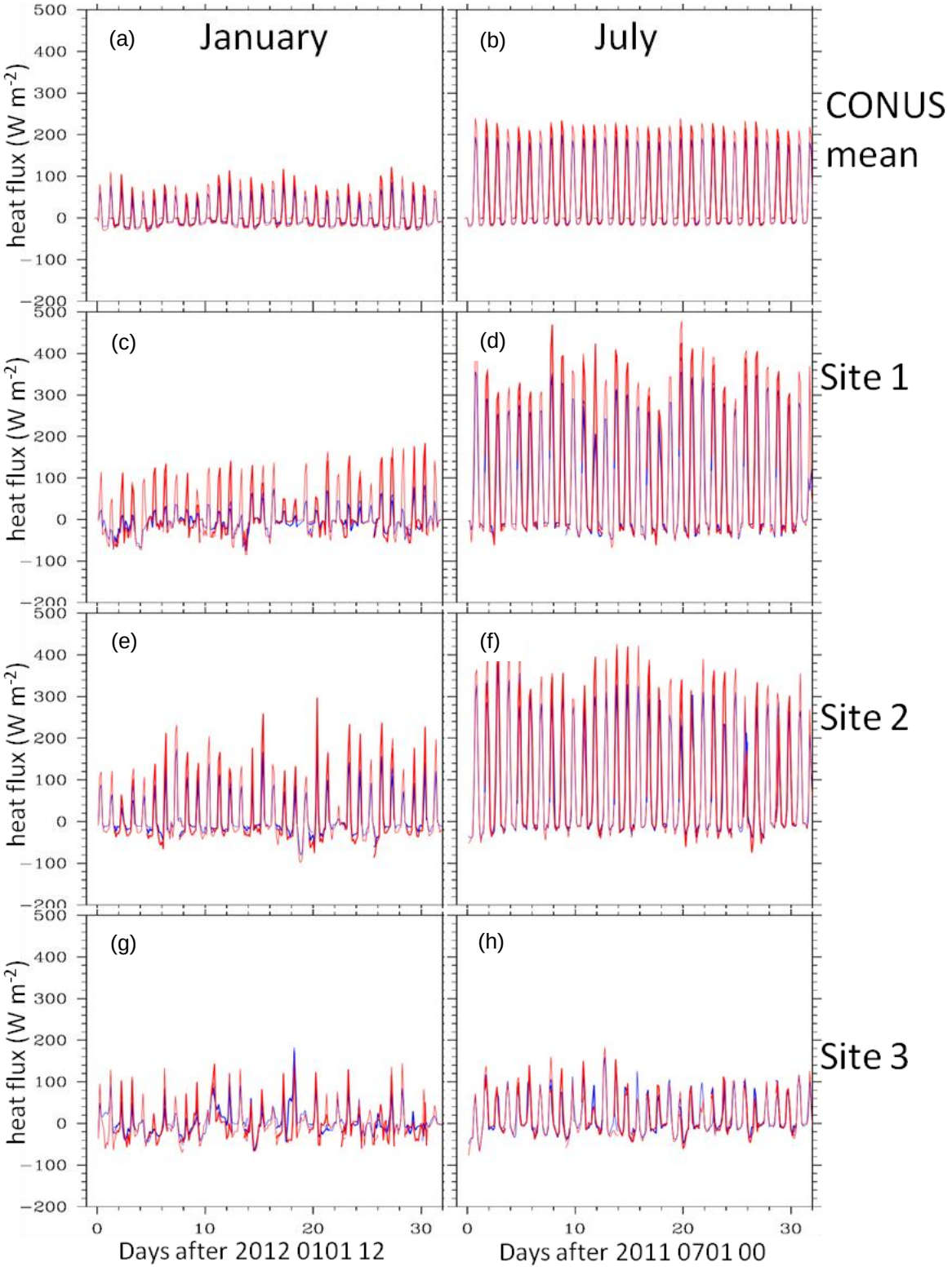


Figure 30. Collective time series of surface sensible heat flux (HFX) for the month of January (left panels) and July (right panels). Each individual curve is for one continuous simulation including 03 - 48h forecasts. From top to bottom panels are for (a-b) CONUS mean, (c-d) Site 1, (e-f) Site 2, and (g-h) Site 3, respectively. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

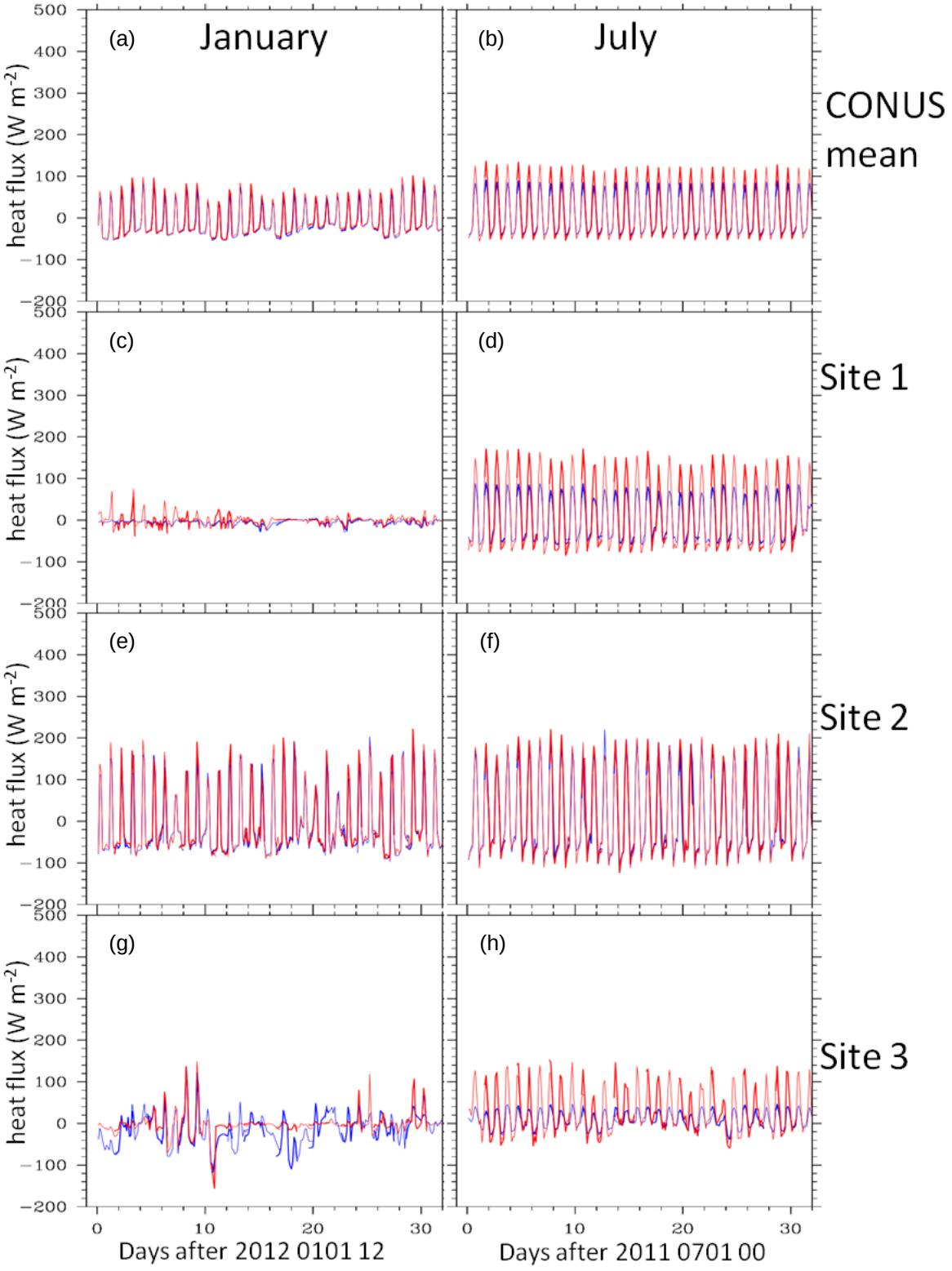


Figure 31. Collective time series of ground heat flux (GRDFLX) for the month of January (left panels) and July (right panels). Each individual curve is for one continuous simulation including 03 – 48 h forecasts. From top to bottom panels are for (a-b) CONUS mean, (c-d) Site 1, (e-f) Site 2, and (g-h) Site 3, respectively. Blue is for the AFWAOC configuration and red is for the NoahMP configuration.

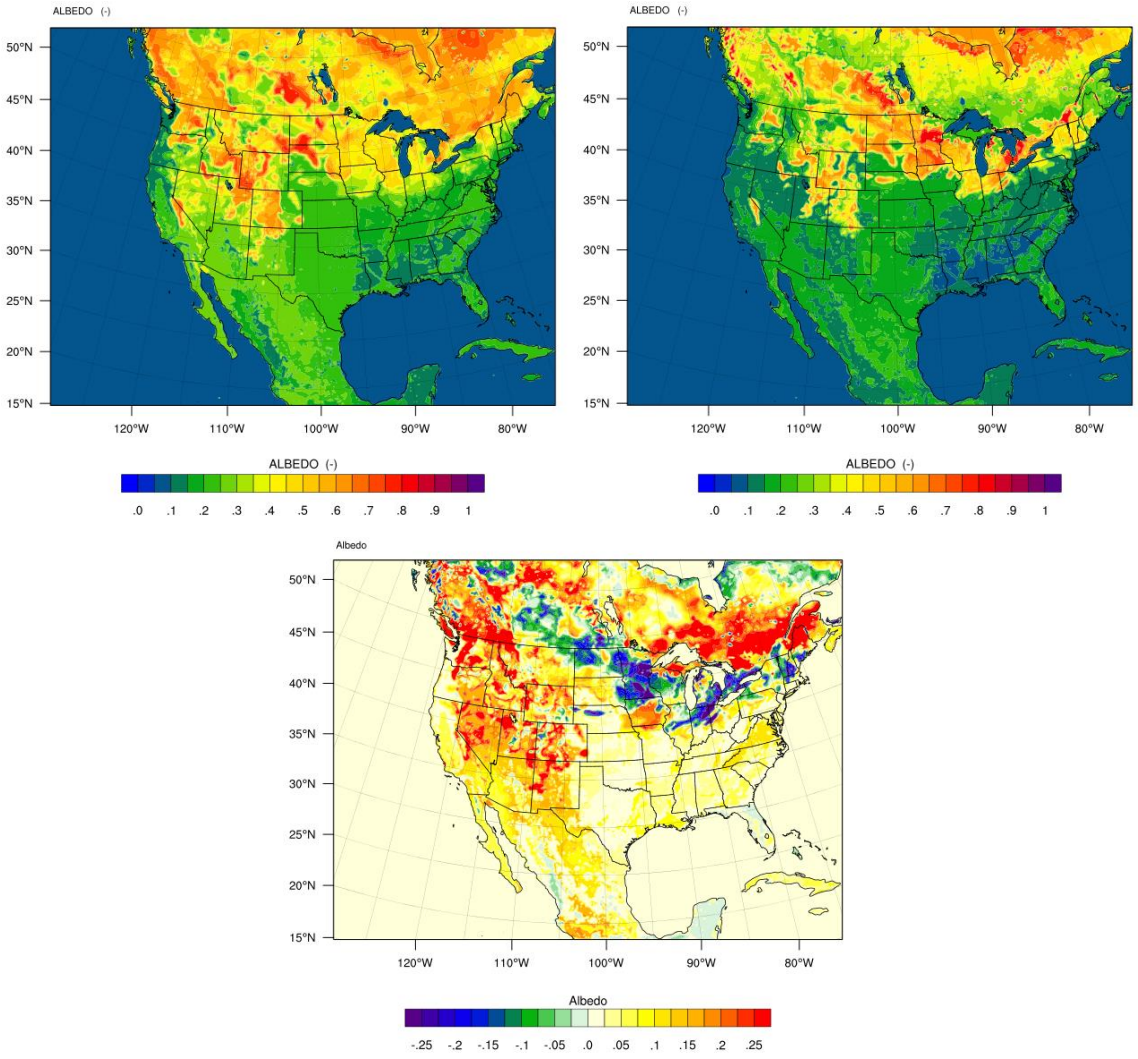


Figure 32. Albedo at the 42-h forecast from the 27 January 2012 00 UTC initialization for (a) AFWAOC, (b) NoahMP, and (c) the difference of the two configurations (AFWAOC – NoahMP).

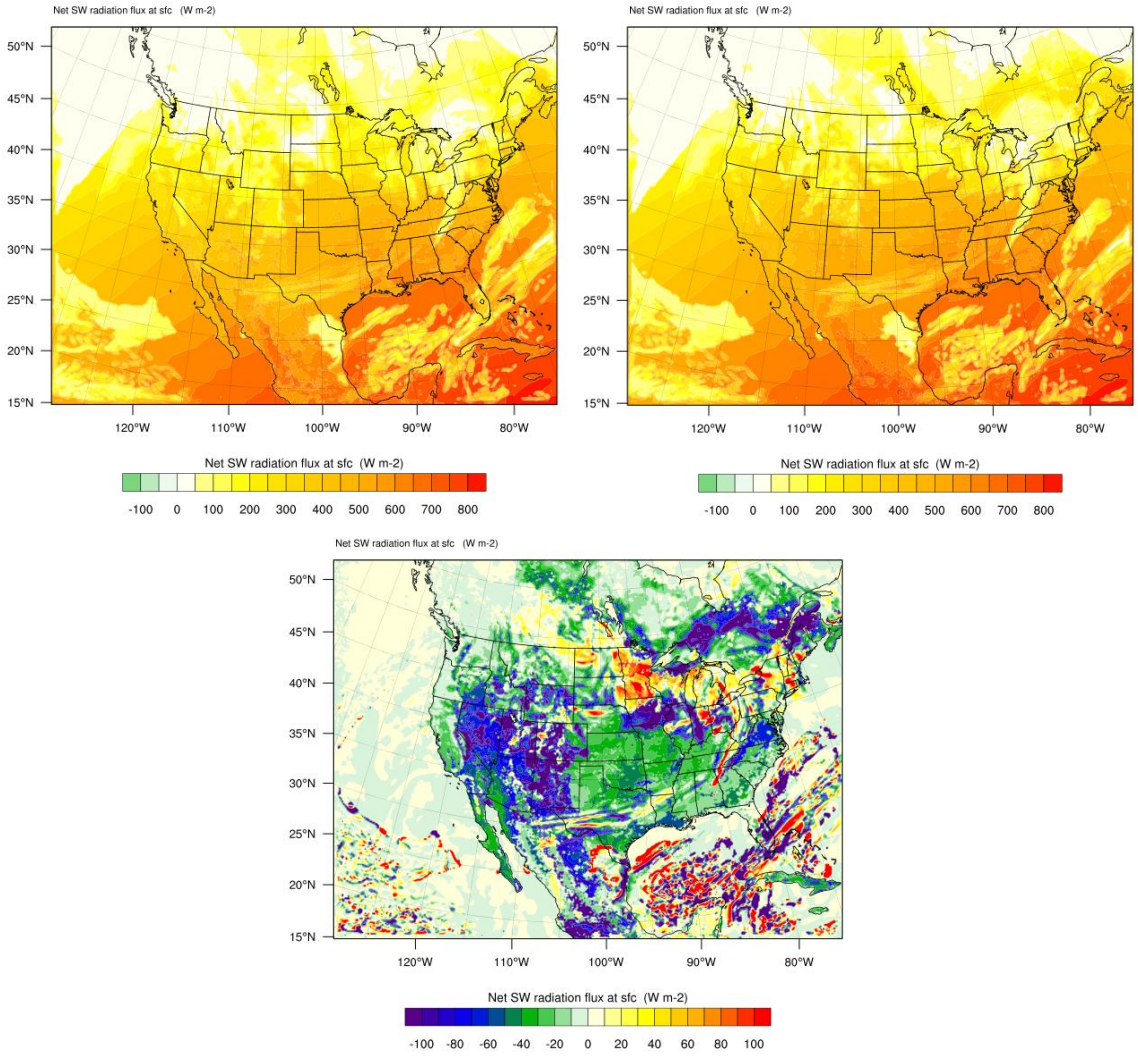


Figure 33. Same as Fig. 32 but for net shortwave radiation flux.

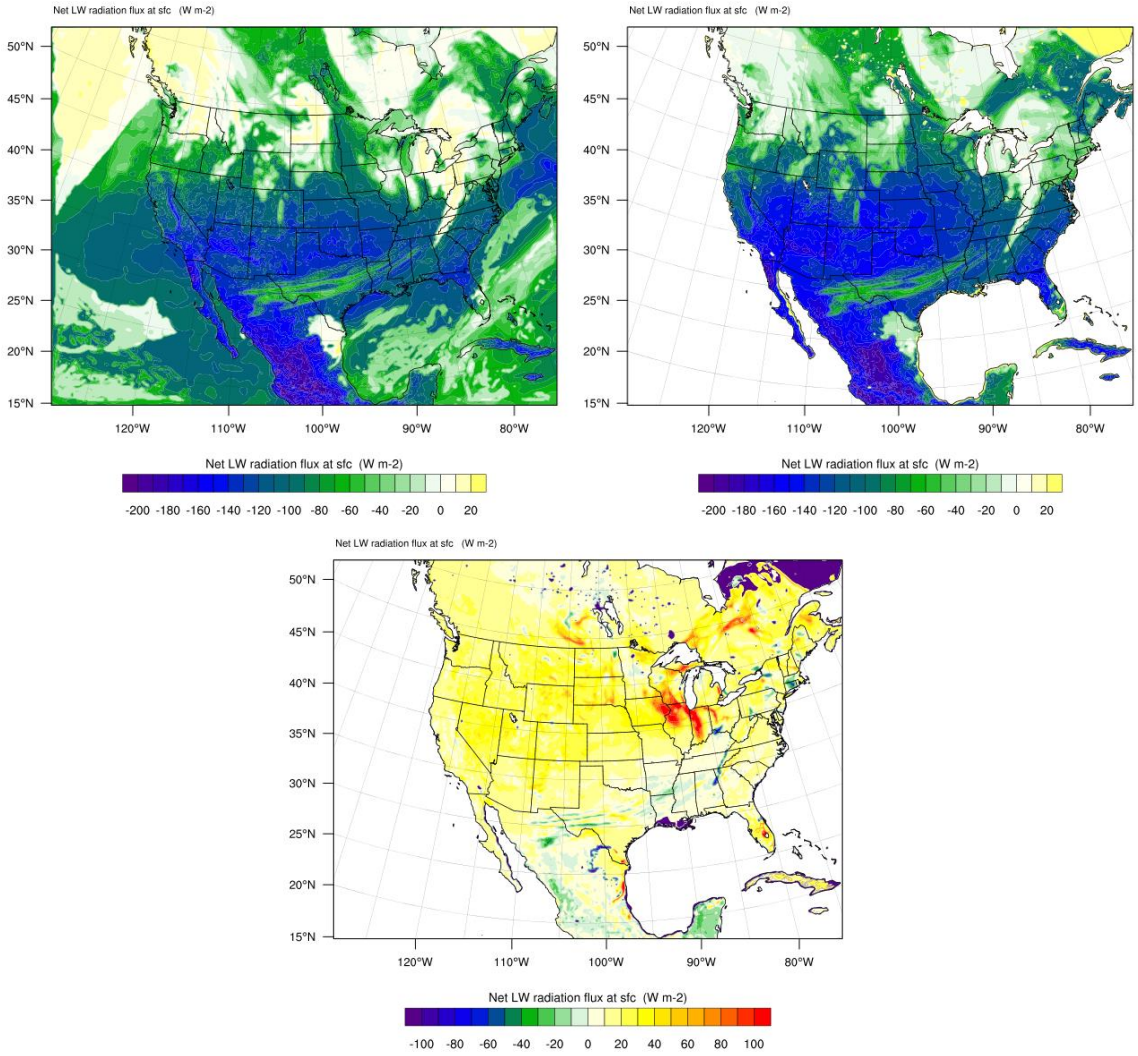


Figure 34. Same as Fig. 32 but for net longwave radiation flux.

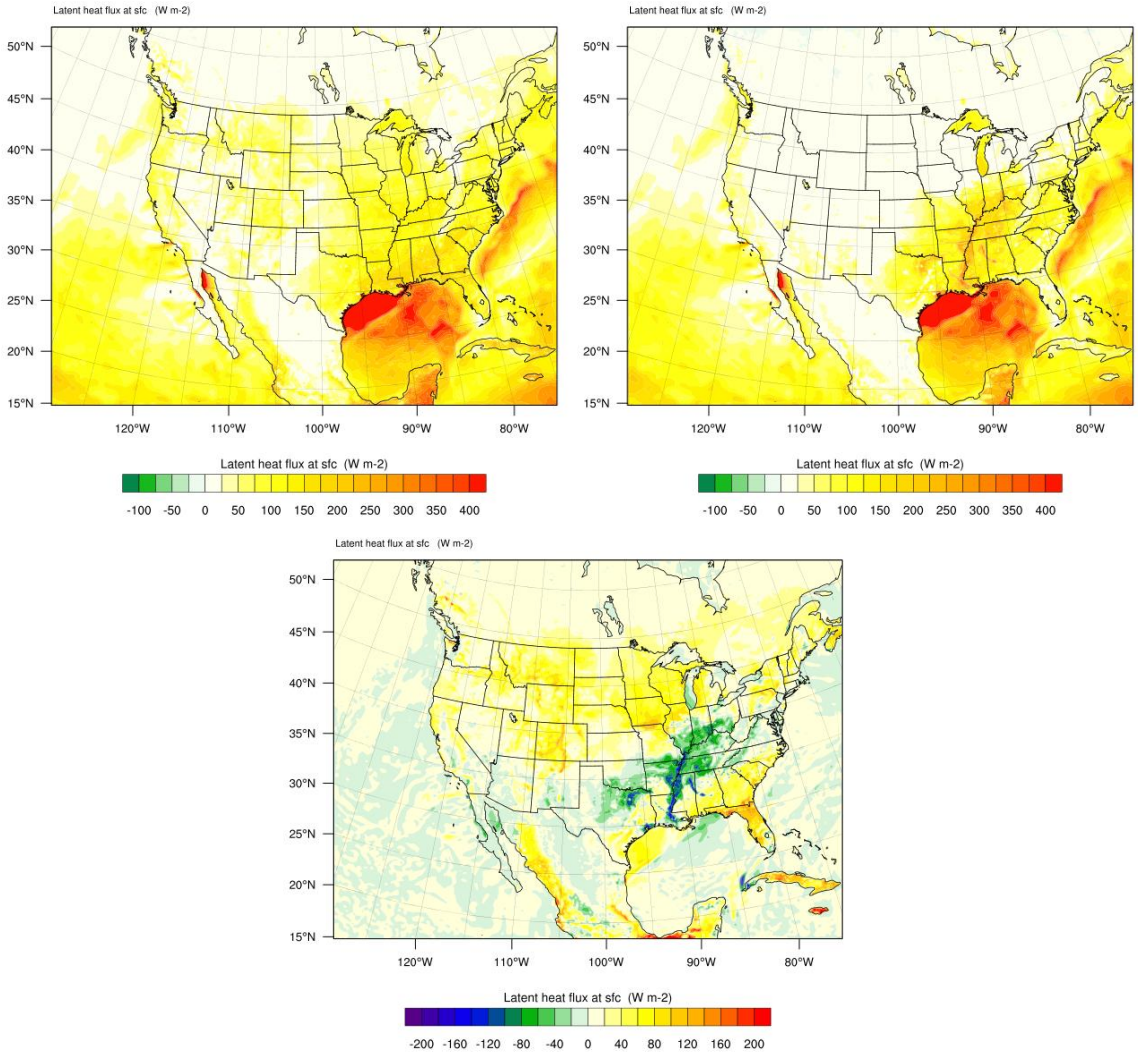


Figure 35. Same as Fig. 32 but for latent heat flux.

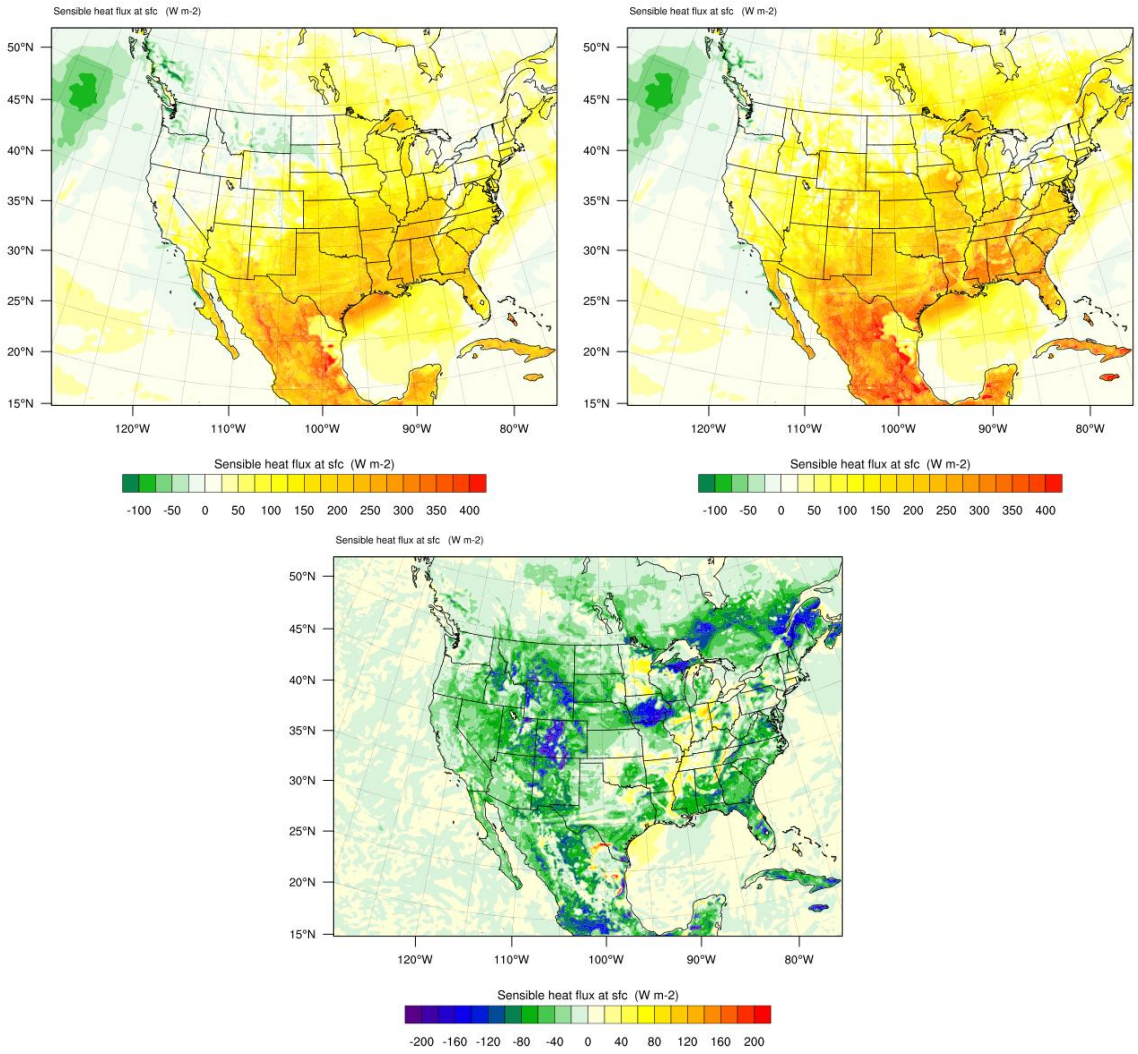


Figure 36. Same as Fig. 32 but for sensible heat flux.

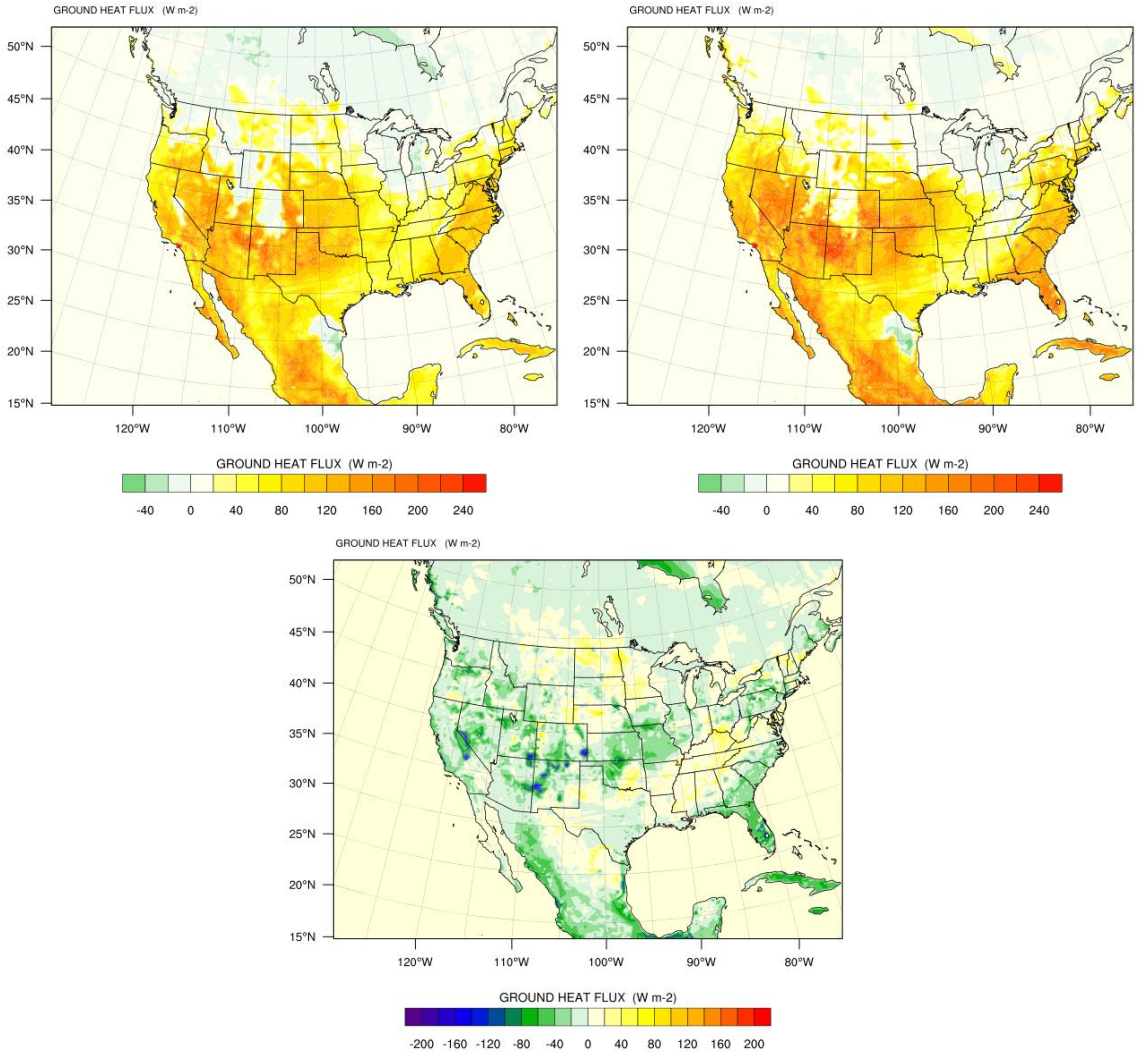


Figure 37. Same as Fig. 32 but for ground heat flux.

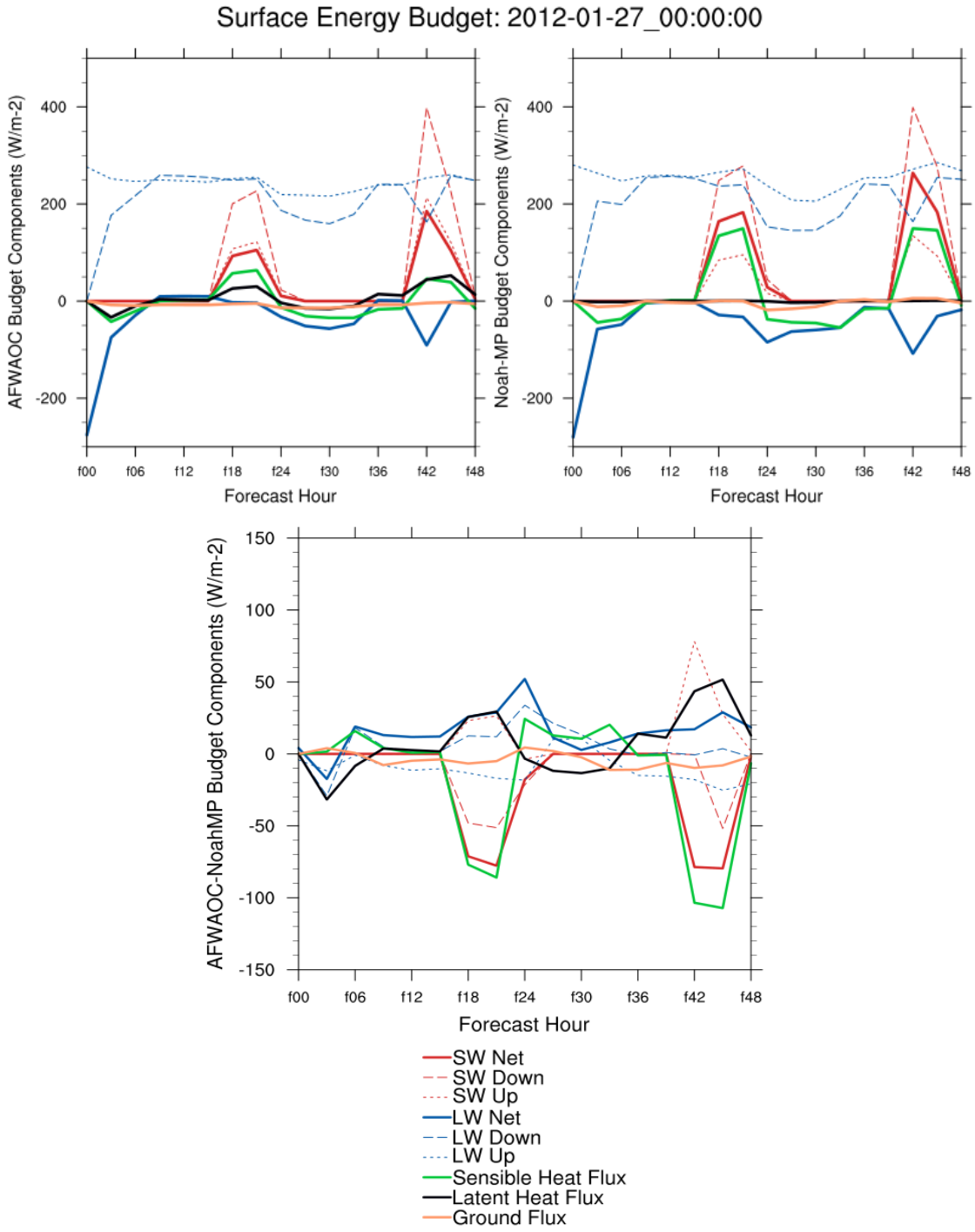


Figure 38. Surface energy budget time series for the 27 January 2012 00 UTC initialization for Site 1 for (a) AFWAOC, (b) NoahMP, and (c) the difference of the two configurations (AFWAOC – NoahMP). The red lines are short-wave radiation fluxes at the surface, where the solid line is net radiation, dashed is downward radiation, and dotted is upward radiation. The blue lines are long-wave radiation fluxes at the surface, where the solid line is net radiation, dashed is downward radiation, and dotted is upward radiation. The solid green line is upward heat flux at the surface. The black solid line is latent heat flux at the surface. The solid orange line is the ground flux.

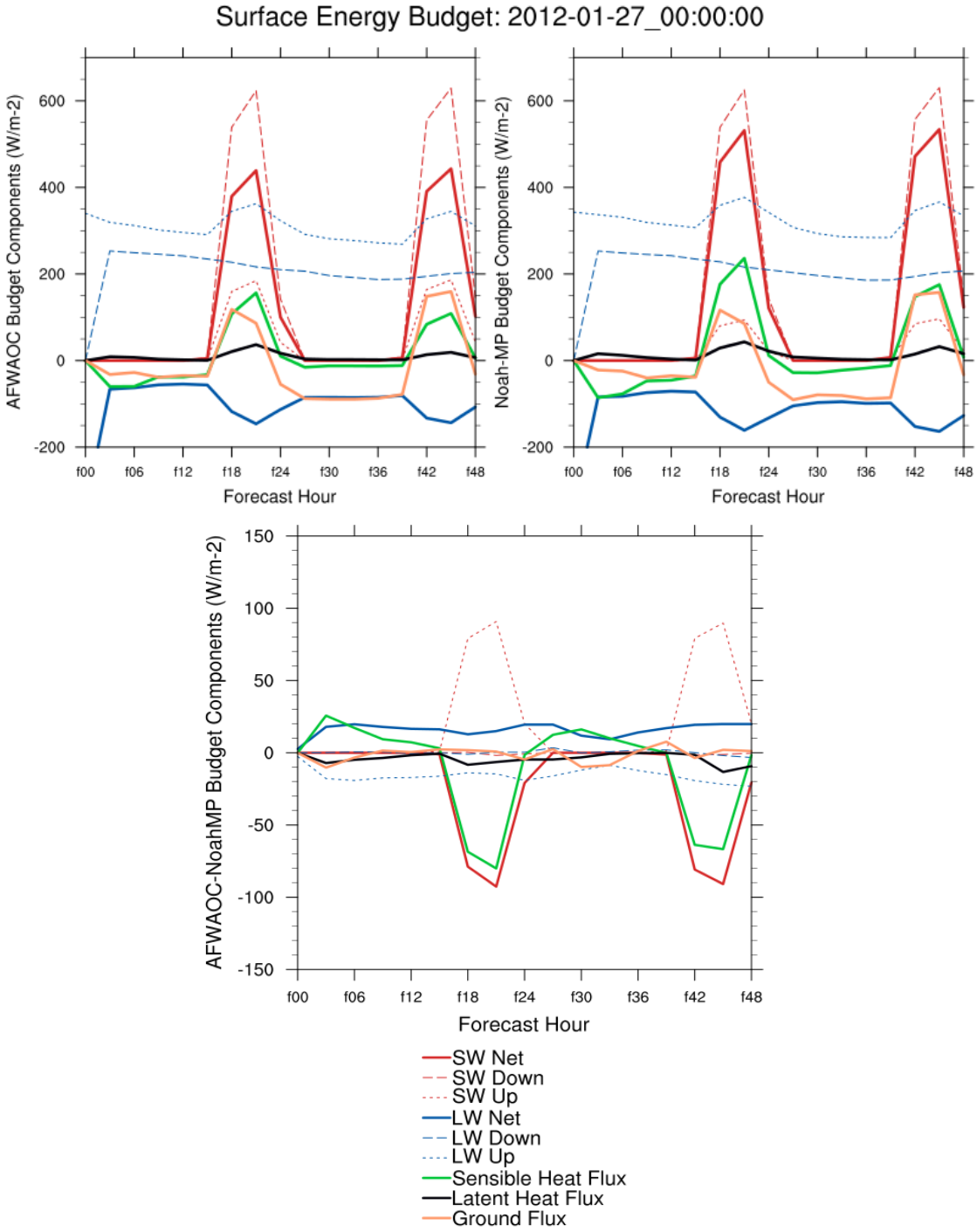


Figure 39. Same as Fig. 38 but for Site 2.

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

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

Appendix B: Subset of the WRF ***namelist.input*** used for this test.

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

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

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

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

&domains
time_step                           = 90,
time_step_fract_num                 = 0,
time_step_fract_den                 = 1,
max_dom                             = 1,
e_we                                = 403,
e_sn                                = 302,
e_vert                              = 57,
num_metgrid_levels                  = 27,
num_metgrid_soil_levels             = 4,
dx                                  = 15000,
dy                                  = 15000,
p_top_requested                     = 1000,
interp_type                         = 1,
lowest_lvl_from_sfc                 = .false.,
lagrange_order                      = 1,
force_sfc_in_vinterp                = 6,
zap_close_levels                    = 500,
adjust_heights                      = .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,
num_land_cat = 28,
mp_zero_out = 2,
/

&noah_mp
Dveg = 4,
opt_crs = 1,
opt_sfc = 1,
opt_btr = 2,
opt_run = 3,
opt_frz = 1,
opt_inf = 1,
opt_rad = 3,
opt_alb = 2,
opt_snf = 1,
opt_tbot = 2,
opt_stc = 1,
/

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

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

Appendix C: Surface verification statistics over the sub-region of MDW

The region of MDW encompasses the states of Minnesota, Iowa, Wisconsin, Michigan, and Ohio, and includes the northern portions of Illinois and Indiana, and the western portions of Pennsylvania and New York. Within the sub-region, the main land-use types are mixed forest in the northern portions of the region, evergreen needleleaf in the extreme northwest and eastern edge, crop and grassland through the central portion, and dryland crop and pasture through the southern portion of the region. Depending on the variable and/or season, the verification statistics for MDW may also be representative of NPL, SPL, and LMV.

Temperature BCRMSE and bias for MDW

Similar to the CONUS aggregations, a diurnal trend with a general increase in error is seen for both configurations (Fig. C.1; 00 UTC initializations only shown). The weakest diurnal signal is observed in the annual aggregation for both configurations, while in the summer aggregation the diurnal signal between the two configurations is out of phase (e.g., peak for NoahMP at 15 UTC where AFWAOC displays a minimum). A number of PS pair-wise differences are noted in the annual and summer aggregations, while only a few pair-wise differences are noted for the winter aggregation. The AFWAOC is consistently the better performer for valid times between 12 – 18 UTC for both the annual and summer; however, during the 6 – 9 UTC valid times in the summer NoahMP is the better performer. While we see a consistent offset in the median values for BCRMSE for the winter aggregation, where NoahMP has higher values, only a few forecast lead times (12-, 15- and 42-h) are PS.

For 2 m temperature bias, the NoahMP configuration generally has a larger cold bias for all temporal aggregations for the MDW region (Fig. C.2); the only exceptions are during the summer aggregation between the 03 – 09 UTC valid times (which are PS regardless of initialization time) and near 21 UTC during the winter aggregation (which are PS for 12 UTC initializations). This leads to the AFWAOC configuration being favored in most cases when pair-wise differences are noted. Similar to the CONUS aggregation, a diurnal signal is noted in the AFWAOC for the MDW region, where the bias tends to be smallest during the early morning hours and largest during the late evening hours (bias increases during the daytime hours). The NoahMP configuration has a similar signal to AFWAOC for the fall aggregation, a weak diurnal signal for the winter aggregation, and a distinct double peak structure during the spring and summer aggregations resulting in lower bias around both 06 and 18 UTC and maximum bias at both 12 and 00 UTC.

Dew Point Temperature BCRMSE and bias for MDW

While the median values tend to be lower for the MDW region as compared to the CONUS-wide aggregation, the BCRMSE distribution is similar for both the 00 and 12 UTC initializations (Fig C.3; 00 UTC initializations only shown) and for all seasonal aggregations. A general increase in error with forecast lead time is also noted. All pair-wise differences between the two configurations are PS and favor AFWAOC with exception to the overnight hours for the summer aggregation, which favor NoahMP.

For the annual aggregation of 2 m dew point temperature bias (Fig. C.4), the diurnal signal is very similar to that of the CONUS aggregation where both configurations have a peak in moisture bias (too wet) around 00 UTC and a minimum around 12 UTC (unbiased). The NoahMP configuration is offset toward a higher moist bias during the overnight hours (03 – 09 UTC) compared to the AFWAOC leading to PS pair-wise differences favoring AFWAOC during that time. The configurations perform similarly during the daytime hours between 12 and 00 UTC valid times. The spring and summer aggregations are most similar to the annual aggregation; though, the NoahMP configuration is favored with PS pair-wise differences during the daytime hours for the summer aggregation. The fall aggregation generally shows the same characteristics; however, the large CIs about the median value result in very few pair-wise differences and several forecast lead times that are not statistically different from an unbiased forecast for both configurations (not shown). The winter aggregation behaves quite differently, and median bias values for the NoahMP configuration are always lower than the AFWAOC. Again the CIs are quite large

and encompass zero for most of the forecast lead times; however, PS pair-wise differences still exist with the favored configuration dependent on the forecast lead time.

Wind Speed BCRMSE and bias for MDW

The BCRMSE distribution for 10 m wind speed aggregated over the MDW region is similar to the CONUS aggregation for all temporal aggregations (Fig. C.5). Very few SS pair-wise differences are noted for the summer, winter, and spring aggregations; however, a larger number are seen in the fall aggregation, all favoring the AFWAOC.

For 10 m wind speed bias across the MDW region, different temporal aggregations show varying behavior (Fig. C.6). For the spring and summer aggregations, the NoahMP configuration frequently has lower median bias values, which makes it the favored configuration given that a consistent high bias is noted for all forecast lead times. Very few overall SS pair-wise differences are noted for the fall and winter aggregations (especially for the 00 UTC initializations). In contrast to other seasons, when SS pair-wise differences are noted for the fall aggregation, the AFWAOC configuration is favored.

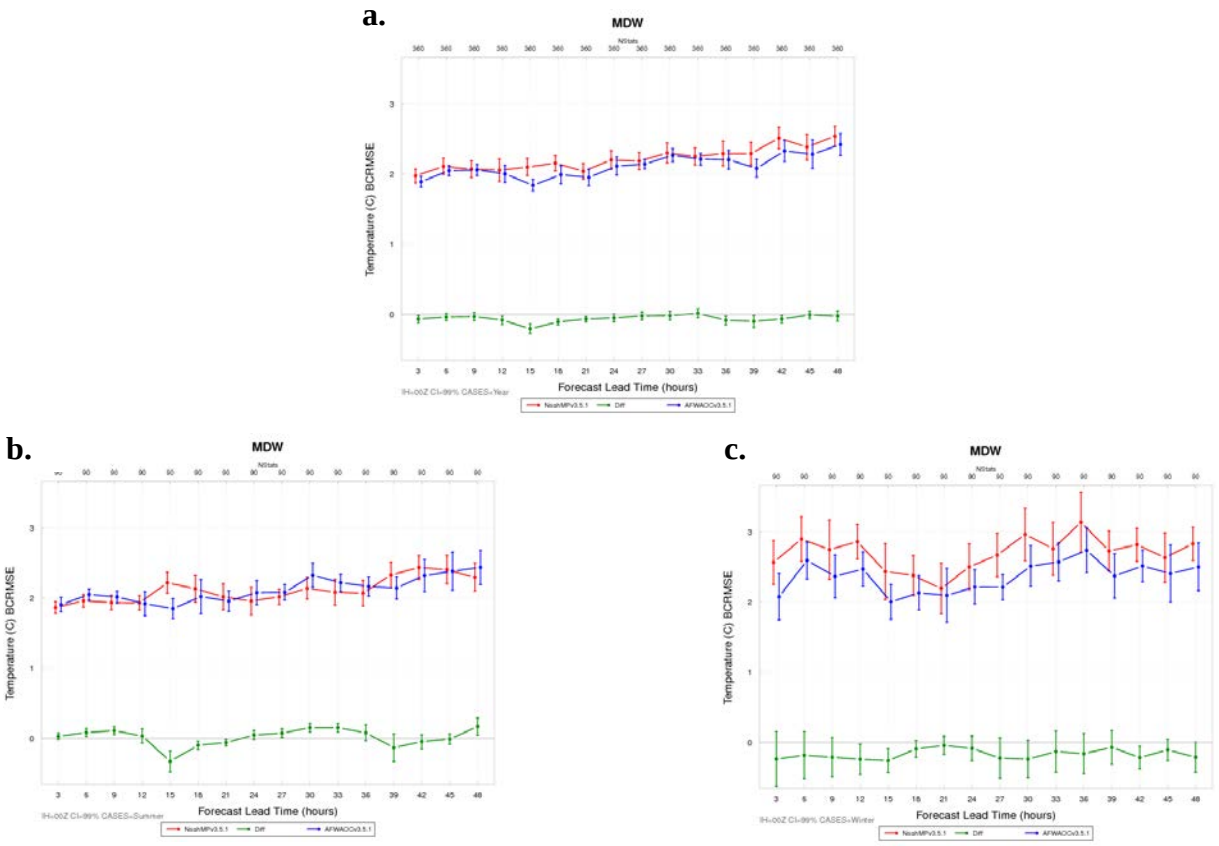


Figure C.1. Time series plot of 2 m AGL temperature (°C) for median BCRMSE for the 00 UTC initializations within the MDW region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

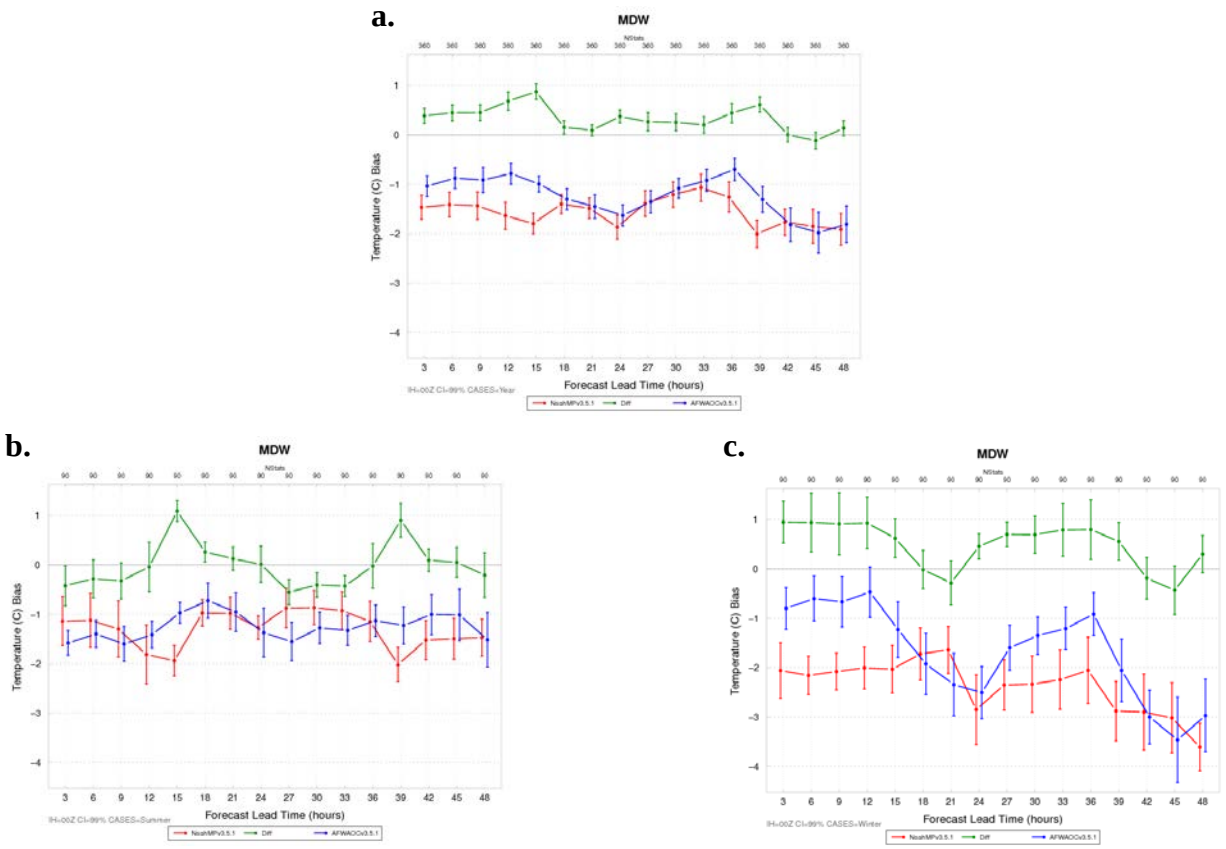


Figure C.2. Time series plot of 2 m AGL temperature (°C) for median mean error (bias) for the 00 UTC initializations within the MDW region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

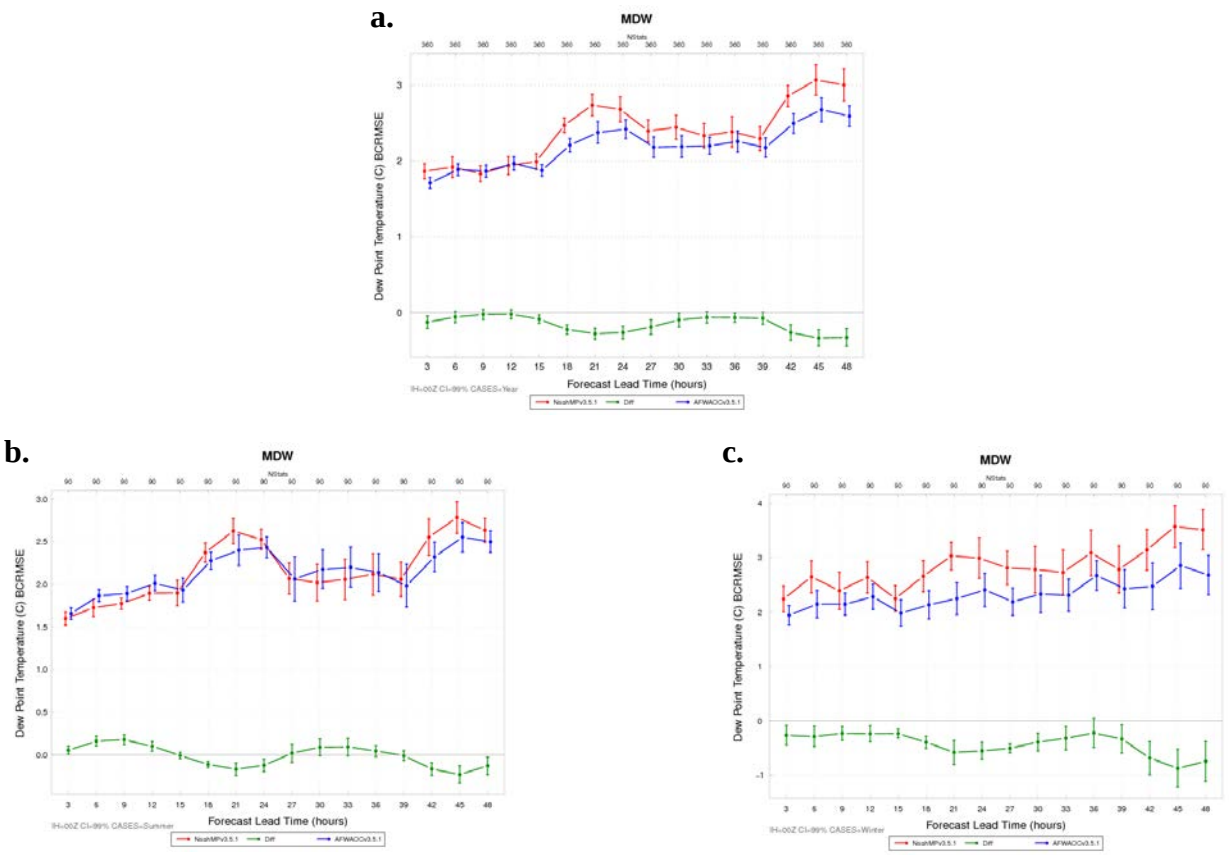


Figure C.3. Time series plot of 2 m AGL dew point temperature (°C) for median BCRMSE for the 00 UTC initializations within the MDW region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

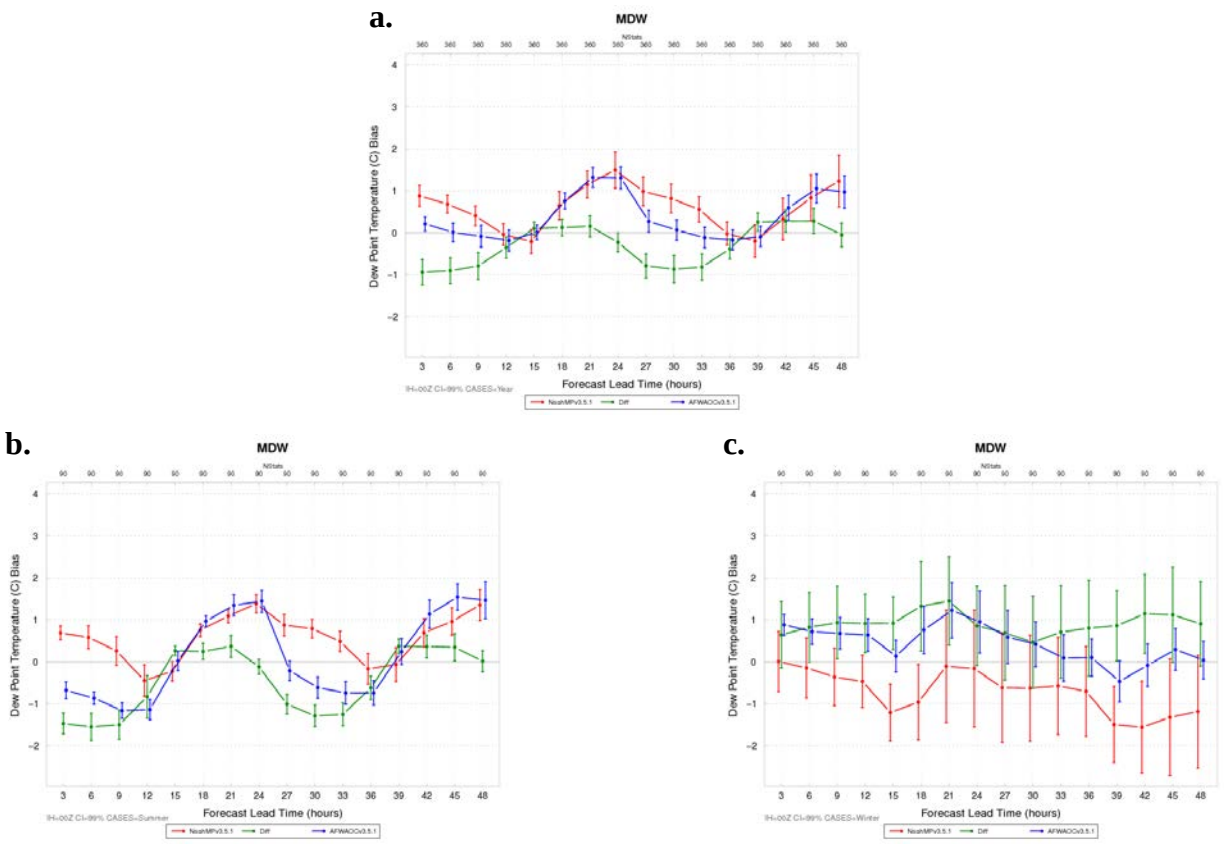


Figure C.4. Time series plot of 2 m AGL dew point temperature (°C) for median mean error (bias) for the 00 UTC initializations within the MDW region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

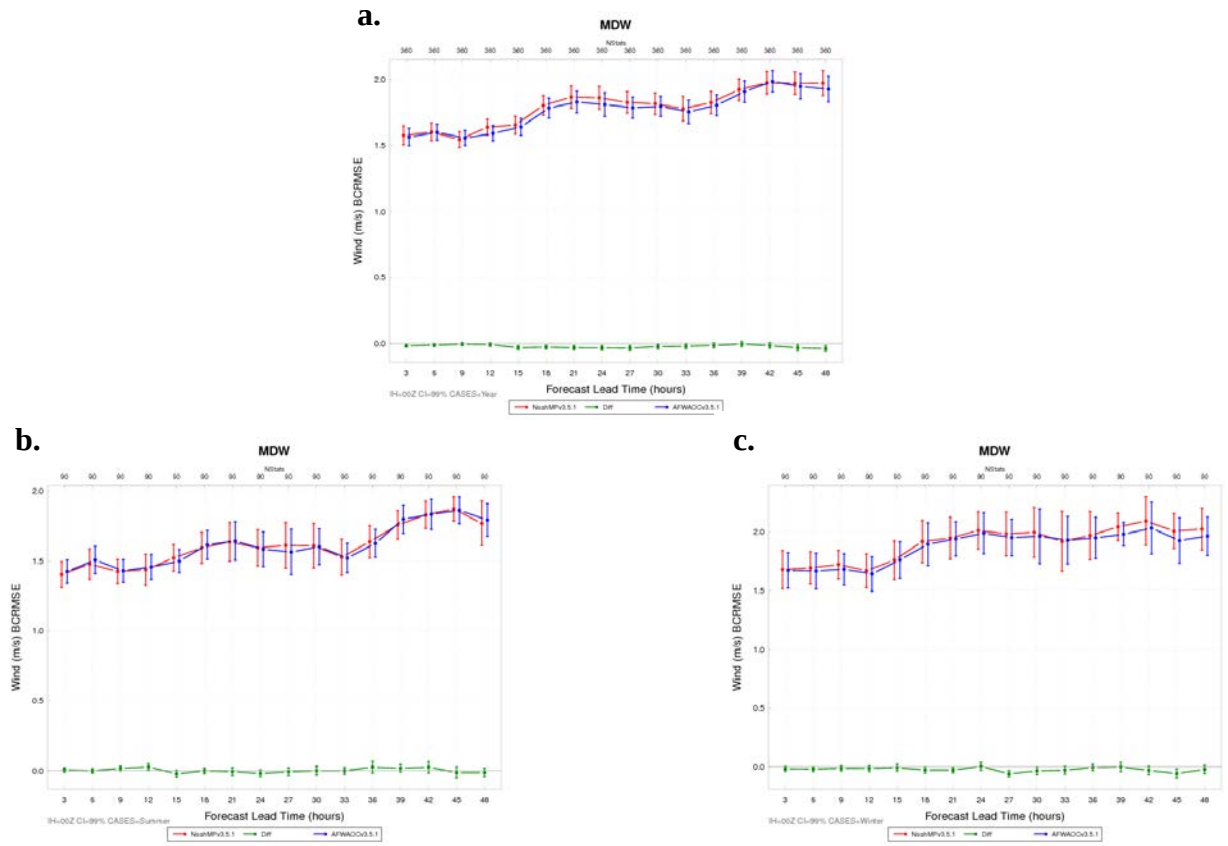


Figure C.5. Time series plot of 10 m AGL wind speed (ms^{-1}) for median BCRMSE for the 00 UTC initializations within the MDW region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

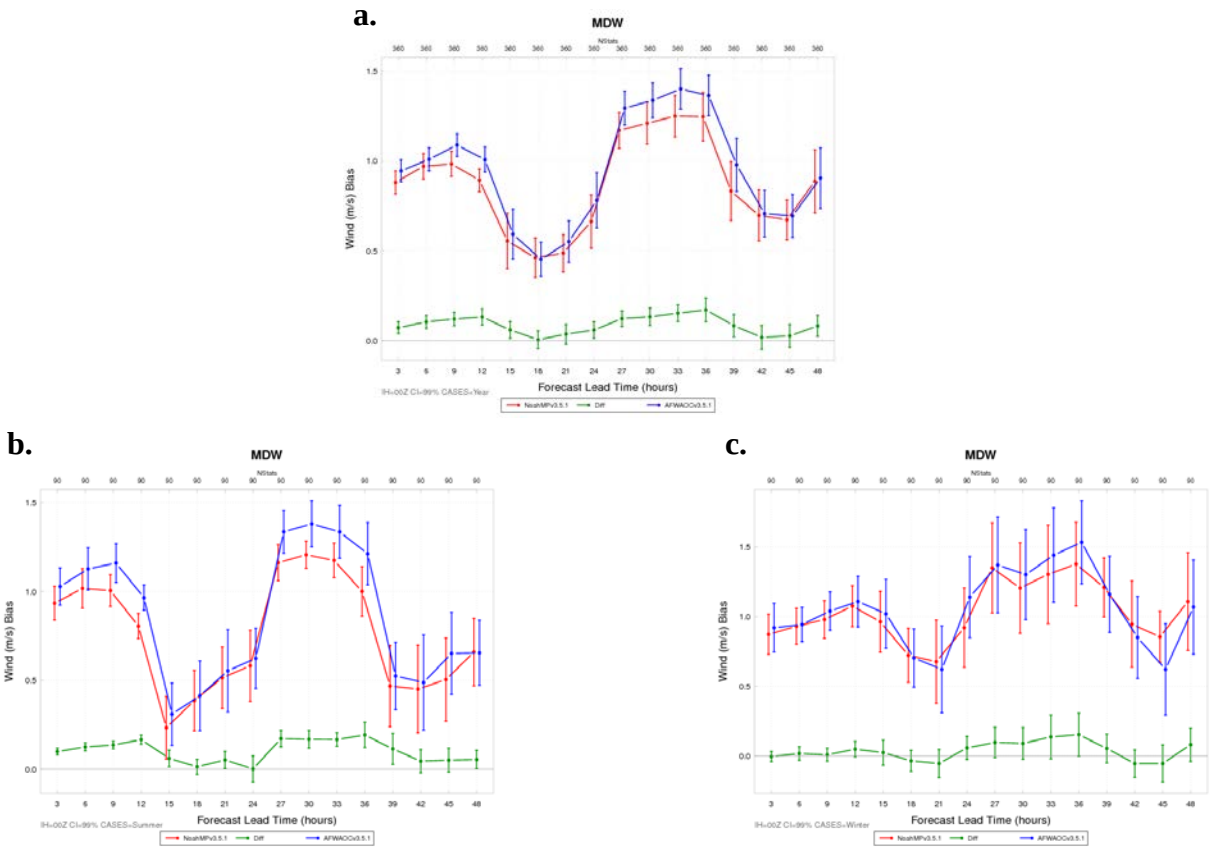


Figure C.6. Time series plot of 10 m AGL wind speed (ms^{-1}) for median mean error (bias) for the 00 UTC initializations within the MDW region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

Appendix D: Surface verification statistics over the sub-region of SMT

The region of SMT centers on the four-corners of Utah, Colorado, New Mexico and Arizona. Within the sub-region, the main land-use types are shrubland in the northwest and south, nestled by some pockets of grassland and savanna in the northeast (over the Rocky Mountains), which is surrounded by areas of evergreen needleleaf. The terrain heights range from 2500 – 3500 m in the northeast and 1200 – 2500 m in the south. Depending on the variable and/or season, the verification statistics for SMT may also be representative of NMT, GRB and SWD.

Temperature BCRMSE and bias for SMT

The median values for 2 m temperature BCRMSE for the SMT region tend to be larger than the corresponding CONUS values for both configurations (Fig. D.1). A more prominent diurnal trend is also seen in BCRMSE over SMT for all seasons. Generally, larger errors are seen at times valid from 00 – 12 UTC and smaller errors seen at times valid from 15 – 21 UTC. The diurnal signal in the summer aggregation is somewhat out of phase with those in the other seasonal aggregations. While the other aggregations have the largest error at 12 UTC and smallest error at 18 UTC, the summer aggregation has the largest error at times valid from 00 – 03 UTC which remains relatively flat at 03 – 12 UTC valid times before decreasing to a minimum at 15 UTC. The differences between NoahMP and AFWAOC are insignificant except at early morning and late afternoon hours, with the NoahMP having PS smaller BCRMSE at 00 UTC valid times (03 UTC in the summer aggregation). In addition, PS larger errors are seen in NoahMP around 09 – 12 UTC in the summer and spring aggregations.

For 2 m temperature bias over SMT, both configurations have cold bias at most forecast lead times, which is generally larger in winter and smaller in summer (Fig. D.2). The bias for all aggregations exhibits diurnal cycles showing largest cold bias in the afternoon (00 UTC for the winter and fall aggregations, and 21 UTC for the summer and spring aggregations) and small cold bias or warm bias before sunrise (12 – 15 UTC). Considering the diurnal variation of the surface temperature, the bias pattern reflects that the model under-estimates the amplitude of the diurnal temperature variations. Between the two configurations, the bias in NoahMP is significantly smaller than that in AFWAOC. Compared to AFWAOC, the 2 m temperature in NoahMP is warmer (smaller cold biases) during daytime (18 – 03 UTC) and similar or slightly colder over nighttime. The 2 m temperature in NoahMP tends to respond to the daytime heating more reasonably than that in AFWAOC.

Dew Point Temperature BCRMSE and bias for SMT

The 2 m dew point temperature BCRMSE for the SMT region is similar in magnitude to the CONUS-wide aggregations (Fig. D.3). A diurnal signal with a gentle increase in errors with time is present for both configurations and all aggregations except the fall (not shown). The diurnal trend generally shows a maximum BCRMSE at 21 – 00 UTC valid times and a minimum at 9 – 15 UTC valid times. For the fall aggregation, no clear diurnal cycle is seen (not shown). Between the two configurations, the NoahMP error is generally larger than in the AFWAOC. PS pair-wise differences favoring AFWAOC mostly appear during nighttime (03 – 09 UTC valid times) in the fall and winter aggregations and at 18 – 00 UTC valid times in the summer aggregation.

The 2 m dew point temperature bias for the SMT region in AFWAOC show ~ 2 ^{o}C wet bias in the afternoon and evening hours (21 – 03 UTC valid times), and smaller wet bias or dry bias overnight through early morning (09 – 15 UTC valid times) (Fig. D.4). There are usually large differences between the two configurations. The dew point temperature forecast by the NoahMP is significantly wetter than the AFWAOC in the spring and summer aggregations (spring not shown) and drier in the fall and winter aggregations (especially during the daytime; fall not shown). The pair-wise differences show that the NoahMP performs better in the winter aggregation and worse in the annual, spring and summer aggregations with mixed performance in the fall aggregation. Among the seasonal aggregations, the NoahMP bias is the driest in the fall and wettest in the spring.

Wind Speed BCRMSE and bias for SMT

The 10 m wind speed BCRMSE for the SMT region tends to be higher as compared to the CONUS-wide aggregation. The diurnal signal is only present in the summer aggregation where the errors are largest at 00 UTC valid time and smallest near 18 UTC (Fig. D.5). The differences between the two configurations are generally small, with a few SS pair-wise differences, mostly noted at 03 – 06 UTC valid times and favoring NoahMP.

A prominent diurnal signal is present in the 10 m wind speed bias over SMT for all temporal aggregations and both initializations, with the largest errors (high bias) seen at times valid 06 – 18 UTC and lowest errors at times valid 21 – 03 UTC (Fig. D.6). In the spring and summer aggregations, the median bias values are below 0 (i.e., low bias) in the daytime hours. Generally the differences between the two configurations are such that NoahMP performs better in the summer and spring aggregations (mainly at nighttime) and worse in the fall and winter aggregations (mainly at daytime).

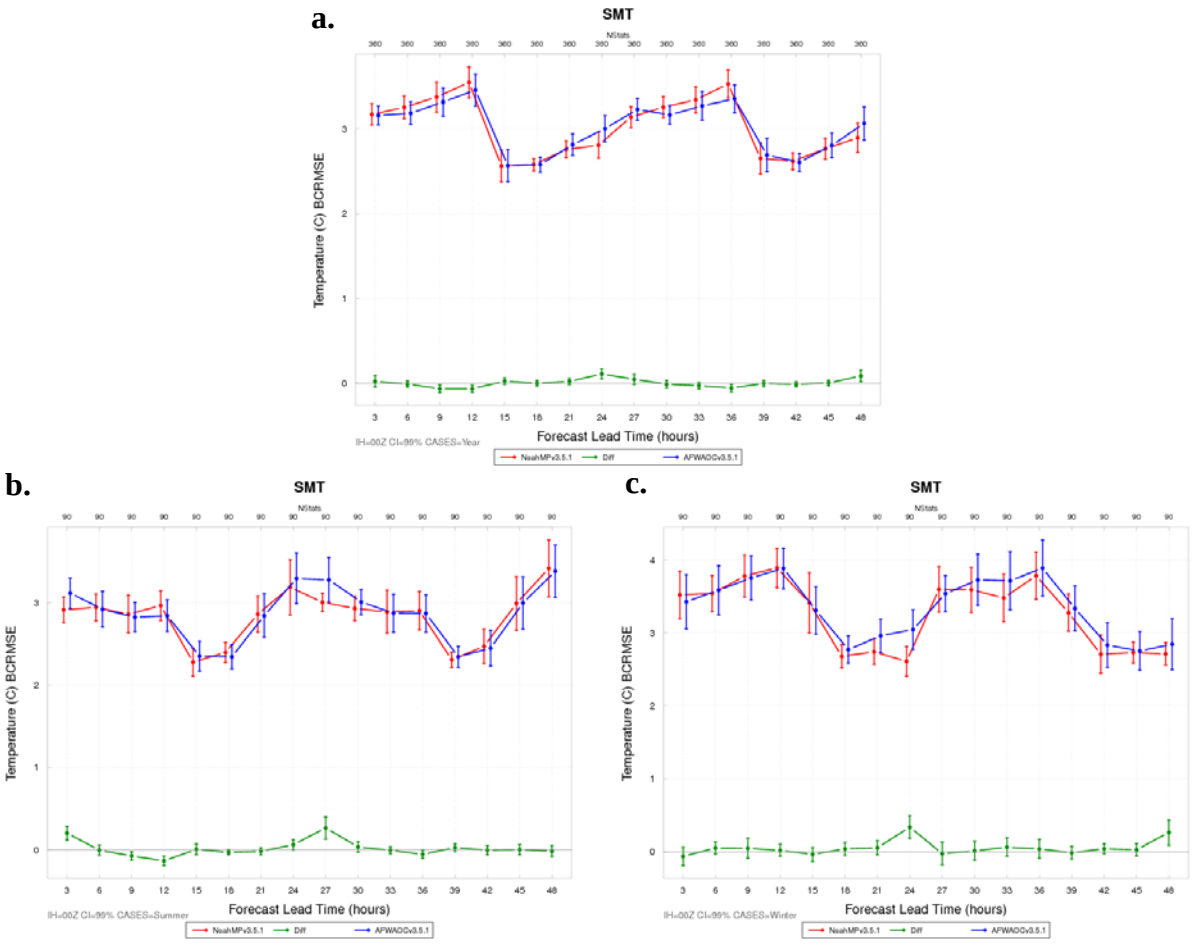


Figure D.1. Time series plot of 2 m AGL temperature (°C) for median BCRMSE for the 00 UTC initializations within the SMT region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

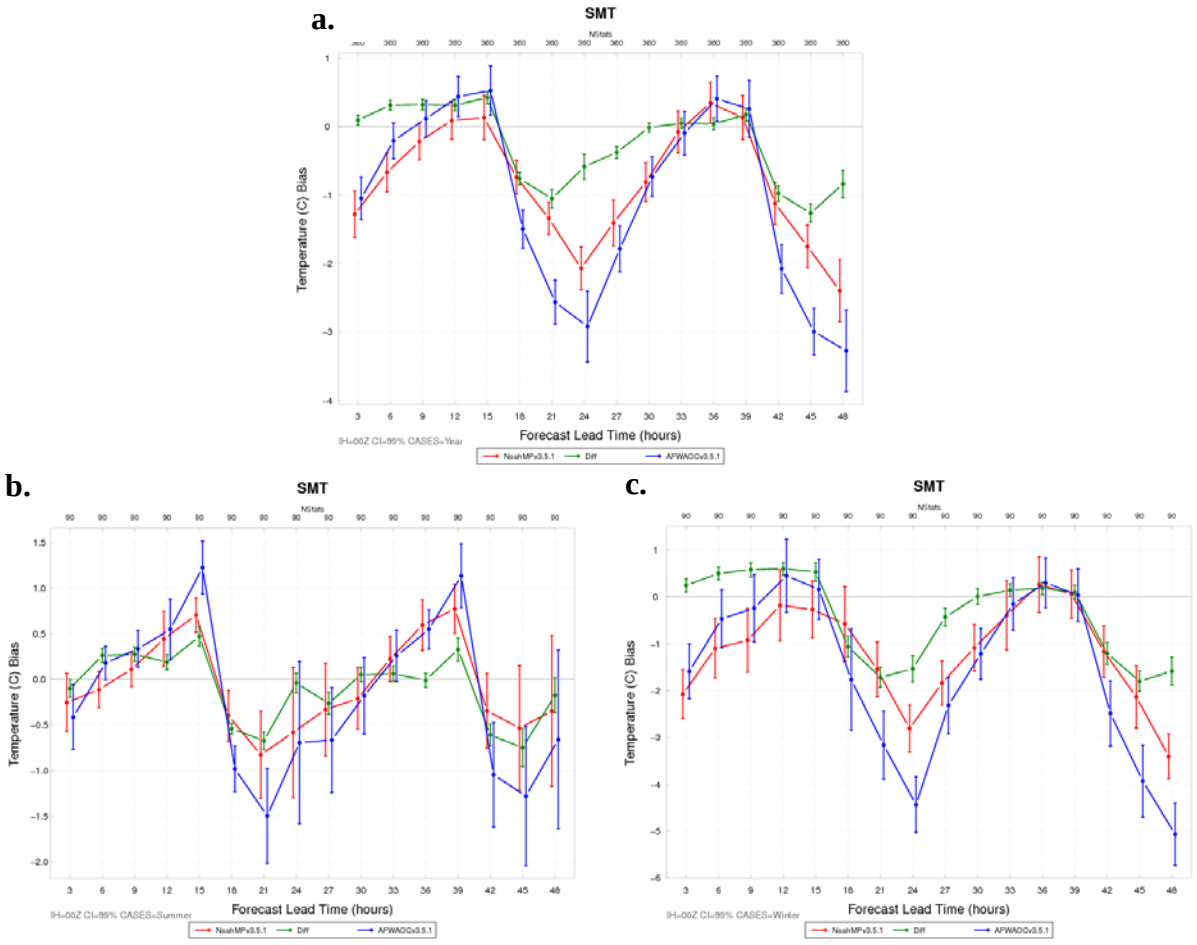


Figure D.2. Time series plot of 2 m AGL temperature (°C) for median mean error (bias) for the 00 UTC initializations within the SMT region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

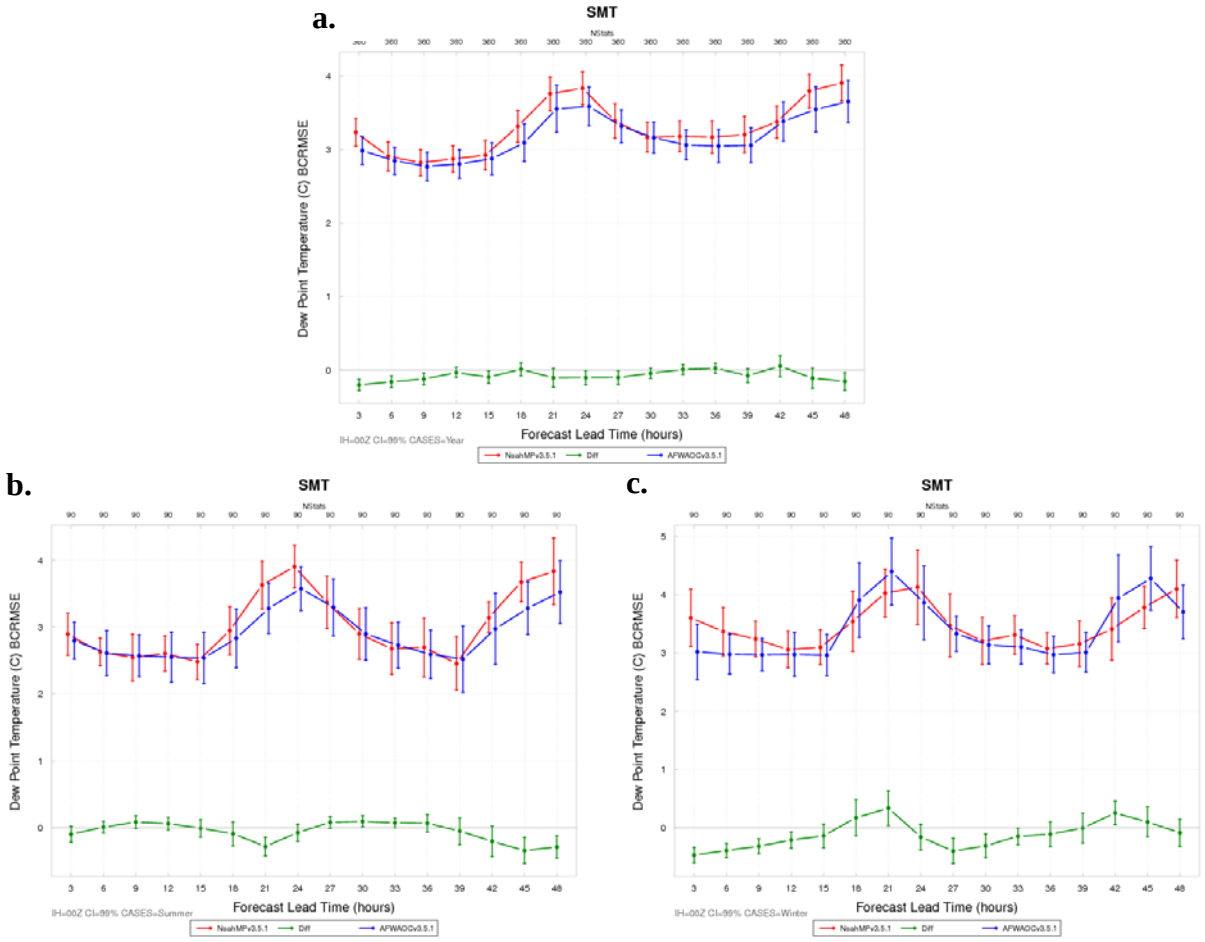


Figure D.3. Time series plot of 2 m AGL dew point temperature (°C) for median BCRMSE for the 00 UTC initializations within the SMT region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

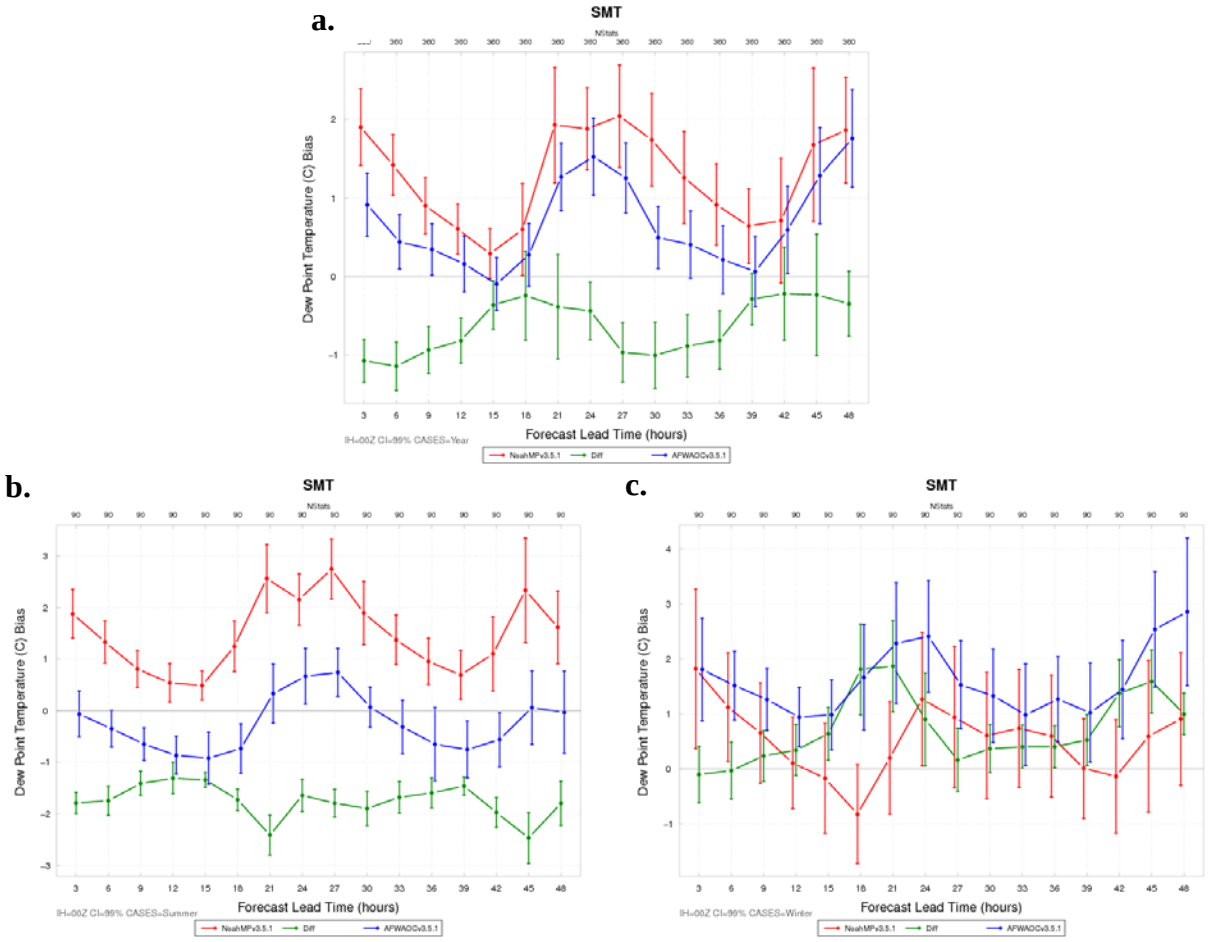


Figure D.4. Time series plot of 2 m AGL dew point temperature (°C) for median mean error (bias) for the 00 UTC initializations within the SMT region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

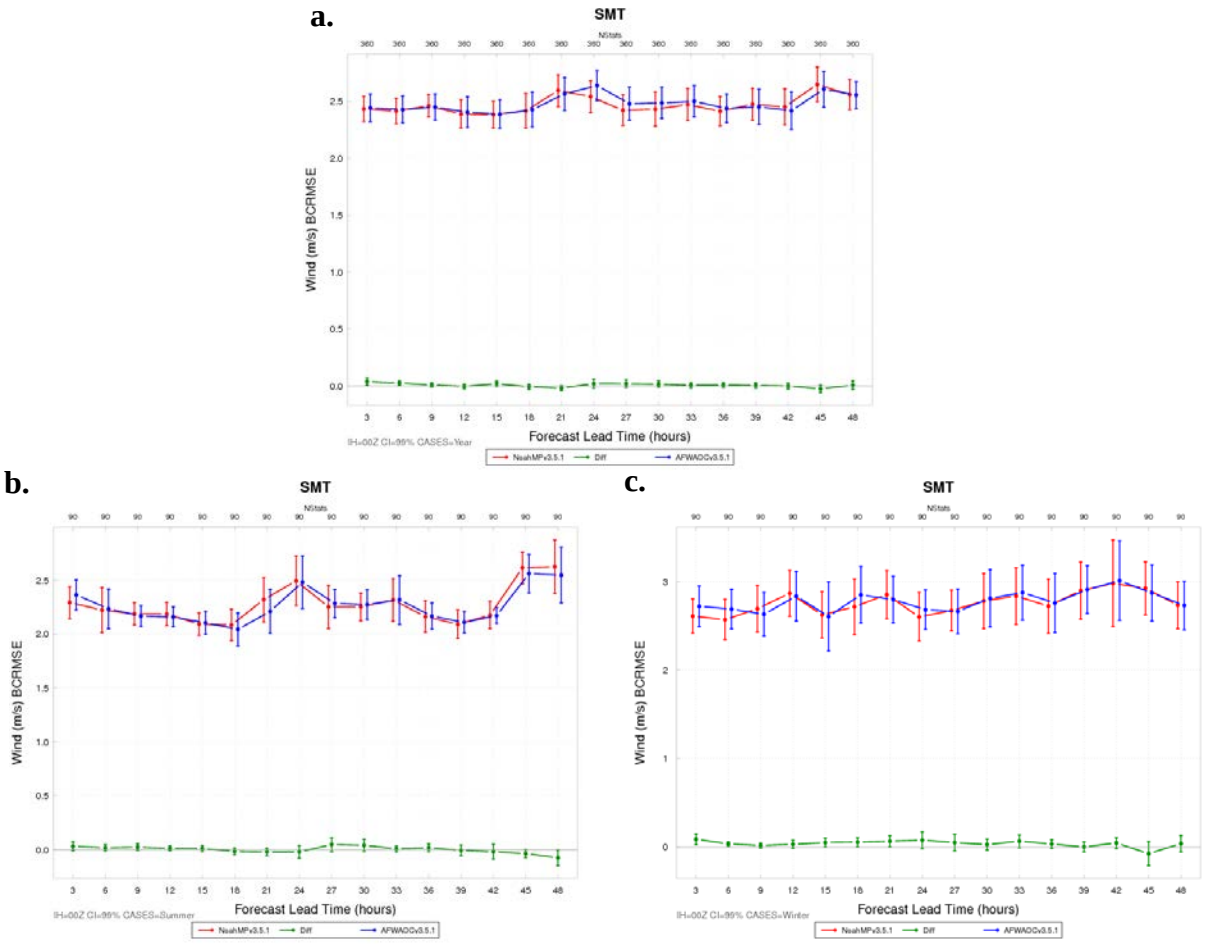


Figure D.5. Time series plot of 10 m AGL wind speed (ms^{-1}) for median BCRMSE for the 00 UTC initializations within the SMT region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

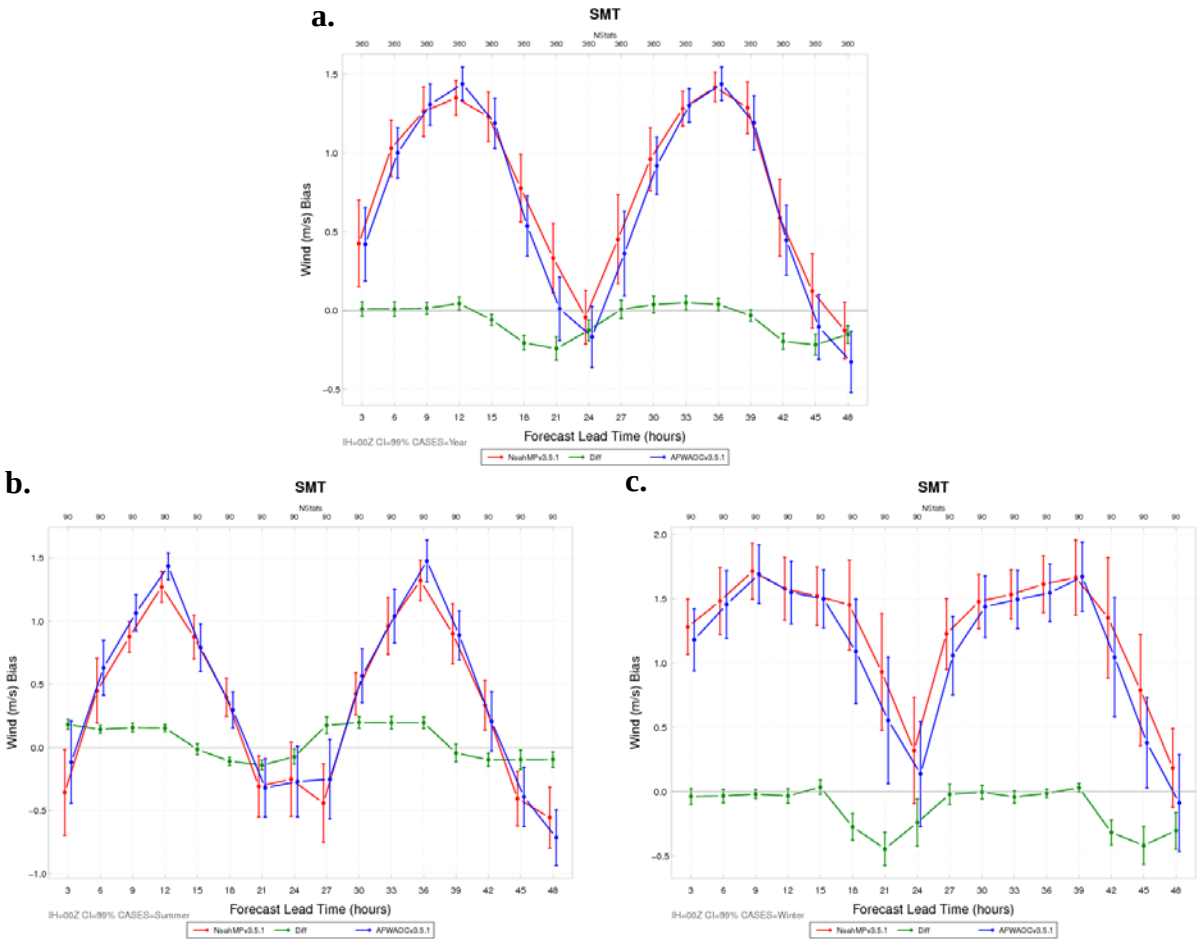


Figure D.6. Time series plot of 10 m AGL wind speed (ms^{-1}) for median mean error (bias) for the 00 UTC initializations within the SMT region aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

Appendix E: Surface verification statistics over the sub-region of GMC

The sub-region of GMC covers coastal areas of southeast TX, LA, southern MS, southern AL, and the panhandle of FL. In the sub-region, the primary land-use types are dryland crop and pasture, crop and grassland mosaic, crop and woodland mosaic, evergreen needleleaf forest, and grassland. The sub-region's topography does not surpassing 250 m and has a strong influence from being situated near coastlines.

Temperature BCRMSE and bias for GMC

Similar to the CONUS aggregation, all seasonal aggregations have a diurnal trend with an overall increase with forecast lead time, with the summer and fall aggregations having the largest diurnal amplitude and winter having the weakest signal (Fig. E.1; only 00 UTC aggregations). Timing of minimum and maximum errors vary by seasonal aggregation; in the annual, summer, and spring aggregations, peak BCRMSE values are centered around time valid at 21 UTC, while in the fall aggregation, the highest errors occur between times valid from 06 – 12 UTC. For the times valid 15 – 00 UTC, AFWAOC typically had lower median BCRMSE values, which is reflected in the PS pair-wise differences, where if differences are noted, AFWAOC is always the better performer. However, with exception to winter, at times valid 06 -12 UTC, NoahMP is generally favored when SS or PS pair-wise differences are observed.

The distributions of temperature bias display a variety of behaviors depending on configuration and temporal aggregation for both 00 and 12 UTC initializations, with the largest divergence in bias occurring in the spring and fall aggregations (Fig. E.2; only 00 UTC aggregations). With exception to the winter aggregation, where median values are generally closer between the two configurations, and a number of times valid at 15 UTC in all temporal aggregations, NoahMP typically has higher median biases than AFWAOC. In the annual and spring aggregations, AFWAOC has a cold bias at nearly all forecast lead times, while NoahMP is unbiased at a number of forecast lead times. In the summer aggregation, both configurations perform well, with a majority of forecasts beyond the 18 h forecast being unbiased; however, at the 42 – 48 h forecast lead times, the CIs become quite large indicating higher uncertainty. Both configurations have a well-defined diurnal signal in the fall and winter aggregations, with the winter aggregation having a cold bias at nearly all forecast lead times; the smaller cold bias values are generally seen in the overnight and early morning (i.e, 03 – 12 UTC), while the larger cold bias values are seen during the day (i.e., 15 – 21 UTC, depending on configuration and season). All pair-wise differences are PS, with a majority showing NoahMP being closer to the observations than AFWAOC. With a focus on the 00 UTC initializations, most differences favoring NoahMP are seen after the 15 hour forecast lead time and occur in the annual, fall, and spring aggregations.

Dew Point Temperature BCRMSE and bias for GMC

Dew point temperature BCRMSE gently grows with forecast lead time for all seasonal aggregations and both 00 and 12 UTC initializations (Fig. E.3; only 00 UTC initializations shown). In addition, a strong diurnal signal is noted for all seasonal aggregations, with exception to the fall season, where only a weak signal is noted. The peak in BCRMSE values is at and around times valid at 21 UTC for all seasonal aggregations, especially for NooahMP, which has a more amplified BCRMSE curve. This translates to AFWAOC being a better performer at times valid 18 – 00 UTC (i.e., when both configurations have peak errors); at times valid from 06 – 12 UTC (i.e., when both configurations generally have smaller errors), if there is a pair-wise difference, it favors NoahMP.

For both configurations and initializations and all seasonal aggregations, there is a diurnal signal in bias; however, the shape and values of the distribution are heavily dependent on configuration and seasonal aggregation (Fig. E.4; only 00 UTC initializations shown). A noted consistency across the seasonal aggregations occurs in the overnight hours (i.e., 03 – 12 UTC valid times), where NoahMP has higher median bias values than AFWAOC, which is typically either unbiased or has a small wet bias; therefore, at these times AFWAOC is typically closer to the observations and is favored. In the annual, summer, fall, and spring aggregations, AFWAOC and NoahMP are slightly out of phase with each other; this results in NoahMP occasionally having lower median high bias values than AFWAOC (i.e., NoahMP is

closer to the observations and is favored) at times valid at and around 15 – 00 UTC. In the winter aggregation, NoahMP has consistently higher median dew point temperature biases than AFWAOC at all forecast lead times, which results in AFWAOC being closer to the observations and favored at a majority of forecast lead times. All pair-wise differences are PS for dew point temperature bias.

Wind Speed BCRMSE and bias for GMC

In all temporal aggregations for wind speed BCRMSE, a diurnal signal with increasing error with lead time is noted, with the annual and summer aggregations having the largest amplitudes of errors (Fig. E.5; only 00 UTC initializations shown). In general, peak errors occur at and around times valid 21 – 00 UTC, with the smallest errors in the overnight and early morning hours (i.e., 06 – 12 UTC, with dependence on seasonal aggregation). No PS pair-wise differences are noted; several of SS pair-wise differences are observed, with AFWAOC being favored more than NoahMP.

Similar to the CONUS, GMC has a very prominent diurnal trend in wind speed bias for both 00 and 12 UTC initializations and for all temporal aggregations (Fig. E.6; only 00 UTC initializations shown). The highest biases are typically at times valid 03 – 12 UTC and are always positive; the smallest biases (i.e., closer to zero) occur at times valid 15 – 21 UTC, and depending on configuration and temporal aggregation, there are high, neutral, and low biases, with AFWAOC typically having lower median bias values than NoahMP. No PS pair-wise differences are noted. However, a number of SS pair-wise differences are noted, with most occurring at times valid 15 – 00 UTC and favoring AFWAOC. Exceptions include times valid at 12 UTC and in the summer aggregation, where NoahMP is more often closer to observations than AFWAOC.

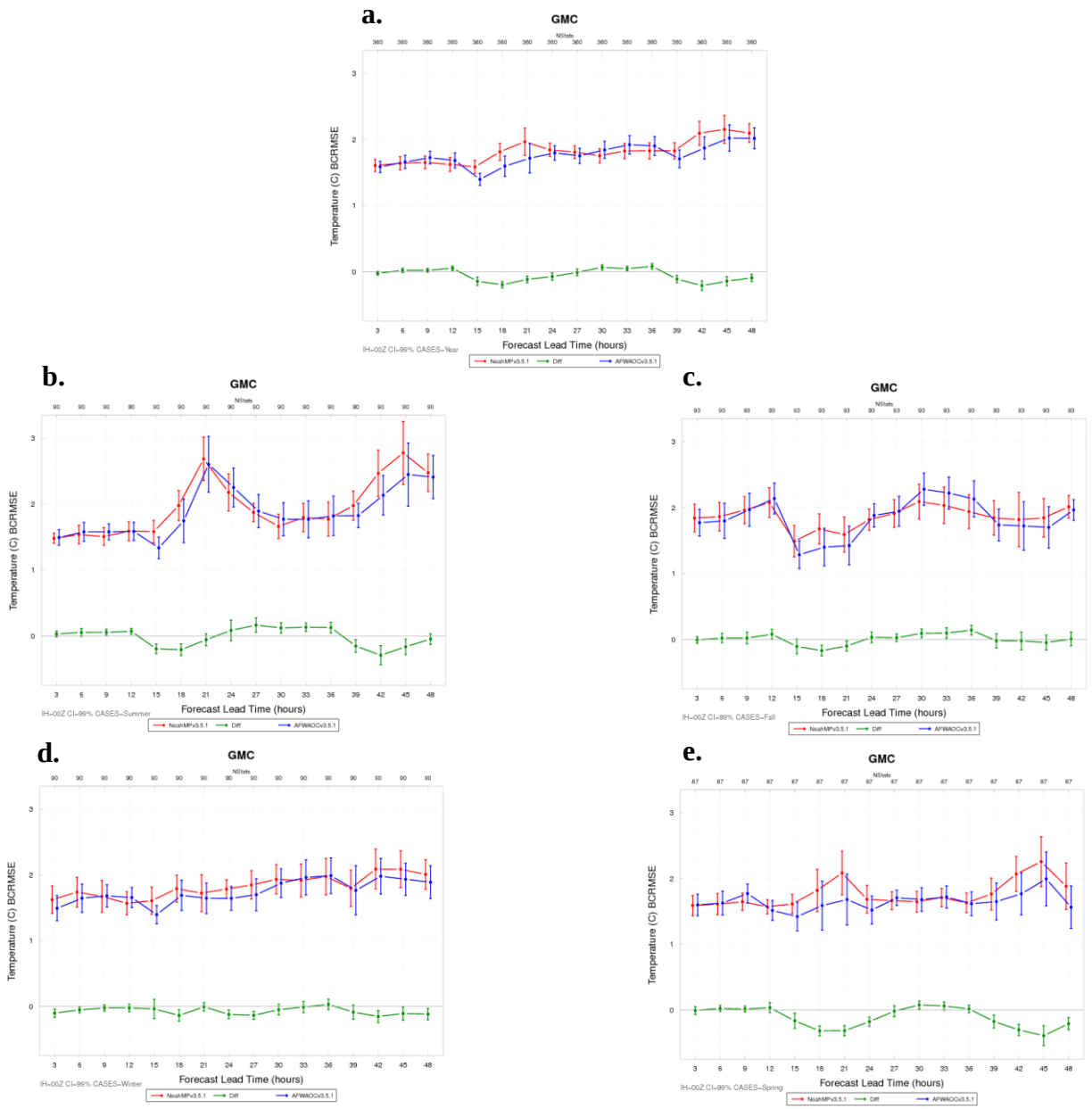


Figure E.1. Time series plot of 2 m AGL temperature (°C) for median BCRMSE for the 00 UTC initializations within the GMC region aggregated across the (a) entire year of cases, (b) summer aggregation, (c) fall aggregation, (d), winter aggregation, and (e) spring aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

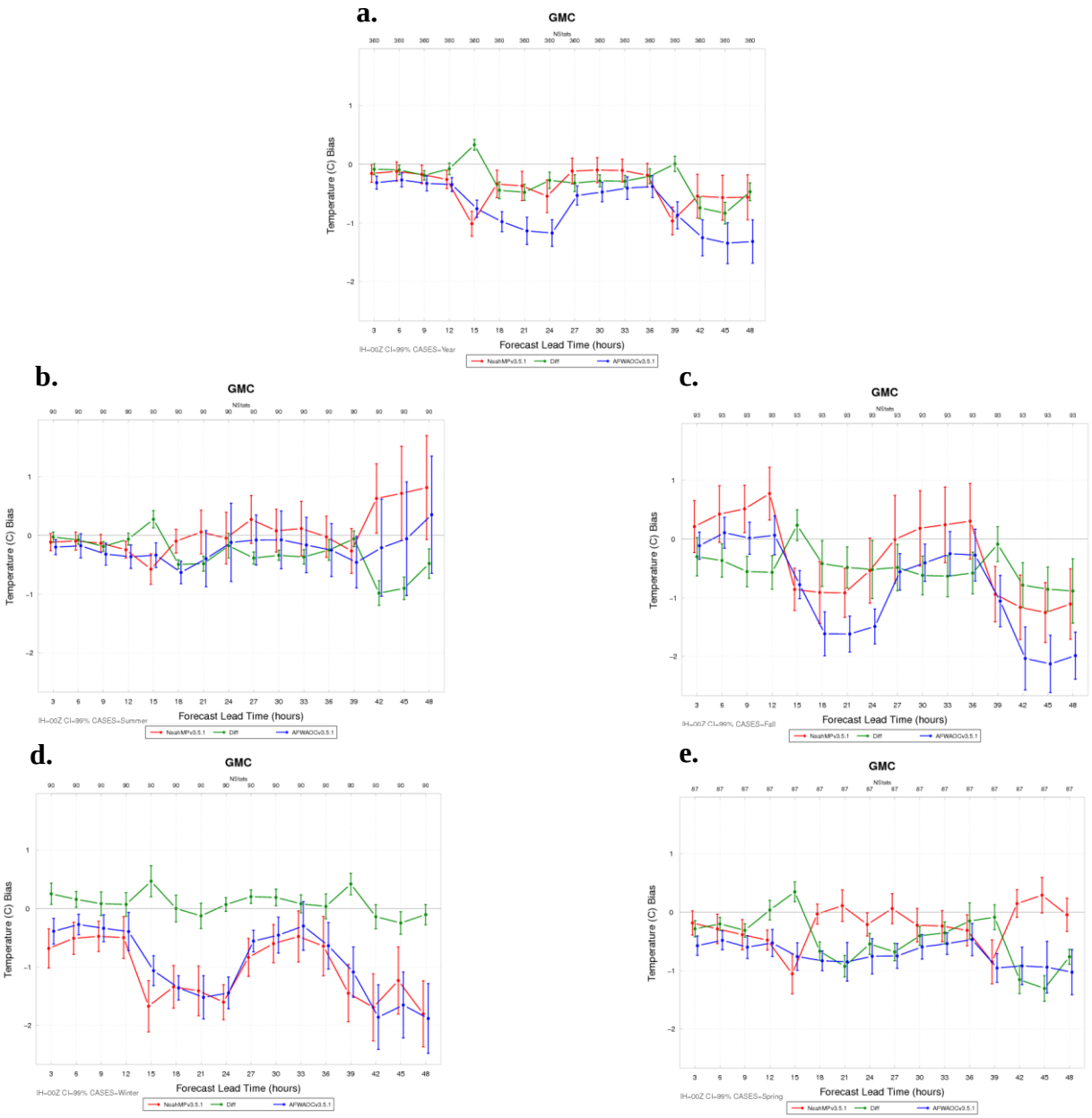


Figure E.2. Time series plot of 2 m AGL temperature (°C) for median bias for the 00 UTC initializations within the GMC region aggregated across the (a) entire year of cases, (b) summer aggregation, (c) fall aggregation, (d), winter aggregation, and (e) spring aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

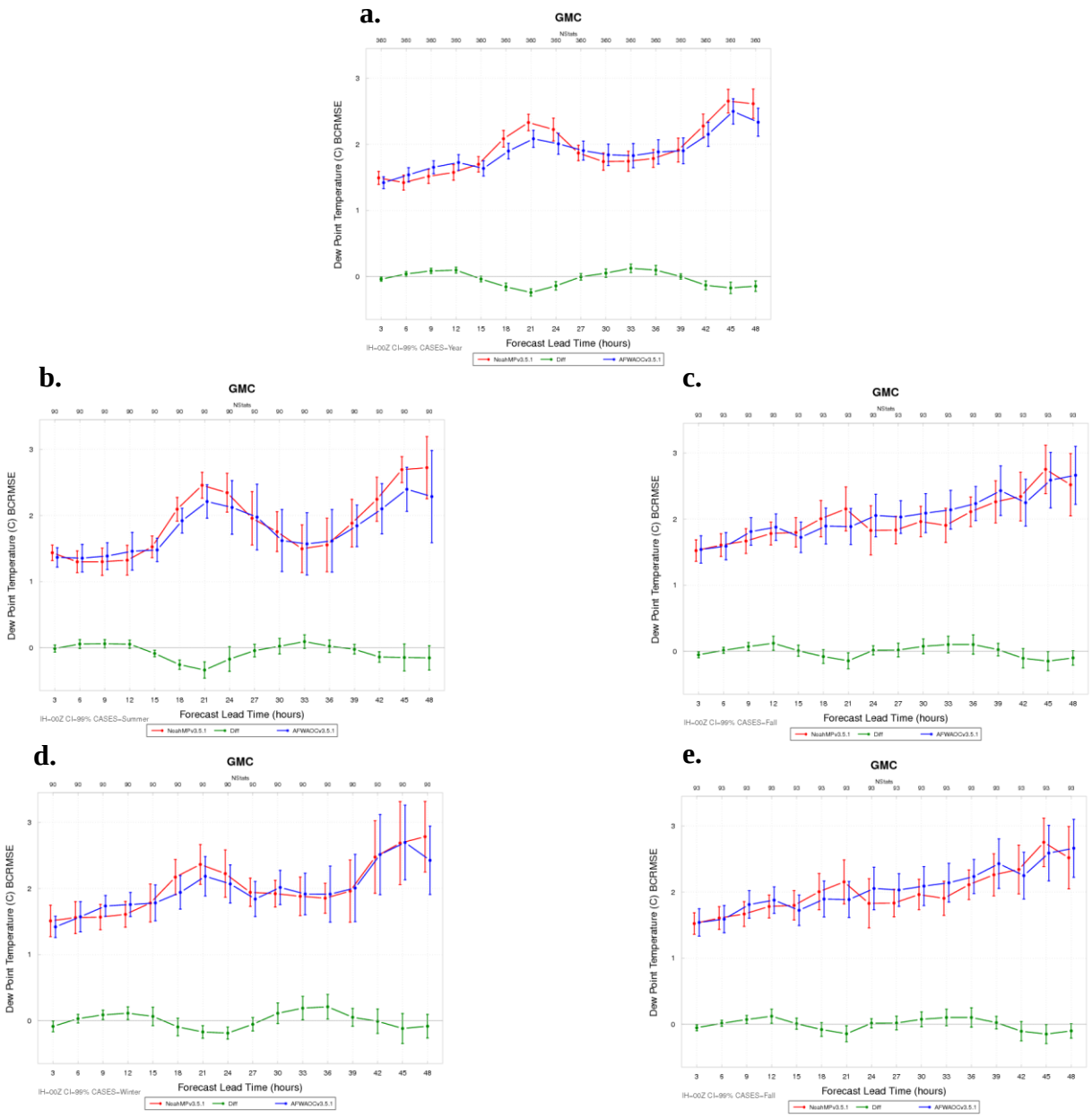


Figure E.3. Time series plot of 2 m AGL dew point temperature (°C) for median BCRMSE for the 00 UTC initializations within the GMC region aggregated across the (a) entire year of cases, (b) summer aggregation, (c) fall aggregation, (d), winter aggregation, and (e) spring aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

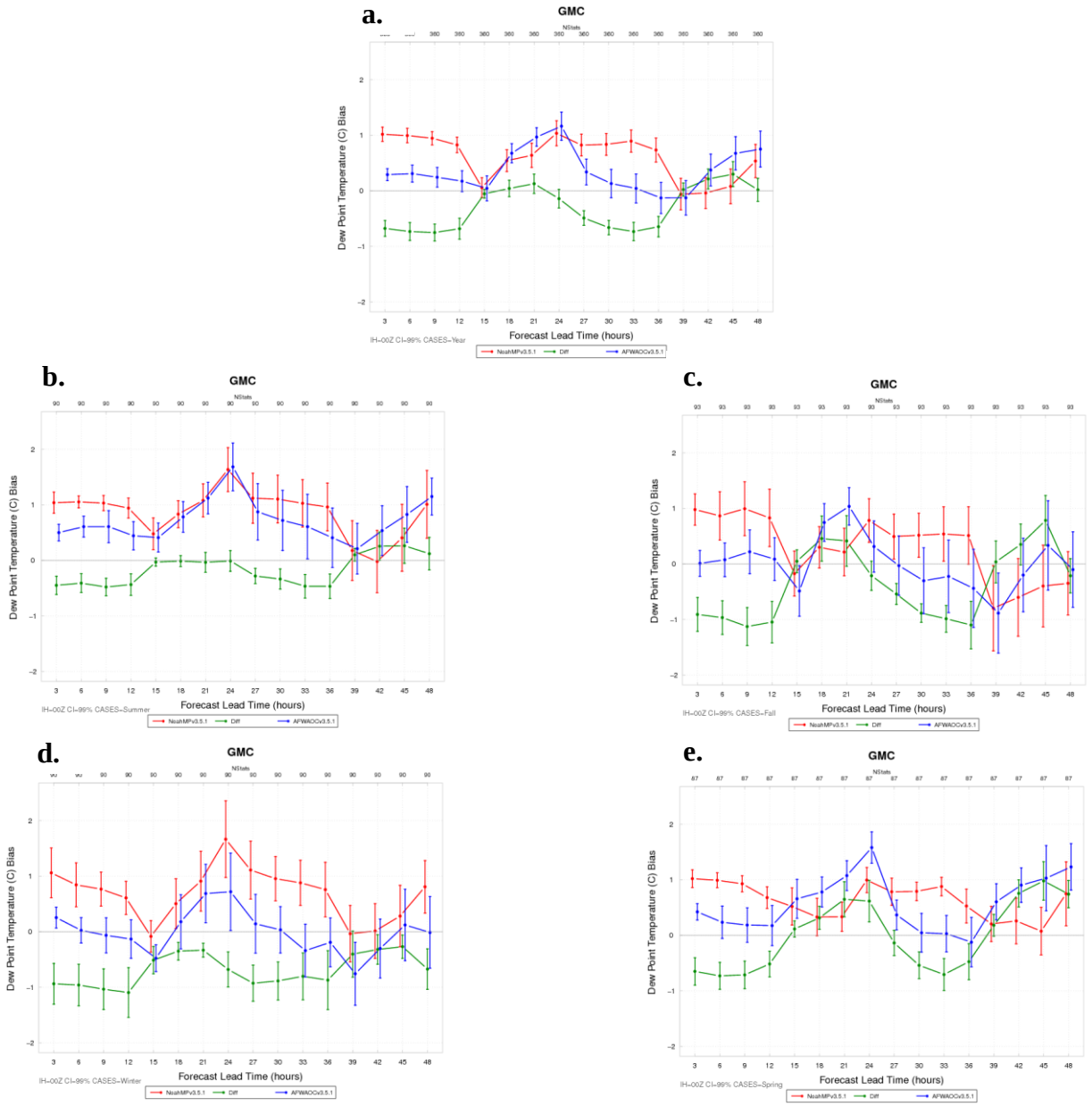


Figure E.4. Time series plot of 2 m AGL dew point temperature (°C) for median bias for the 00 UTC initializations within the GMC region aggregated across the (a) entire year of cases, (b) summer aggregation, (c) fall aggregation, (d), winter aggregation, and (e) spring aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

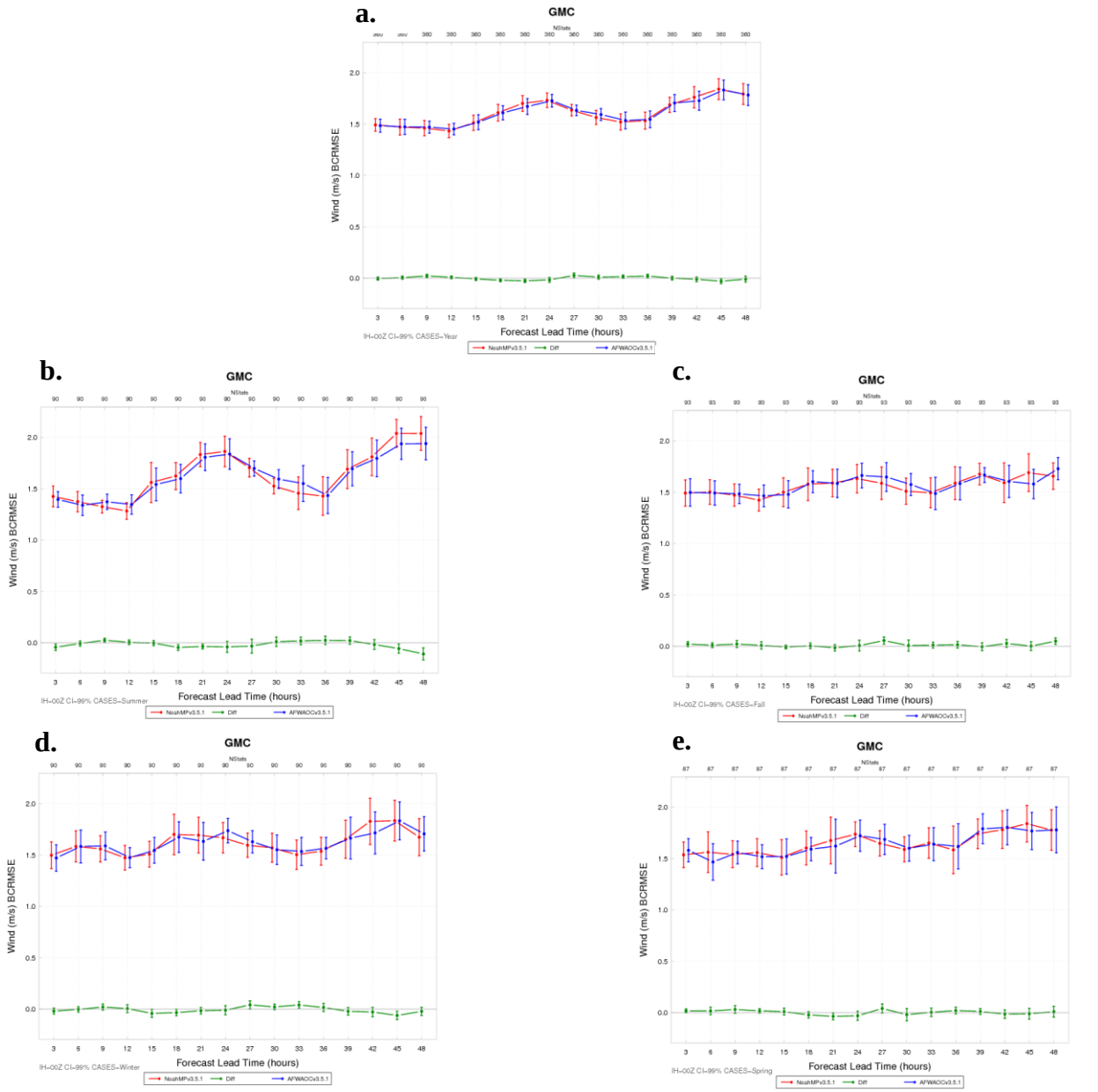


Figure E.5. Time series plot of 10 m AGL wind speed (m s^{-1}) for median BCRMSE for the 00 UTC initializations within the GMC region aggregated across the (a) entire year of cases, (b) summer aggregation, (c) fall aggregation, (d), winter aggregation, and (e) spring aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.

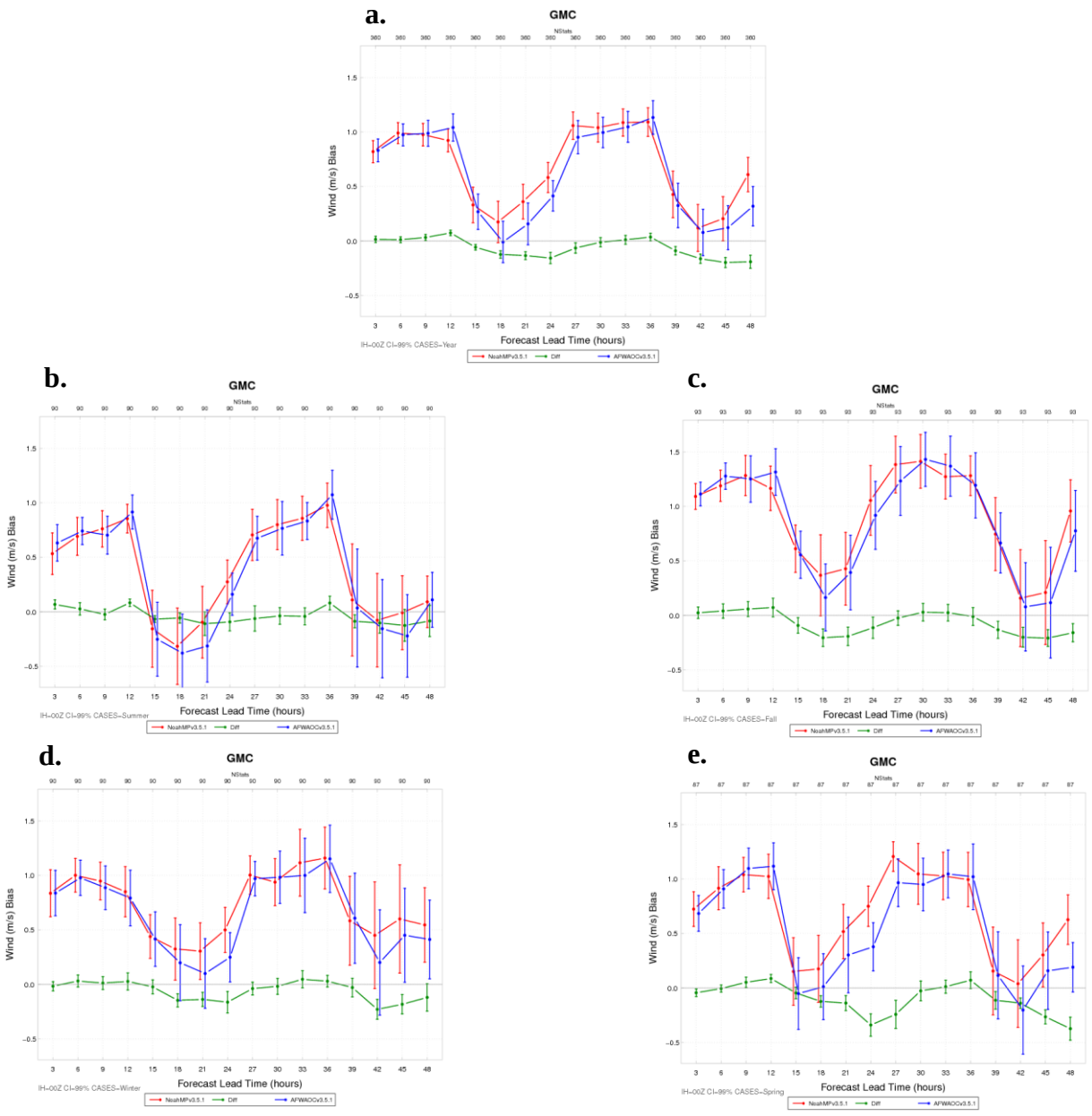


Figure E.6. Time series plot of 10 m AGL wind speed (m s^{-1}) for median bias for the 00 UTC initializations within the GMC region aggregated across the (a) entire year of cases, (b) summer aggregation, (c) fall aggregation, (d), winter aggregation, and (e) spring aggregation. AFWAOC is in blue, NoahMP in red, and the difference (AFWAOC-NoahMP) in green. The vertical bars attached to the median represent the 99% CIs.