

MEMORANDUM

Date: 30 June 2025

To: New Mexico Environmental Department, Air Quality Bureau

From: Jeremiah Johnson, Blake Himes, Jelena Popovic, and Ralph Morris

Subject: WRF 2022 4 km Meteorological Modeling for New Mexico and Model Performance Evaluation

INTRODUCTION

The New Mexico Environmental Department (NMED), Air Quality Bureau (aqb) would like to understand the causes of the elevated ozone concentrations within the New Mexico portion of the Permian Basin. To estimate the contributions of emissions from oil and gas emissions in the Permian Basin to ozone concentrations in New Mexico, Ramboll is conducting ozone source apportionment modeling. The Comprehensive Air-quality Model with extensions (CAMx) Photochemical Grid Model (PGM) is used for this purpose. The Weather Research and Forecasting (WRF) meteorological model, Advanced Research WRF (ARW) core, provides meteorological inputs required by CAMx. WRF is a mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs (Skamarock et al., 2019).

Ramboll is conducting ozone source apportionment modeling for the NMED/AQB in two Tasks. Task 1 conducts 2026 ozone source apportionment modeling of oil and gas emissions in the Permian Basin, and other major Source Sectors, using the existing National Emissions Collaborative (NEC) 2022 Emissions Modeling Platform (EMP) 2022 version 1 (2022v1) 12 km CAMx modeling platform. Task 2 develops a more refined CAMx 2022 12/4 km modeling platform with the 4 km domain focused on New Mexico and the Permian Basin. Also under Task 2 we envision conducting CAMx future year ozone source apportionment modeling. The work is being performed in two phases. Phase I is being conducted in FY2024-2025 to be completed by June 30, 2025 and consist of all of Task 1 and start developing the CAMx 2022 12/4 modeling platform for Task 2. Phase II will be performed in FY2025-2026 starting after June 30, 2025 and will perform the remainder of the Task 2 work effort including completion of the CAMx 2022 12/4 km modeling platform development and performing future year CAMx 12/4 km ozone source apportionment modeling.

This Memorandum describes the Phase I Task 2 2022 12/4 km meteorological modeling and performance assessment following the procedures set forth in the Work Plan (Ramboll, 2025a). Details on the WRF datasets, model settings, and model performance approach are discussed. A separate report presents the Phase I Task 1 CAMx 2022 base case modeling and model performance evaluation, and 2026 12 km ozone source apportionment modeling using the EMP 2022v1 platform (Ramboll, 2025b). And a separate Memorandum presents a preliminary ozone evaluation of the CAMx 2022 4 km ozone model performance using the WRF 2022 4 km meteorological inputs whose development are described in this Memorandum (Ramboll, 2025c).

WRF CONFIGURATION

Meteorological inputs for the New Mexico CAMx application were developed from a summer 2022 WRF simulation on a 4 km grid covering New Mexico, west Texas, small portions of surrounding states including Colorado, Oklahoma, Kansas, Arizona and Utah and small portion of Mexico (Figure 1). The 4 km WRF grid was nested within a North American 12 km grid equivalent to that used by EPA to develop meteorological inputs for the 2022 12US2 modeling platform (MP). The 4 km grid aligns with the 12US2 grid with sufficient extent to fully include the 4 km CAMx modeling grid while maintaining the 5-grid point buffer recommended in EPA's photochemical modeling guidance (EPA, 2018). The WRF vertical grid structure was identical to the 2022 MP (Table 1). Ramboll applied most of the same science configuration, inputs, and application methodology as in the NMED Ozone Assessment Initiative (OAI) modeling (Ramboll, 2021). We changed the source of boundary conditions and analyses fields for data assimilation to the European Center for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) as in more recent years we have found superior WRF performance using this dataset in other western U.S. applications, especially for atmospheric moisture. We also used the WRF default MODIS-based landcover datasets to define terrestrial surface conditions. Details on the WRF configuration are listed in Table 2.

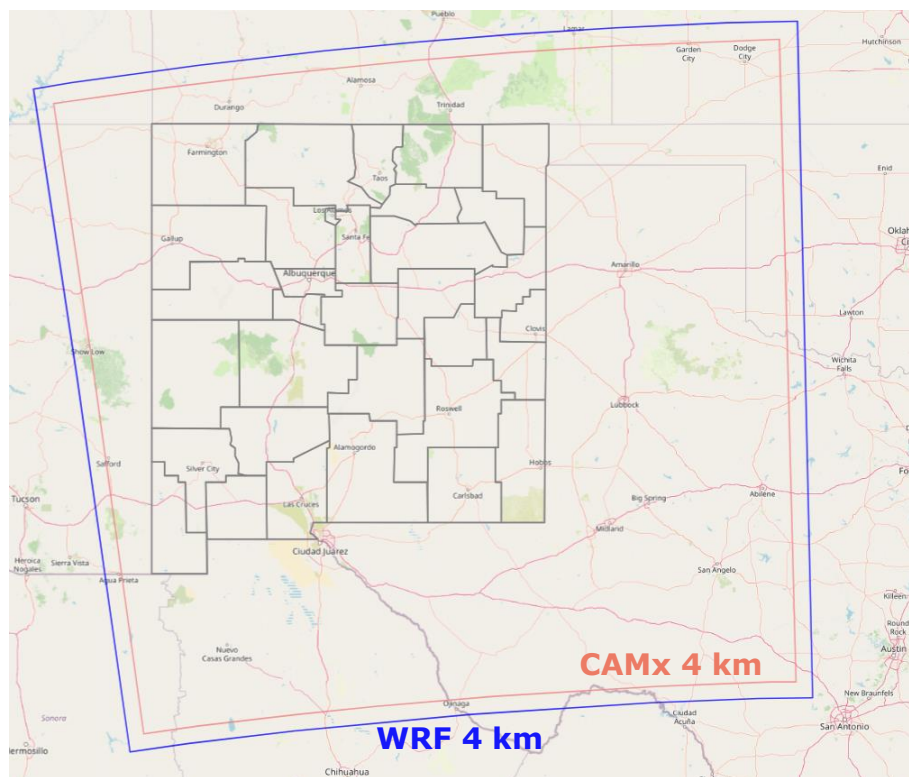


Figure 1. Extent of the 4 km WRF domain covering New Mexico, Western Texas and portions of the surrounding states. The domain meshes with the 2022 EMP 12US2 grid.

Table 1. WRF vertical grid structure for the EPA 2022 MP and the current modeling application.

Layer	eta	Pressure (mb)	Height (m)
35	0.00	50	20576
34	0.05	98	16297
33	0.10	146	13766
32	0.15	194	11961
31	0.20	243	10555
30	0.25	291	9372
29	0.30	339	8337
28	0.35	387	7416
27	0.40	435	6583
26	0.45	483	5822
25	0.50	532	5120
24	0.55	580	4467
23	0.60	628	3857
22	0.65	676	3284
21	0.70	724	2743
20	0.74	763	2331
19	0.77	792	2033
18	0.80	821	1744
17	0.82	840	1555
16	0.84	859	1370
15	0.86	878	1188
14	0.88	898	1010
13	0.90	917	835
12	0.91	927	748
11	0.92	936	662
10	0.93	946	577
9	0.94	955	493
8	0.95	965	409
7	0.96	975	326
6	0.97	984	243
5	0.98	994	162
4	0.985	999	121
3	0.990	1004	81
2	0.995	1008	40
1	0.9975	1011	20
0	1.0000	1013	0

Table 2. WRF configuration for the NMED WRF 2022 12/4 km meteorological modeling.

Model Component	Specifications and Datasets
Model Code	WRF v4.6.1
Modeling Period	May 1 – October 16, 2022
Application	Chain of 5-day segments
<u>Horizontal Grid</u>	
Map Projection	Lambert Conic Conformal
Projection Center	-97°N / 40°W
True Latitudes	33°N and 45°N
Grid Points	255 x 237
SW Corner coordinate	-1212, -1140
NE Corner coordinate	-192, -192
<u>Vertical Grid</u>	
Layers	35, up to 50 hPa (~20 km AGL)
Coordinate	Default hybrid eta coordinate
Initial/Boundary Conditions	ECMWF Reanalysis v5 (ERA5)
Nudging Analyses	ECMWF Reanalysis v5 (ERA5)
Landcover	Moderate Resolution Imaging Spectroradiometer- International Geosphere-Biosphere Programme (MODIS-IGBP)
Sea Surface Temperature	Fleet Numerical Meteorology and Oceanography Center Sea Surface Temperatures (FNMOC SST)
Lightning Data	None
<u>Physics</u>	
Short/Longwave Radiation	Rapid Radiation Transfer Model - Global (RRTMG)
Resolved Clouds	Thompson microphysics
Surface Model	Noah
Surface Layer	MM5 Similarity
Boundary Layer	Yonsei University (YSU)
Sub-grid Clouds	Multiscale Kain-Fritsch
<u>Data Assimilation/Nudging</u>	
3-D Wind	$3 \times 10^{-4} \text{ s}^{-1}$ (all layers) 12-km grid only
3-D Temperature	$3 \times 10^{-4} \text{ s}^{-1}$ (above PBL) 12-km grid only
3-D Moisture	$1 \times 10^{-5} \text{ s}^{-1}$ (above PBL) 12-km grid only
Surface Analysis Nudging	None
Observation Nudging	None
Soil Nudging	None
Lightning Assimilation	None

SUMMARY OF RESULTS

Ramboll has determined that the WRF run performed very well overall during the May-October 2022 modeling period, meeting statistical benchmarks at airport monitoring sites across the region spanning the Permian Basin, including all of New Mexico and the majority of western Texas, plus the panhandle of Oklahoma and the southernmost ~50 miles of Colorado. Meteorological monitoring sites were aggregated into the following regions for conducting a subregional model evaluation (number of included monitoring sites): New Mexico of the Permian Basin, NM_PB (5); Texas portion of the Permian Basin, TX_PB (24); Northern San Juan Basin in Colorado and southern San Juan Basin in New Mexico, CO-NSJ+NM-SSJ (7); and Southern New Mexico and El Paso in Texas, NM-S_ELPASO, (10). Statistical performance is substantially better than achieved in prior WRF modeling that supported the Ozone Assessment Initiative. The analyses described here confirm that the WRF run is applicable for CAMx photochemical modeling over the May-October 2022 ozone modeling period, for the following reasons:

- Model performance in replicating surface temperature and winds at local to regional scales is outstanding relative to other recent WRF modeling performed in the western U.S., meeting statistical benchmarks.
- The model captures observed wind speed and direction shifts and performs particularly well during stronger winds, while properly exhibiting more variable hour-by-hour stochastic variability during weaker winds. We note that there are some occasional instances of missing wind direction observations due to calm wind speeds during high ozone periods.
- Surface humidity tends to be simulated appropriately over all 6 months, generally the model replicates the day-to-day variation in humidity with the passing of synoptic scale weather systems. The effects on ozone from errors in water vapor are far less important than for similar errors in temperature.
- During ozone exceedance days when rainfall occurred within the CAMx domain, WRF mostly tracked the observed conditions and spatial patterns but tended to simulate less precipitation than reported by PRISM.

Temperature performance is critical for ozone modeling: temperature drives the diurnal evolution of mixing depth and influences temperature-sensitive emissions and chemical rates that drive ozone formation. Use of ERA5 in combination with the multi-scale Kain-Fritsch sub-grid convection parameterization led to less convective activity and overall precipitation than observed but maintained very good performance for temperature and winds. We do not see substantial and widespread errors in winds and temperature in the current modeling.

WRF indicates more spotty convection-oriented patterns, while PRISM indicates more widespread rainfall. However, the pattern in WRF may be more reasonable (though likely not accurate in time and space) as PRISM spatially interpolates between rain gauge data, which are lacking in rural areas, and thus may overstate the spatial extent of rainfall. Therefore, the PRISM precipitation interpolation scheme works better for organized synoptic weather systems than for the stochastic summer convective showers, which tend to be spotty and intermittent. This is the primary reason why the analysis of precipitation performance remains a qualitative comparison of spatial patterns and magnitudes.

Neither PRISM nor WRF can be expected to replicate the stochastic nature of actual summer convective rainfall in time and space. It is therefore perhaps more desirable for the meteorological simulation to tend toward a drier environment, as spurious convective precipitation events can drive large local errors in wind and temperature patterns. Such meteorological errors often lead to poor replication of ozone patterns during those times.

MODEL PERFORMANCE EVALUATION

Evaluation of the WRF simulation included quantitative and qualitative methods. The quantitative evaluation statistically compares WRF predictions against surface hourly meteorological observations matched by time and location. We also conducted a qualitative phenomenological evaluation that assesses whether the model replicates the meteorological conditions important for ozone formation in the area. Qualitative evaluations compare regional or "synoptic" maps of key meteorological variables, time series of modeled wind speed/direction, temperature, and humidity to observations at specific sites, and monthly rainfall spatial patterns and accumulations. The evaluation was conducted for meteorological observation sites over New Mexico portion of Permian Basin, Texas portion of Permian Basin, northern San Juan Basin in Colorado and southern San Juan Basin in New Mexico, and Southern New Mexico and El Paso in Texas to assess performance within the CAMx domain.

Approach

Quantitative model performance was evaluated using graphical and statistical analyses for surface winds, temperatures, and humidity using Ramboll's in-house set of model evaluation software. The purpose of these evaluations is to establish a first-order acceptance/rejection of the simulation based on adequate replication of the weather phenomena in the study area. Thus, this approach screens for obvious model flaws and errors. Statistical measures include mean observation and prediction, prediction signed error (bias), and prediction unsigned error (absolute or gross error). It is useful to also look at a measure of correlation, such as Pearson's correlation coefficient or the "index of agreement."

Mean observation (M_o) is calculated using values from 1 or many sites over a given period:

$$M_o = \frac{1}{IJ} \sum_{j=1}^J \sum_{i=1}^I O_j^i$$

where O_j^i is the individual observed quantity at site i and time j , and the summations are over all sites (I) and over time periods (J).

Mean Prediction (M_p) is calculated from simulation results that are interpolated to each observation site used to calculate the mean observation for a given period:

$$M_p = \frac{1}{IJ} \sum_{j=1}^J \sum_{i=1}^I P_j^i$$

where P_j^i is the individual predicted quantity at site i and time j . Note the predicted mean wind speed and mean resultant direction are derived from the vector-average of the east-west component u and north-south component v that are output by WRF.

Bias (B) is calculated as the mean signed difference in prediction-observation pairings with valid data within a given analysis region and for a given period:

$$B = \frac{1}{IJ} \sum_{j=1}^J \sum_{i=1}^I (P_j^i - O_j^i)$$

Gross Error (E) is calculated as the mean absolute difference in prediction-observation pairings with valid data within a given analysis region and for a given period:

$$E = \frac{1}{IJ} \sum_{j=1}^J \sum_{i=1}^I |P_j^i - O_j^i|$$

Note that the bias and gross error for winds are calculated from the predicted-observed residuals in speed and direction (not from vector components u and v). The direction error for a given prediction-observation pairing is limited to range from 0 to $\pm 180^\circ$.

Root Mean Square Error (RMSE), which is another form of unsigned error, is calculated as the square root of the mean squared difference in prediction-observation pairings with valid data within a given analysis region and for a given period:

$$RMSE = \left[\frac{1}{IJ} \sum_{j=1}^J \sum_{i=1}^I (P_j^i - O_j^i)^2 \right]^{\frac{1}{2}}$$

The RMSE, as the gross error, is a good overall measure of model performance. However, since large errors are weighted heavily (due to squaring), large errors in a small sub-region may produce a large RMSE for the entire region even though the errors may be small and quite acceptable elsewhere.

To put the statistical performance of a meteorological model simulation into context for air quality model applications, specific statistics are compared to performance benchmarks. The purpose of the benchmarks is to understand how good or poor the results are relative to the history of other model applications throughout the United States. Table 3 lists the meteorological model performance benchmarks that were considered in this study. The simple benchmarks (Emery et al., 2001) were developed by analyzing well-performing meteorological model results in areas of mostly flat terrain and simple meteorological conditions (e.g., stationary high pressure). The complex benchmarks (Kemball-Cook et al., 2004) were developed during the 2002 Western Regional Air Partnership (WRAP) visibility modeling and are appropriate for applications in complex terrain and more variable meteorological conditions. McNally et al. (2008) analyzed multiple annual runs that included complex terrain conditions and suggested an alternative set of benchmarks for temperature under more complex conditions. The complex benchmarks in Table 3 represent the maximum among those proposed by Kembell-Cook and McNally.

Table 3. Meteorological model performance benchmarks for simple conditions (Emery et al., 2001) and complex conditions (Kemball-Cook et al., 2004; McNally et al., 2008).

Parameter	Simple	Complex
Temperature Bias	$\leq \pm 0.5$ K	$\leq \pm 2.0$ K
Temperature Error	≤ 2.0 K	≤ 3.5 K
Humidity Bias	$\leq \pm 0.8$ g/kg	$\leq \pm 1.0$ g/kg
Humidity Error	≤ 2.0 g/kg	≤ 2.0 g/kg
Wind Speed Bias	$\leq \pm 0.5$ m/s	$\leq \pm 1.5$ m/s
Wind Speed RMSE	≤ 2.0 m/s	≤ 2.5 m/s
Wind Dir. Bias	$\leq \pm 10$ degrees	--
Wind Dir. Error	≤ 30 degrees	≤ 55 degrees

*Dashes indicate that the parameter was not addressed by the referenced study.

The WRF application was statistically evaluated against these benchmarks, including bias and gross error in temperature, wind direction and mixing ratio, and bias and RMSE in wind speed. Observations for WRF verification and evaluation were obtained from the National Climate Data Center's (NCDC) global-scale, quality-controlled DS472¹ integrated surface hourly dataset. Global hourly and synoptic observations are compiled from numerous sources into a single common text format and common data model. The DS472 database contains records of most official surface meteorological stations primarily from civil and military airports dating from 1976 to the present. NCDC airport data are consistent with the timescales of WRF output (near instantaneous at the top of each hour) and uniformly adhere to strict siting/exposure and instrumentation standards.

The WRF surface meteorological model performance metrics were compared against the simple and complex model performance goals using "soccer plots." Soccer plots present WRF statistics as symbols in X/Y space (e.g., temperature bias as X, and temperature error as Y) along with the performance benchmarks plotted as a box or "goal". The closer the symbols are to zero, the better the model performance. Statistical symbols within the goal indicate that WRF performs generally better than most WRF simulations conducted for air quality modeling applications. Statistical symbols outside the goal indicate that WRF performs generally worse. We note that the benchmarks were designed for evaluating monthly performance across multiple sites within a region. In general, bias and error statistics look better when calculated for larger data populations (Emery et al., 2001). Therefore, we expect that statistical performance evaluated over shorter periods and fewer sites will result in a wider range relative to benchmarks.

The hourly prediction and observation data that feed into the statistical calculations described above were plotted as time series. Either site-specific or site-aggregated time series can be developed. These types of plots were qualitatively reviewed to assess the ability of WRF to replicate intra-diurnal and inter-daily variations in temperature, winds, and humidity. These surface assessments focused on periods when ozone was high or exceeded the NAAQS to evaluate the extent to which meteorology was properly characterized locally and across regional transport routes.

¹ Meteorological Development Laboratory/Office of Science and Technology/National Weather Service/NOAA/U.S. Department of Commerce. 1987, updated half-yearly. TDL U.S. and Canada Surface Hourly Observations. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. <https://rda.ucar.edu/datasets/d472000/>. Accessed June 9, 2025.

A proper simulation of precipitation is also critically important for modeling ozone formation within, and regional transport into New Mexico. Plots were generated to assess precipitation patterns and rates relative to measured conditions. Oregon State University (OSU) publishes precipitation analysis fields based on observations that can be used to qualitatively evaluate the WRF precipitation fields. The Parameter-elevation Relationships on Independent Slopes Model (PRISM) is used to generate the precipitation analysis fields (Daly et al., 2008). The PRISM interpolation method develops data sets that reflect, as closely as possible, the current state of knowledge of spatial climate patterns in the United States (note that PRISM does not include any analysis fields outside of the U.S.). PRISM calculates a climate – elevation regression for each digital elevation model (DEM) grid cell, and rain gauge stations entering the regression are assigned weights based primarily on the physiographic similarity between the station and the grid cell. Factors considered are location, elevation, coastal proximity, topographic facet orientation, vertical atmospheric layer, topographic position, and orographic effectiveness of the terrain.

Spatial plots of the WRF daily and monthly precipitation fields were compared with the PRISM spatial maps in a qualitative model evaluation. We focused on WRF performance for daily summer convective precipitation events because historically, WRF has overstated convective activity, in which case the model can suppress ozone formation and improperly influence wind, temperature, and moisture patterns. However, the PRISM precipitation interpolation scheme works better for organized synoptic weather systems than for stochastic convective showers, which tend to be spotty and intermittent. This is the primary reason why the analysis of precipitation performance remains a qualitative comparison of spatial patterns and magnitudes.

WRF Evaluation Results

Surface Statistical Performance

We statistically compared WRF model performance against routine airport observations within the WRF 4-km grid modeling domain during May-October 2022. We evaluated modeled temperature, humidity, and winds against observations across the entire modeling domain and within four groups of airport monitoring sites. These groups are identified by symbol color in Figure 2 and are defined as follows:

- Entire 4 km modeling domain (all stations show in Figure 1)
- New Mexico portion of the Permian Basin (NM_PB) with 5 stations:
K4MR, KATS, KCNM, KHOB, KROW
- Texas portion of the Permian Basin (TX_PB) with 24 stations:
K6R6, KABI, KBPG, KDYS, KE11, KE38, KECU, KFST, KGDP, KGNC, KINK, KJCT, KLBB, KMAF, KMDD, KMRF, KODO, KOZA, KPEQ, KPVW, KSJT, KSNK, KSOA, KSWW
- Northern San Juan Basin in Colorado and southern San Juan Basin in New Mexico (CO-NSJ+NM-SSJ) with 7 stations:
KCEZ, KCPW, KDRO, KFMN, KGUP, KPSO, KRQE
- Southern New Mexico and El Paso in Texas (NM-S_ELPASO) with 10 stations:
KALM, KBIF, KDMN, KELP, KHMN, KLRU, KONM, KSRR, KSVC, KTCS

We present monthly statistical results as soccer goal plots showing bias and error statistics along with the simple and complex performance benchmarks defined previously. Statistics averaged over all sites within the WRF 4 km modeling domain are shown in Figure 3, while

Site Groups

- CO-NSJ+NM-SSJ Sites
- NM-S_ELPASO Sites
- NM-PB Sites
- TX-PB Sites
- Other Sites

WRFx 4km Domain
CAMx 4km Domain

0 30 60 120 180 240 Miles

Figure 3 shows excellent performance for wind speeds in all 6 months across the modeling grid, with near zero bias well within simple benchmark ($\leq \pm 0.5$ m/s) and RMSE within simple benchmark (≤ 2.0 m/s) for three months (Aug-Oct) and within complex benchmarks (≤ 2.5 m/s) for three months (May-Jul), May being exactly at the complex benchmark of 2.5 m/s. Wind direction performance shows that all 6 months also meet the simple/complex benchmark for bias ($\leq \pm 10$ degrees), and all months are within the complex error benchmark (≤ 55 degrees). Temperature bias is quite good, ranging within ± 1.1 K, with slight over predictions in most months and minimal under prediction in May. Temperature error is outstanding, remaining within the simple benchmark (≤ 2.0 K) in all 6 months. Water vapor mixing ratio performance shows a negative bias for the first 3 months and positive bias for the last 3 months, and all 6 months meet the simple benchmark for bias ($\leq \pm 0.8$ g/kg). All months are within the simple/complex benchmark for error (≤ 2.0 g/kg).

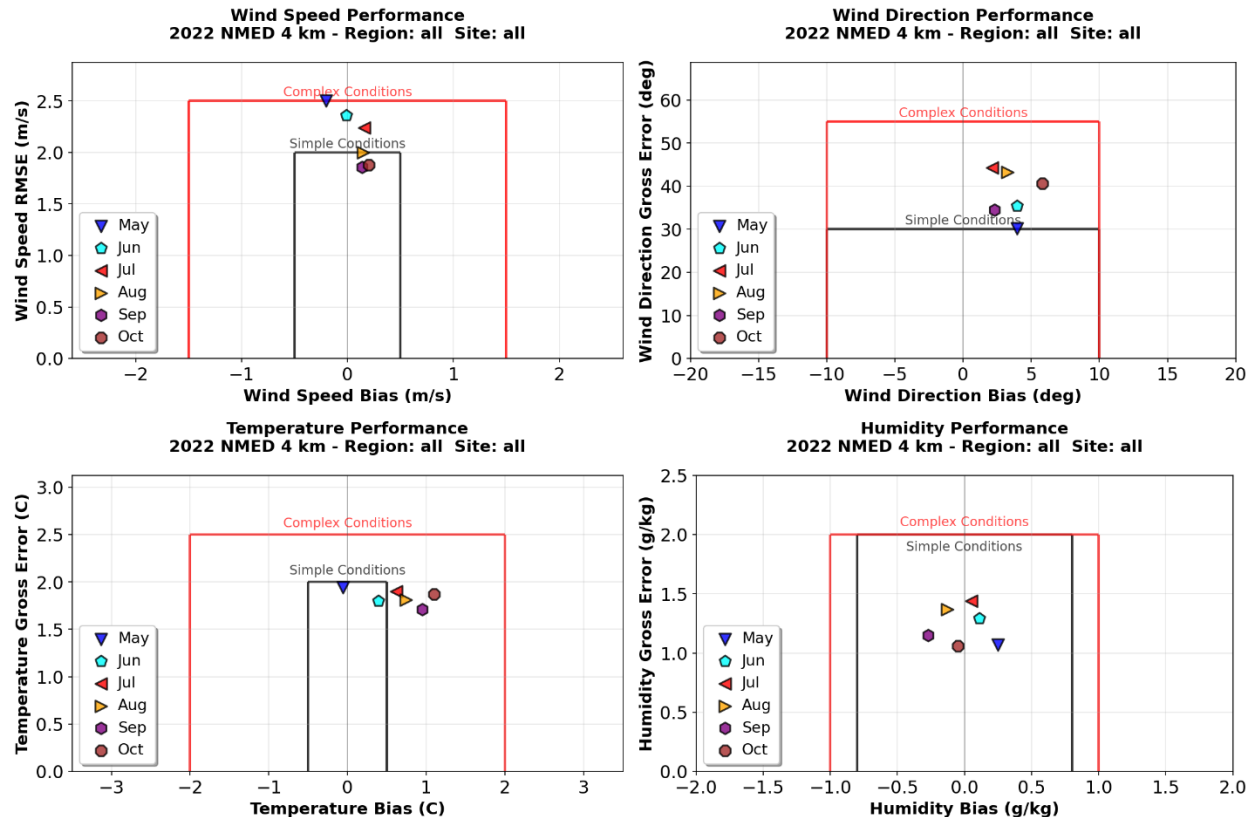


Figure 3. Soccer plots comparing WRF model performance statistical metrics against simple and complex benchmarks averaged over all 75 sites within the CAMx modeling domain: 10-m wind speed (top left), 10-m wind direction (top right), 2-m temperature (bottom left) and 2-m water vapor mixing ratio (bottom right). Symbols refer to each of the six simulation months in 2022.

For the New Mexico portion of the Permian Basin, Figure 4 shows excellent performance for wind speeds across all 6 months, with bias near zero, well within the simple benchmark ($\leq \pm 0.5$ m/s) and RMSE within the simple benchmark (≤ 2.0 m/s) for three months (Aug-Oct) and within the complex benchmark (≤ 2.5 m/s) for the other three months (May-Jul). Wind direction performance shows a positive bias, with 5 months meeting the simple/complex benchmark ($\leq \pm 10$ degrees) and October slightly exceeding it at 11.6 degrees, Wind direction error is within the complex benchmark (≤ 55 degrees) for all months. Temperature bias is within ± 1.0 K, with three months meeting the simple benchmark ($\leq \pm 0.5$ K), being positive in three months and negative in the other three. Temperature error is within the simple benchmark (≤ 2.0 K) for all months except for May, when it falls within the complex benchmark (≤ 2.5 K). Water vapor mixing ratio performance shows a negative bias for all 6 months with 4 months meeting the simple benchmark ($\leq \pm 0.8$ g/kg), one meeting the complex benchmark ($\leq \pm 1.0$ g/kg), and one outside the complex benchmark. All months are within the simple/complex benchmark for error (≤ 2.0 g/kg).

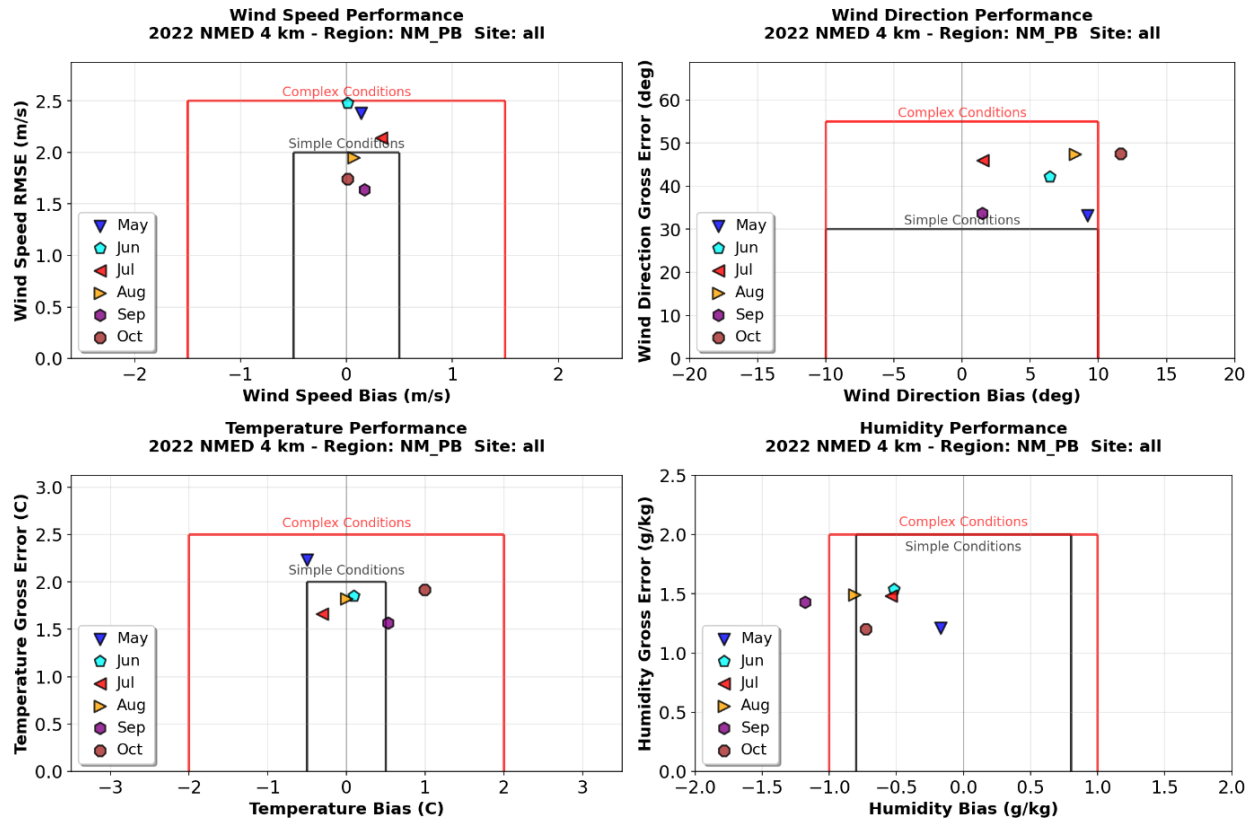


Figure 4. Soccer plots comparing WRF model performance statistical metrics against simple and complex benchmarks averaged over the 5 NM_PB sites: 10-m wind speed (top left), 10-m wind direction (top right), 2-m temperature (bottom left) and 2-m water vapor mixing ratio (bottom right). Symbols refer to each of the six simulation months in 2022.

For the Texas portion of the Permian Basin, Figure 5 also shows excellent performance for wind speeds across all 6 months, with bias well within the simple benchmark ($\leq \pm 0.5$ m/s) and RMSE within the simple benchmark (≤ 2.0 m/s) for three months (Aug-Oct) and within the complex benchmarks (≤ 2.5 m/s) for the other three months (May-Jul). Wind direction performance show slight overprediction with all 6 months meeting the simple/complex bias benchmark ($\leq \pm 10$ degrees). Wind direction error is within the simple error benchmark (≤ 30 degrees), for three months, and slightly above it, but still within the complex benchmark (≤ 55 degrees) for the other three. Temperature is overpredicted for two months, marginally exceeding the simple bias benchmark ($\leq \pm 0.5$ K), and slightly underpredicted for four months, remaining within the simple benchmark. Temperature error stays within the simple benchmark (≤ 2.0 K) throughout the six months. Water vapor mixing ratio shows a positive bias for 5 months and a negative bias for one, with five months meeting the simple benchmark for ($\leq \pm 0.8$ g/kg), July meeting the complex benchmark ($\leq \pm 1.0$ g/kg). Error is within the simple/complex benchmark (≤ 2.0 g/kg) for all months.

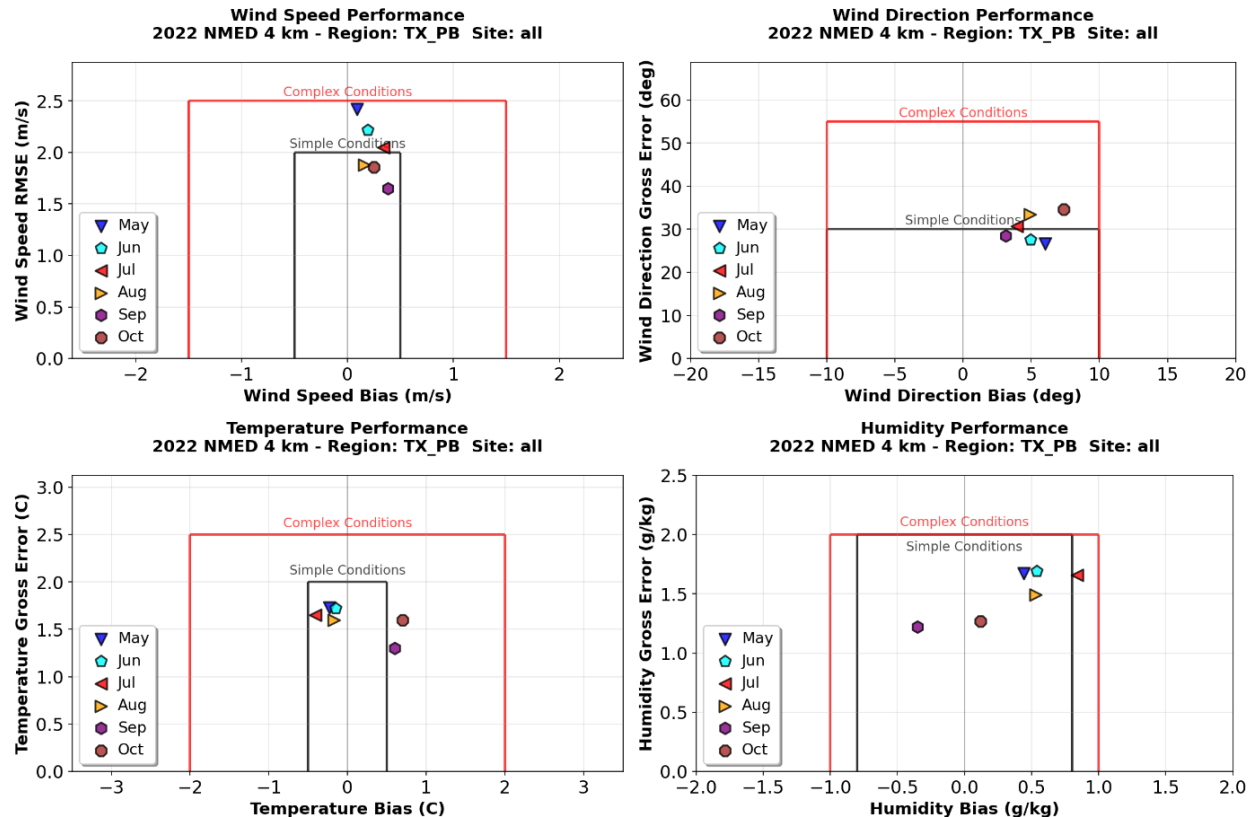


Figure 5. Soccer plots comparing WRF model performance statistical metrics against simple and complex benchmarks averaged over the 24 TX_PB sites: 10-m wind speed (top left), 10-m wind direction (top right), 2-m temperature (bottom left) and 2-m water vapor mixing ratio (bottom right). Symbols refer to each of the six simulation months in 2022.

Model performance for Northern San Juan Basin in Colorado and southern San Juan Basin in New Mexico is shown in Figure 6. Wind speed bias is well within the simple benchmark ($\leq \pm 0.5$ m/s) for 5 months and within the complex benchmark ($\leq \pm 1.5$ m/s) for May. Wind speed RMSE meets the simple benchmark (≤ 2.0 m/s) for four months (July-October) and exceeds the complex benchmark (≤ 2.5 m/s) in May and June. Wind direction performance in this region is lower than in other regions, showing consistent overpredictions. July and August meet the simple/complex bias benchmark ($\leq \pm 10$ degrees) while the remaining 4 months exceed it. Wind direction error is within the complex error benchmark (≤ 55 degrees) for two months and above it for the other four. Temperature is over predicted for five months and under predicted in May, with bias remaining within the complex benchmark ($\leq \pm 2.0$ K) across all months. Temperature error meets the complex benchmark (≤ 2.5 K) for five months and exceeds it in May. The water vapor mixing ratio performance shows a negative bias for 4 months, no bias in June and slightly positive bias in May, with all 6 months meeting the simple benchmark ($\leq \pm 0.8$ g/kg). The error for water vapor mixing ratio is within the simple/complex benchmark (≤ 2.0 g/kg) for all months.

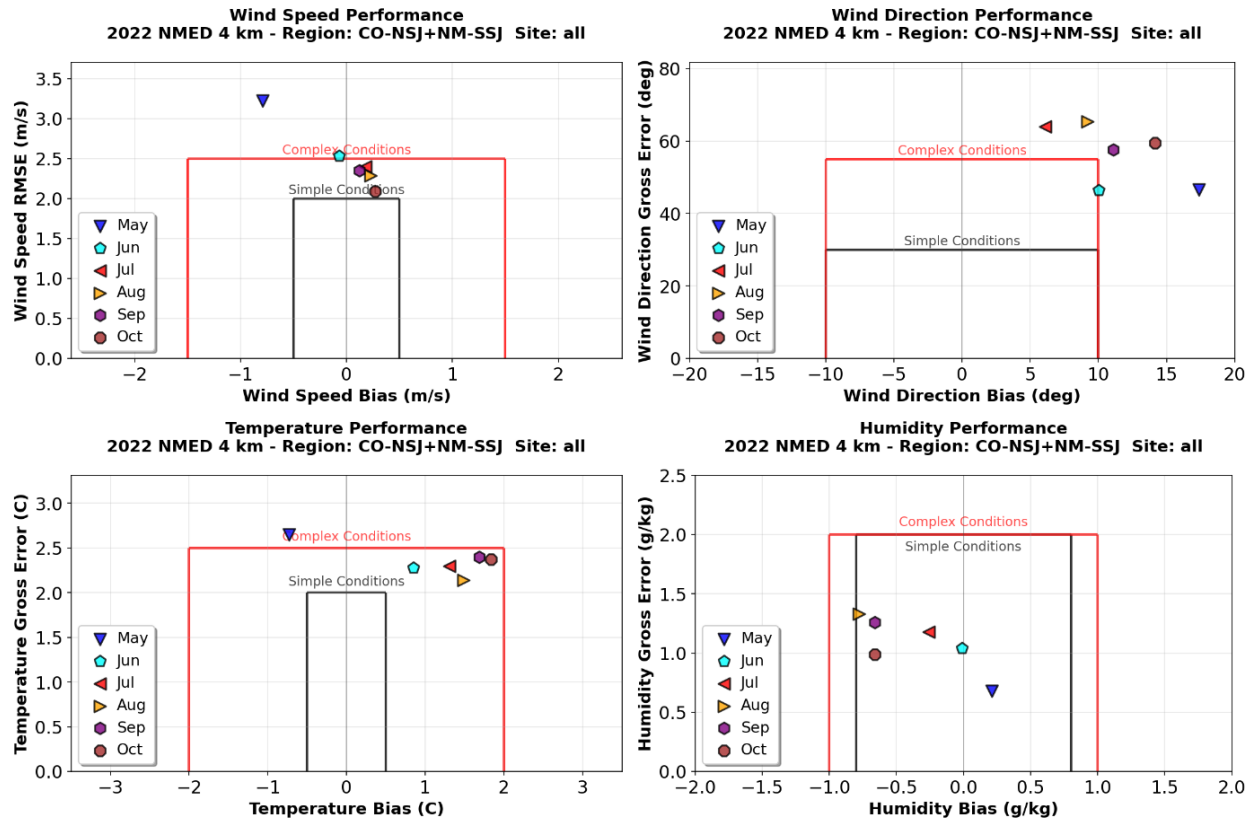


Figure 6. Soccer plots comparing WRF model performance statistical metrics against simple and complex benchmarks averaged over the 7 CO-NSJ+NM-SSJ sites: 10-m wind speed (top left), 10-m wind direction (top right), 2-m temperature (bottom left) and 2-m water vapor mixing ratio (bottom right). Symbols refer to each of the six simulation months in 2022.

Figure 7 shows model performance for Southern New Mexico and El Paso in Texas. Wind speed bias is well within the simple benchmark ($\leq \pm 0.5$ m/s) for 5 months and just slightly above the simple benchmark in June (0.52 m/s). Wind speed RMSE is within the simple benchmark (≤ 2.0 m/s) for two months, within the complex benchmark (≤ 2.5 m/s) for three months, and exceeds the complex benchmark in one month. Wind direction shows overprediction for all 6 months, with 5 months meeting the simple/complex bias benchmark ($\leq \pm 10$ degrees) while August slightly exceeds this threshold. Wind direction error is within the complex error benchmark (≤ 55 degrees) for three months, and above it for the other three. Temperature is overpredicted for five months and underpredicted in May, remaining within the complex bias benchmark ($\leq \pm 2.0$ K) for five months and within the simple benchmark ($\leq \pm 0.5$ K) in June. Temperature error meets the simple benchmark for four months and the complex benchmark for the other two. Water vapor mixing ratio shows a negative bias for 5 months, and a slightly positive bias in May. Two months meet the simple benchmark ($\leq \pm 0.8$ g/kg), and four meet the complex benchmark ($\leq \pm 1.0$ g/kg). All months fall within the simple/complex benchmark for error (≤ 2.0 g/kg).

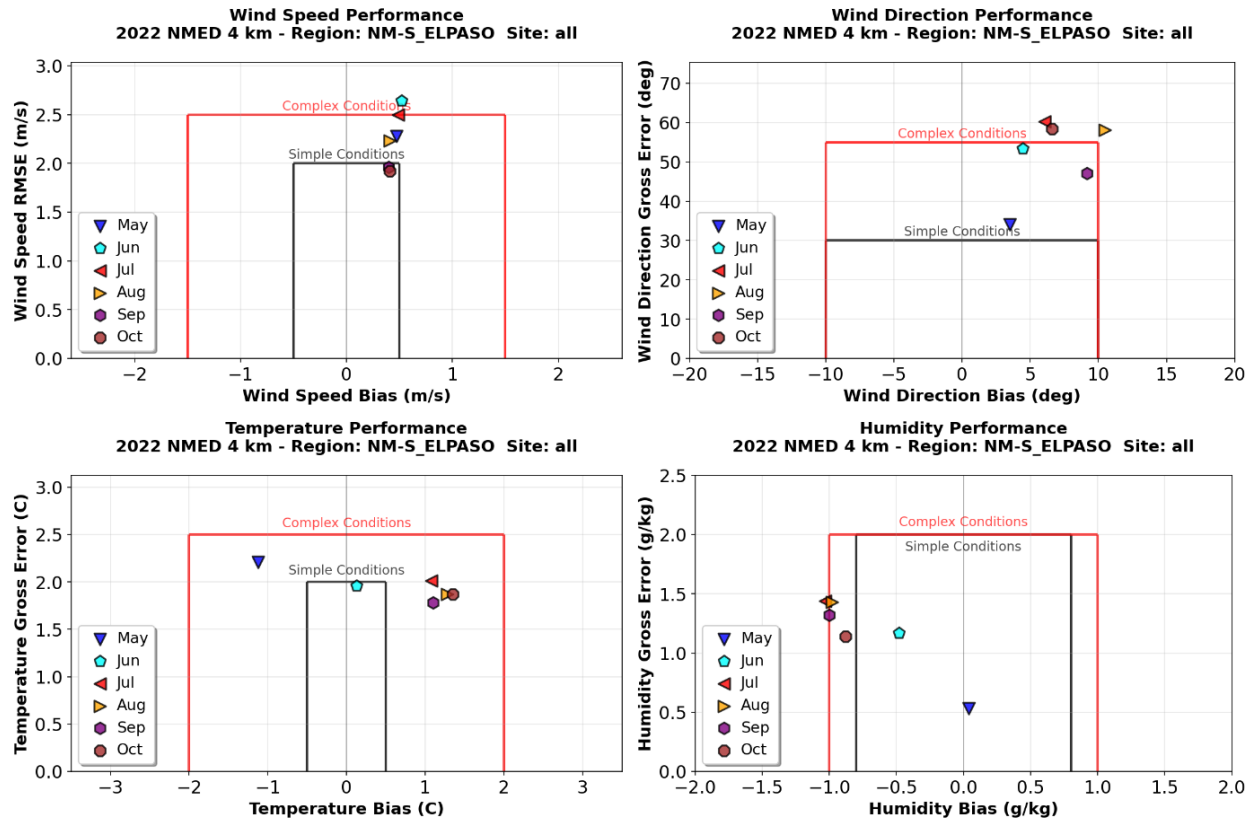


Figure 7. Soccer plots comparing WRF model performance statistical metrics against simple and complex benchmarks averaged over the 10 NM-S_ELPASO sites: 10-m wind speed (top left), 10-m wind direction (top right), 2-m temperature (bottom left) and 2-m water vapor mixing ratio (bottom right). Symbols refer to each of the six simulation months in 2022.

Time Series

Soccer plots shown above represent statistics averaged over all stations within the modeling domain or over a subset of stations within the four defined regions. Soccer plots showing the monthly average statistics for each individual station were also created and analyzed and are being provided in a package accompanying this report. To compare WRF model results with observations on a more granular level at the specific location, we present the hourly time series for one station per region.

Hourly time series plots of the four meteorological parameters in July and August are presented at the stations depicting representative examples for the region, showing statistics that is similar to the average statistics for the entire region.

The time series of predicted and observed hourly surface meteorological variables at the Cavern City Airport in Carlsbad, New Mexico (KCNM), shown in Figure 8 and Figure 9, are separated into 4 sets of plots for July and August 2022, respectively. This site is representative of the New Mexico portion of the Permian Basin. High ozone periods each month are highlighted in yellow.

Overall, the model-observation agreement is quite good for wind, temperature and humidity on all high ozone days. As seen in the statistical results presented in the soccer plots as well as in these time series plots, WRF characterizes observed variation in wind speed rather well over the entire period. The model captures observed wind direction shifts and performs particularly well during periods of stronger winds, while properly exhibiting more variable hour-by-hour stochastic variability during weaker winds. We note that missing wind direction observations due to calm wind speeds are frequent during high ozone periods. Temperatures are generally comparable, although WRF slightly underestimates the highest temperatures and overestimates the lowest. Humidity is also comparable between model and observations, with some days in a model being slightly drier than observed and others slightly wetter, with average slightly negative bias overall.

The time series at the Winkler County Airport (KINK) in Texas just south of the New Mexico southeast border, shown in Figure 10 and Figure 11, are separated into 4 sets of plots for July and August 2022, respectively. This site is representative of the Texas portion of the Permian Basin.

Similar to the KCNM station, KINK station also shows quite good model-observation agreement for all three examined parameters on all high ozone days. WRF captures the observed variation in wind speed very well over the entire period, with occasional mismatch on certain days when the model misses large intermittent spikes in observed wind speed. Modeled wind direction aligns well with observations, particularly during periods of stronger winds, but less so during calm or weak winds when capturing the observed hour-by-hour stochastic variability is more challenging. Modeled daily temperature patterns at this station closely match observations, except during nighttime hours, when minimum temperatures are slightly overestimated by WRF. The model also reproduces the daily humidity pattern reasonably well, though it occasionally simulates greater diurnal fluctuations than observed.

The time series at Durango-La Plato County Airport (KDRO), a site representative of the northern San Juan Basin in Colorado and the southern San Juan Basin in New Mexico, are shown in Figure 12 and Figure 13 for July and August 2022, respectively. The time series at Deming Municipal Airport (KDMN) in Luna County just west of Dona Ana County, a site representative of the southern New Mexico and El Paso in Texas, are shown in Figure 14 and Figure 15 for the same months.

Model-observation agreement on all high-ozone days is also quite good at both sites. Examination of the time series plots for both sites reveal similar conclusions to those described for the previous two. In general, WRF captures wind speed variations well, with occasional underestimation of high observed wind speeds or overestimation of low wind speeds. Wind directions are also well resolved, typically better during stronger winds, and less so during weaker, more variable conditions. Diurnal variations in temperature and humidity are characterized well. However, modeled temperature extremes are generally underestimated with WRF, resulting in less pronounced diurnal temperature variation. Conversely, modeled humidity extremes are generally overestimated, resulting in a more pronounced diurnal humidity variation.

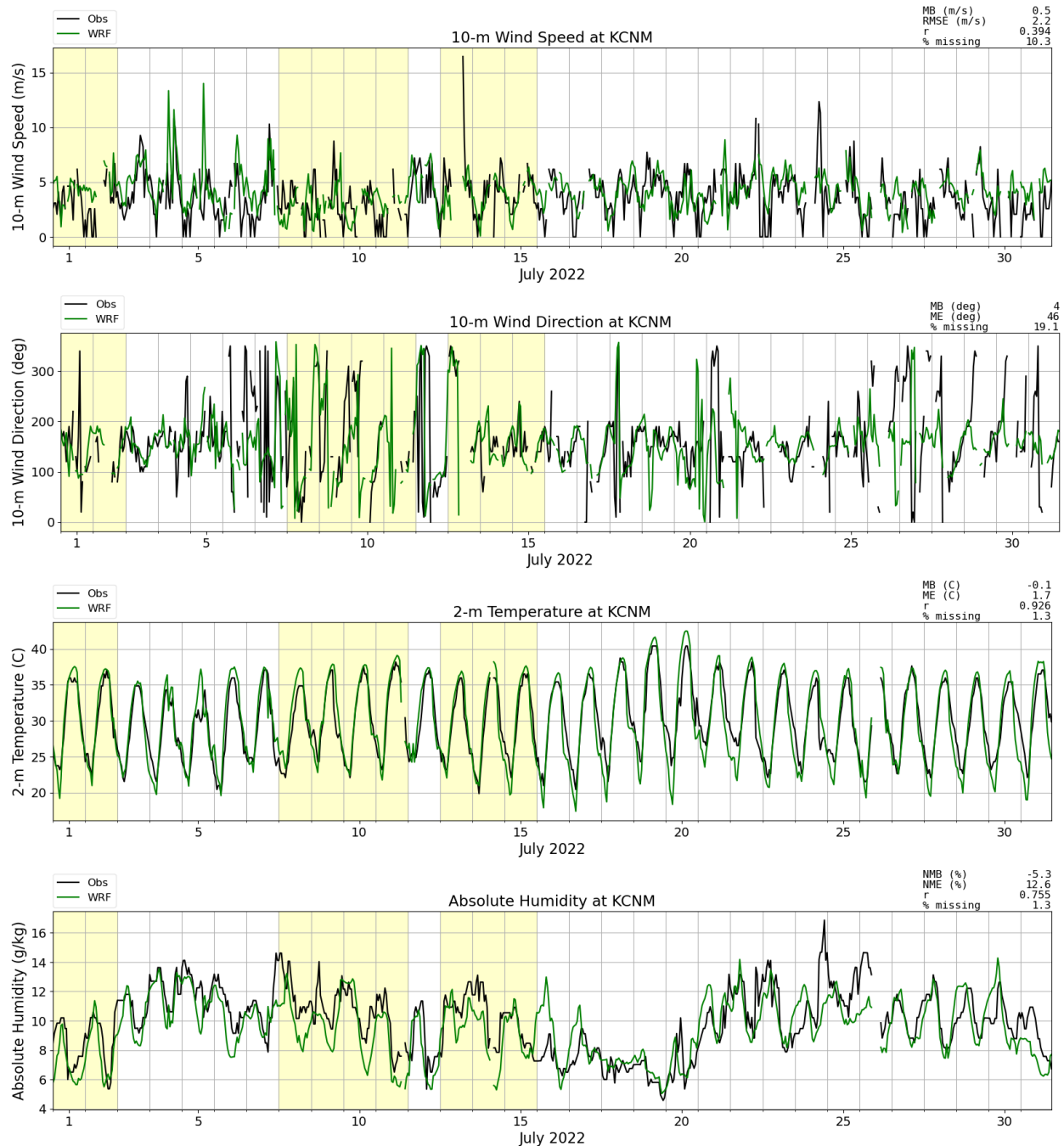


Figure 8. Time series at KCNM of observed (black line) and WRF (green) 10-m wind speed (top panel), 10-m wind direction (second panel), 2-m temperature (third panel), and 2-m water vapor mixing ratio (bottom panel) during July 2022. High ozone periods are noted by yellow shading.

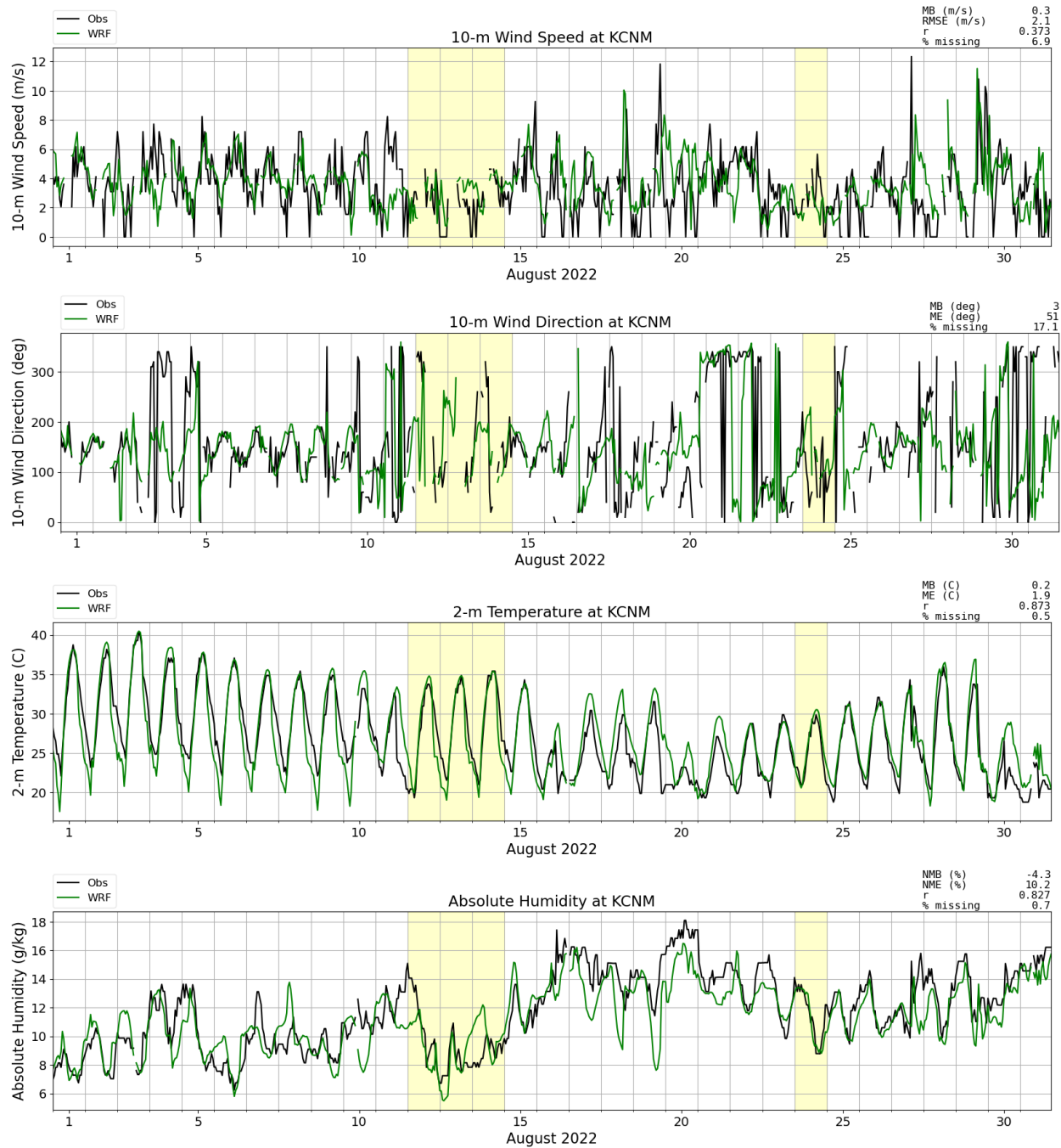


Figure 9. Same as Figure 8 for KCNM but during August 2022.

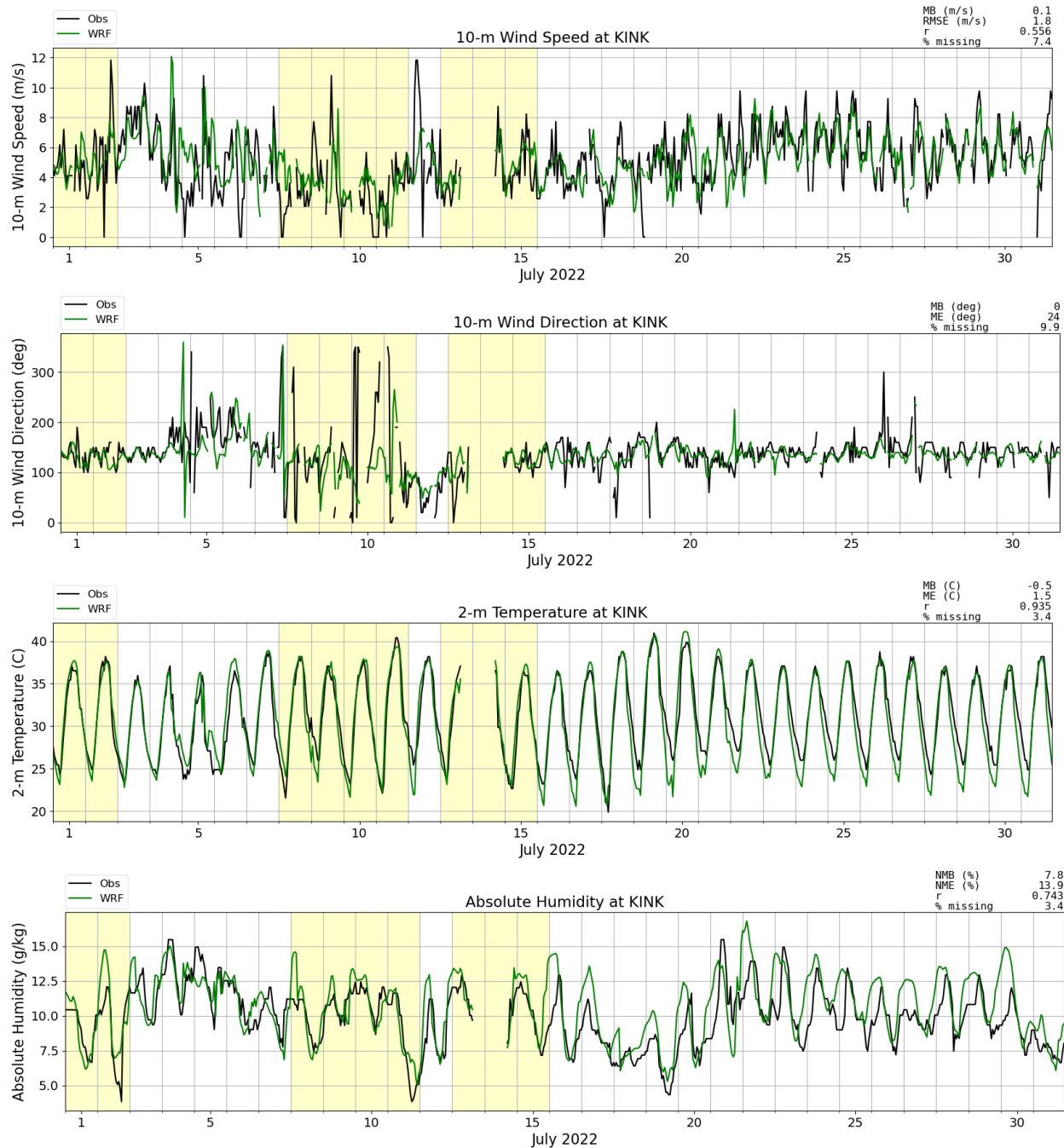


Figure 10. Time series at KINK of observed (black line) and WRF (green) 10-m wind speed (top panel), 10-m wind direction (second panel), 2-m temperature (third panel), and 2-m water vapor mixing ratio (bottom panel) during July 2022. High ozone periods are noted by yellow shading.

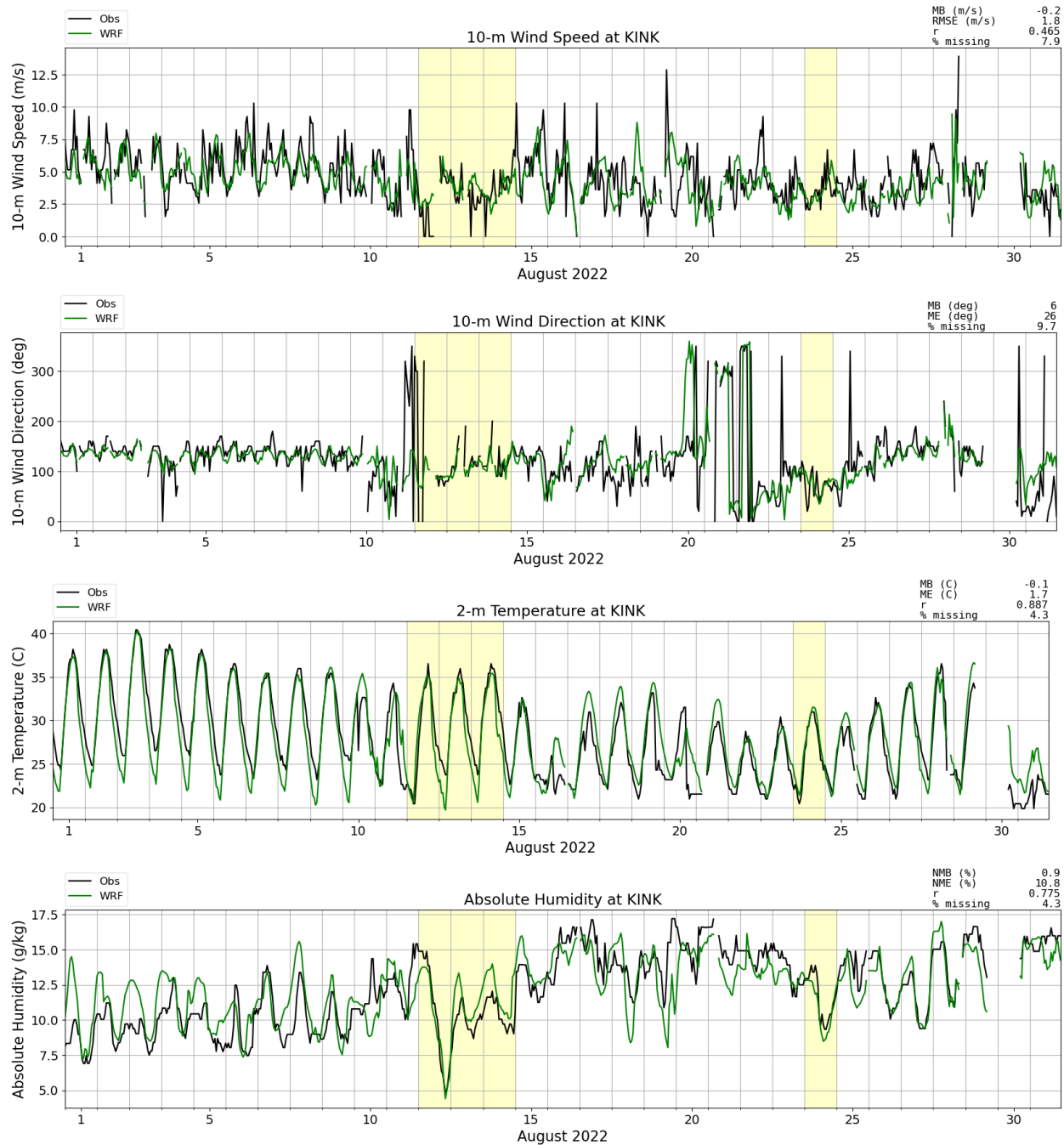


Figure 11. Same as Figure 10 for KINK but during August 2022.

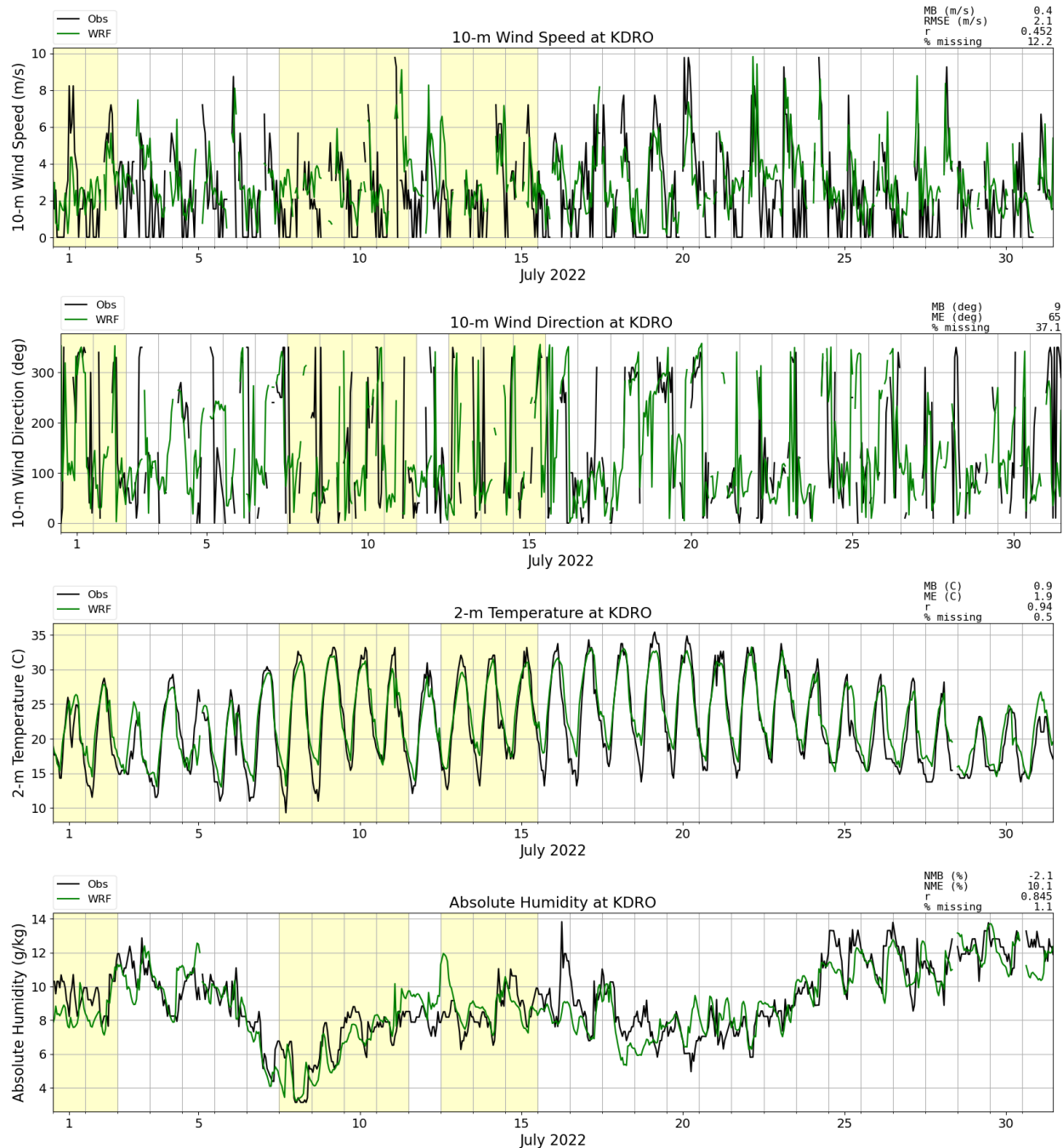


Figure 12. Time series at KDRO of observed (black line) and WRF (green) 10-m wind speed (top panel), 10-m wind direction (second panel), 2-m temperature (third panel), and 2-m water vapor mixing ratio (bottom panel) during July 2022. High ozone periods are noted by yellow shading.

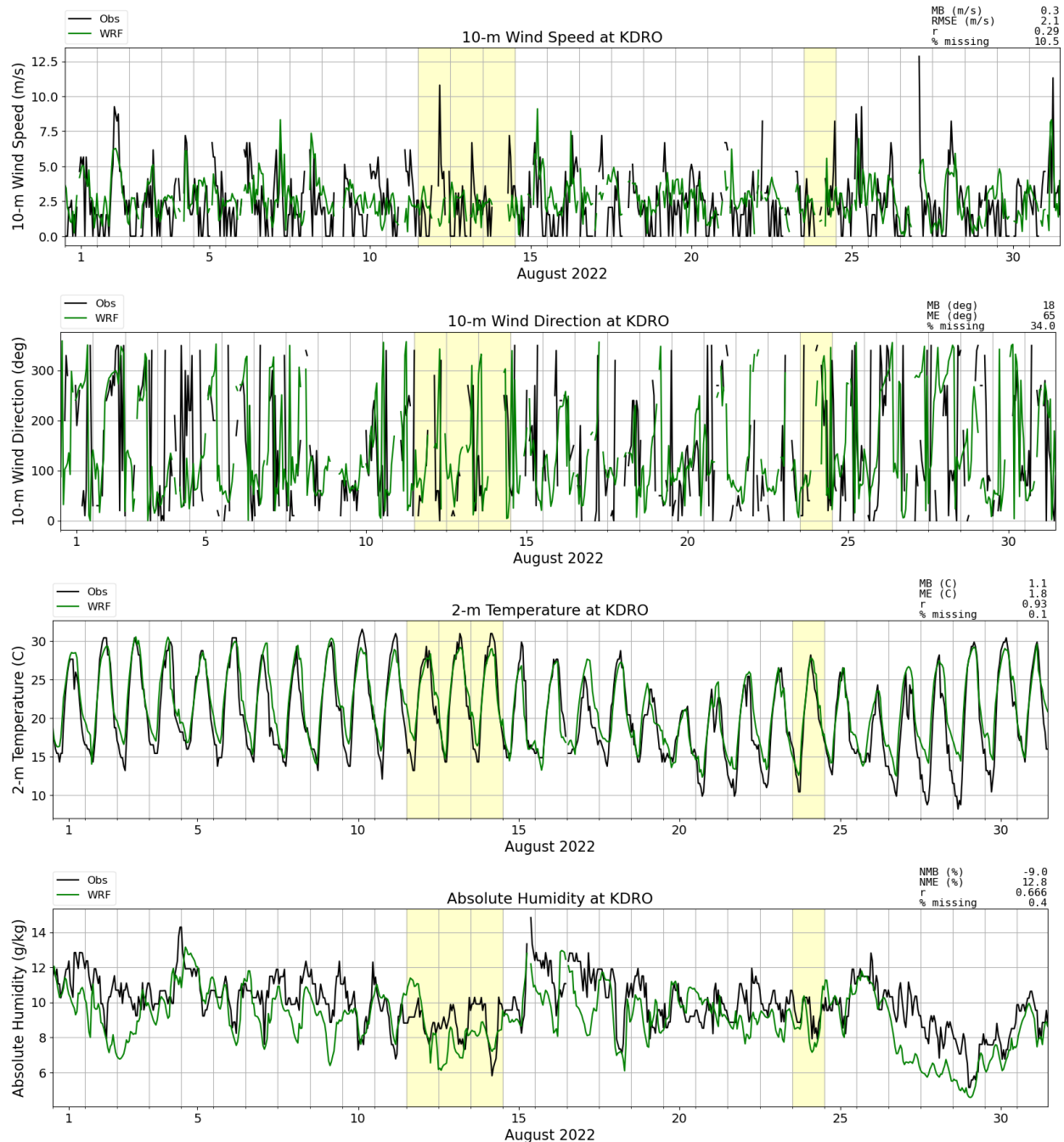


Figure 13. Same as Figure 12 for KDRO but during August 2022.

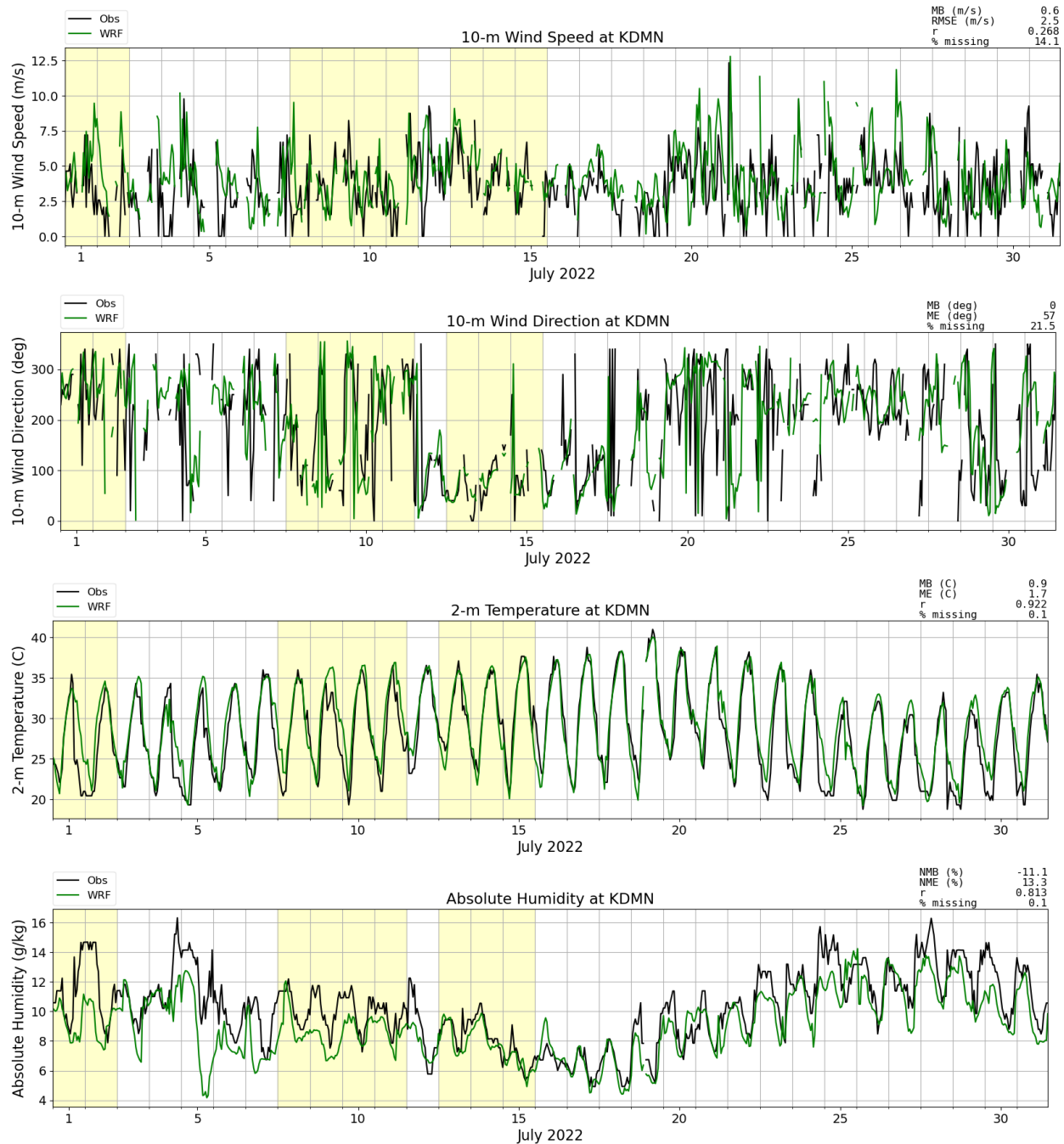


Figure 14. Time series at KDMN of observed (black line) and WRF (green) 10-m wind speed (top panel), 10-m wind direction (second panel), 2-m temperature (third panel), and 2-m water vapor mixing ratio (bottom panel) during July 2022. High ozone periods are noted by yellow shading.

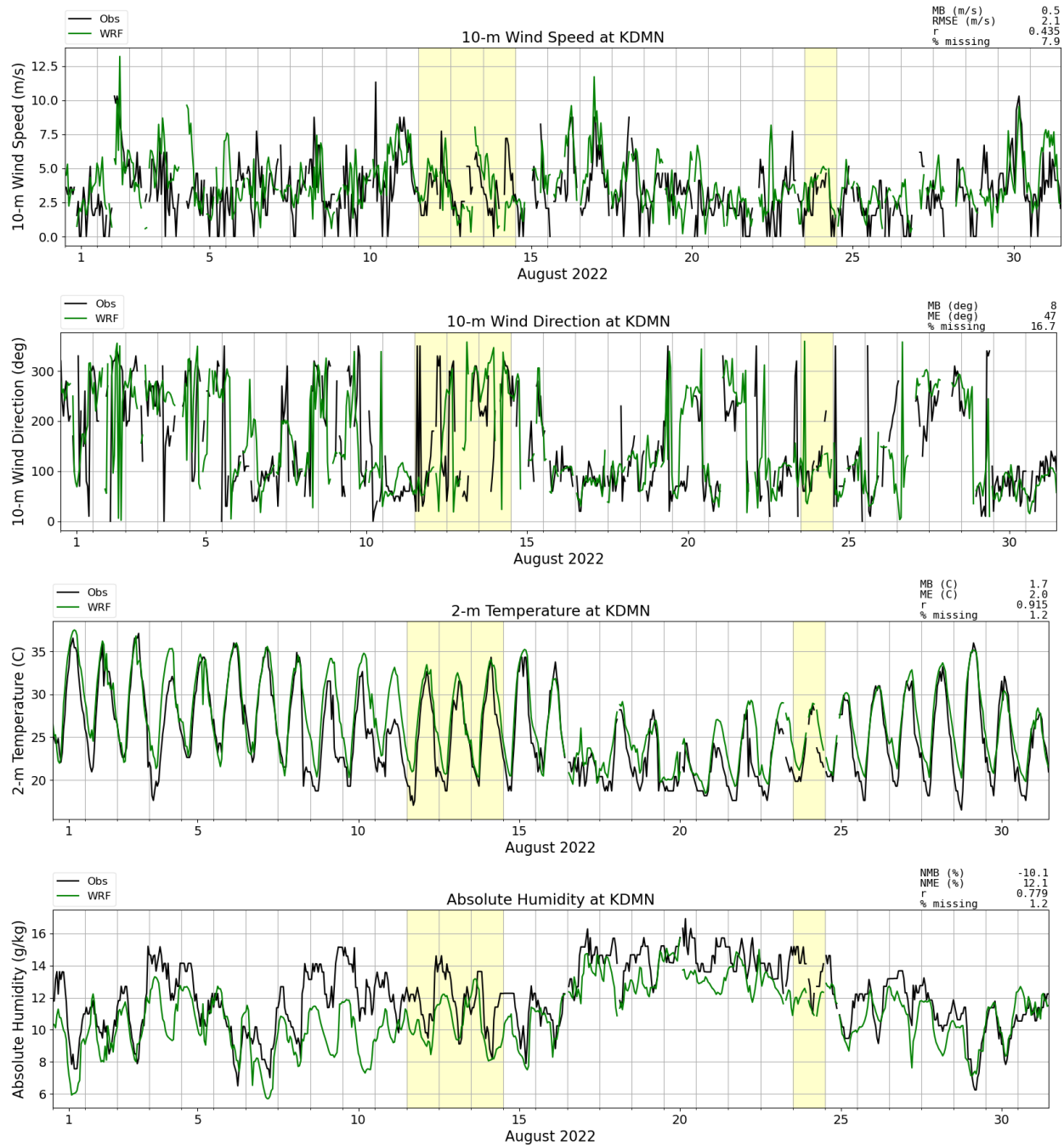


Figure 15. Same as Figure 14 for KDMN but during August 2022.

Qualitative Evaluation for Precipitation

We graphically compared monthly simulated precipitation patterns for May through September in the NMED CAMx 4 km domain against PRISM analyses, as well as daily precipitation patterns on ozone exceedance days when either WRF or observations reported significant precipitation in the region. As mentioned previously, the PRISM precipitation interpolation scheme works better for organized synoptic weather systems than for the stochastic summer convective showers, which tend to be spotty and intermittent, especially in areas with sparse rain gauge data. This is the primary reason why the analysis of precipitation performance remains a qualitative comparison of spatial patterns and magnitudes.

Figure 16 to Figure 19 compare PRISM (left) and simulated (right) total precipitation patterns during each of the five complete simulation months (June-September). WRF generally simulated the precipitation patterns well on the monthly scale across the domain, although in various places the PRISM rain exceeded the WRF total and extent, such as far western Texas during all months except August. On many of the high ozone days, WRF succeeded in replicating the overall patterns but fell short in the totals or had regions lacking any simulated rain. Neither PRISM nor WRF can be expected to replicate the stochastic nature of actual summer convective rainfall in time and space, as illustrated by the different patterns of the heaviest areas of precipitation in each month. It is therefore perhaps more desirable for the meteorological simulation to tend toward a drier environment, as spurious convective precipitation events can drive large local errors in wind and temperature patterns. Such meteorological errors often lead to poor replication of ozone patterns during those times.

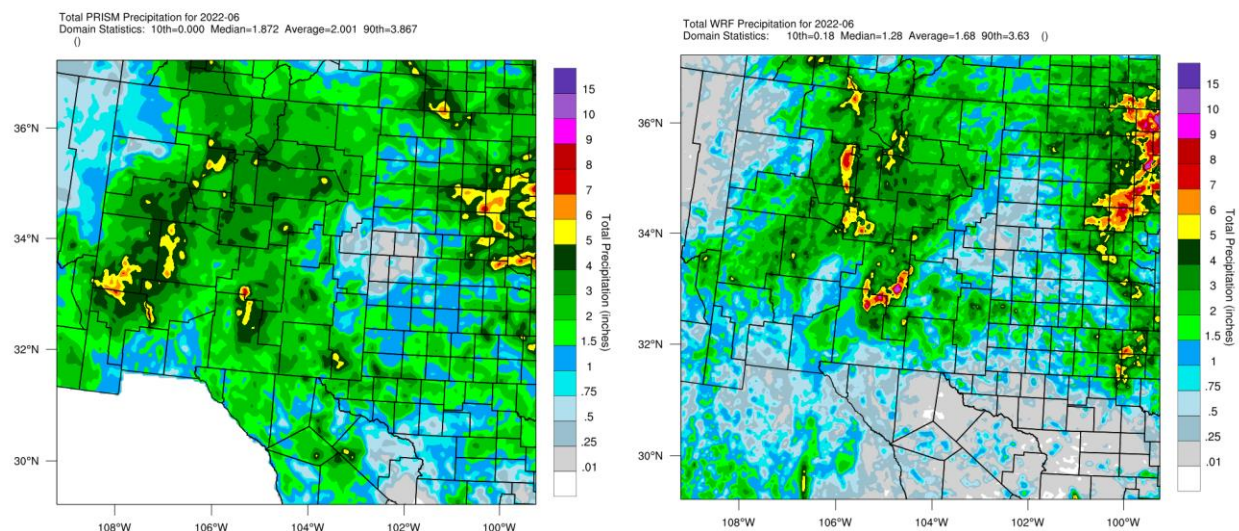


Figure 16. June total precipitation patterns from PRISM based on observations (left) and modeled by WRF (right).

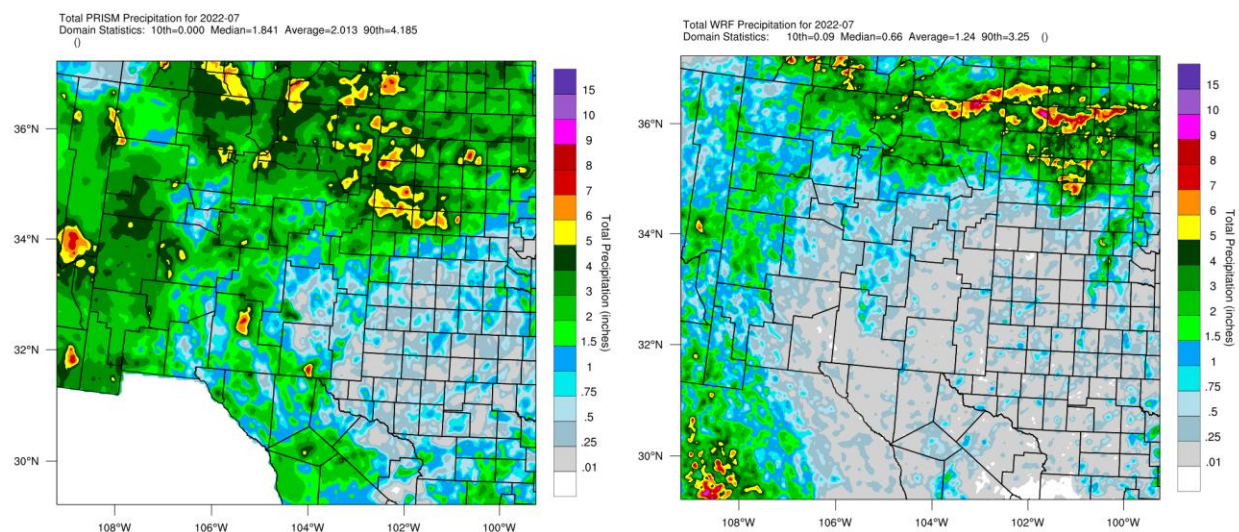


Figure 17. July total precipitation patterns from PRISM based on observations (left) and modeled by WRF (right).

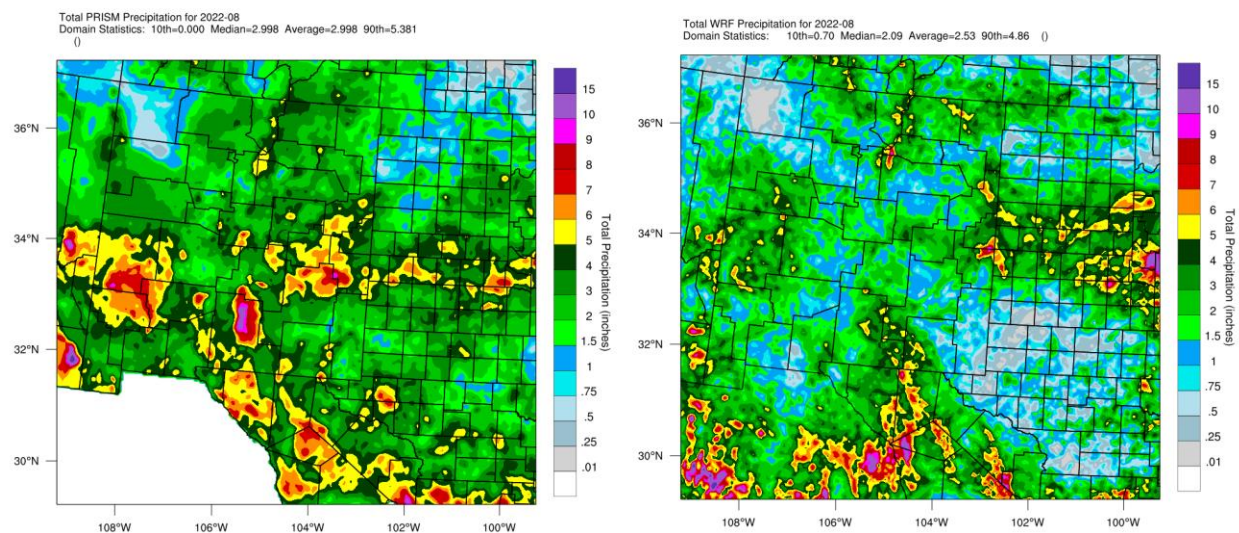


Figure 18. August total precipitation patterns from PRISM based on observations (left) and modeled by WRF (right).

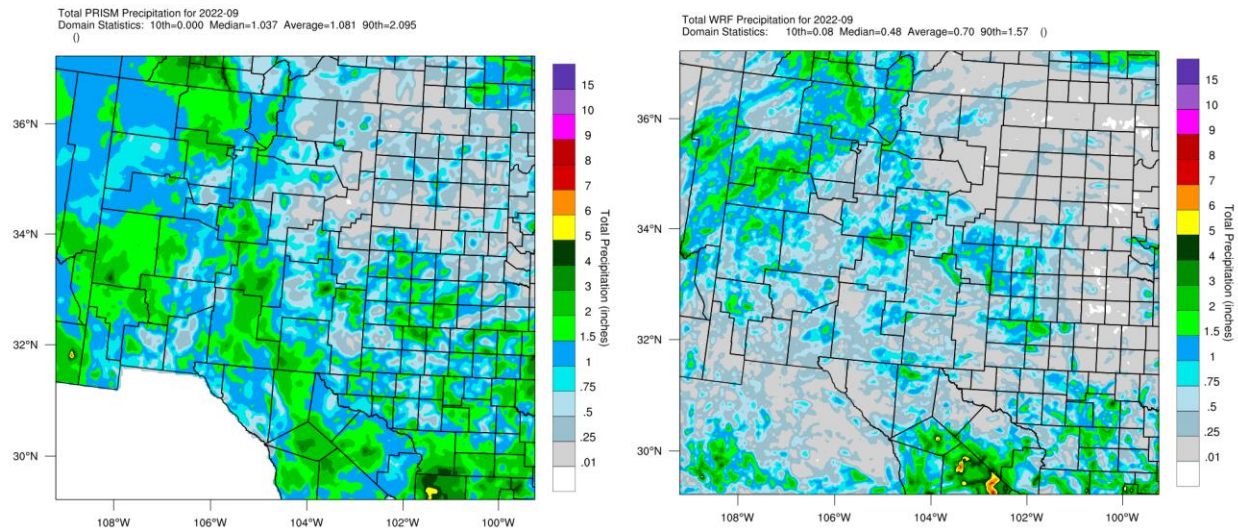


Figure 19. September total precipitation patterns from PRISM based on observations (left) and modeled by WRF (right).

Figure 20 through Figure 31 focus on July and August ozone exceedance days when measurable precipitation occurred over the model domain either in the PRISM data or simulated by WRF: July 1, 2, 8, 9, 10, 11, 13, 14, 15, and August 12, 13, 14. Note that all days except for July 8 were impacted by light-density wildfire smoke. For these daily plots, 24-hour precipitation sums are reported over 12 to 12 UTC (6 AM to 6 AM MDT) and on the plot headers, labeled for the ending time (i.e., the morning following the day of interest). The 24-hour period spans 18 hours on one day and 6 hours on the following day. In the discussion below, when referring to a specific day, for example July 1, we are referring to the 24-hr period that began at 6 am on July 1 and ended at 6 am on July 2, while the plot header for July 1 refers to the end time of the 24-hour period at 6 am on July 2.

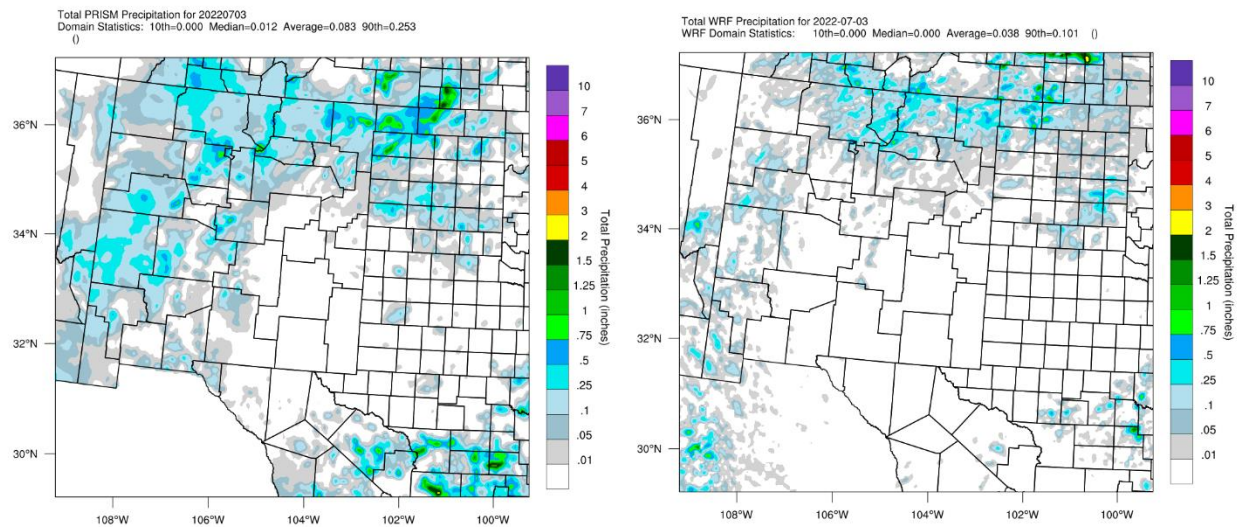
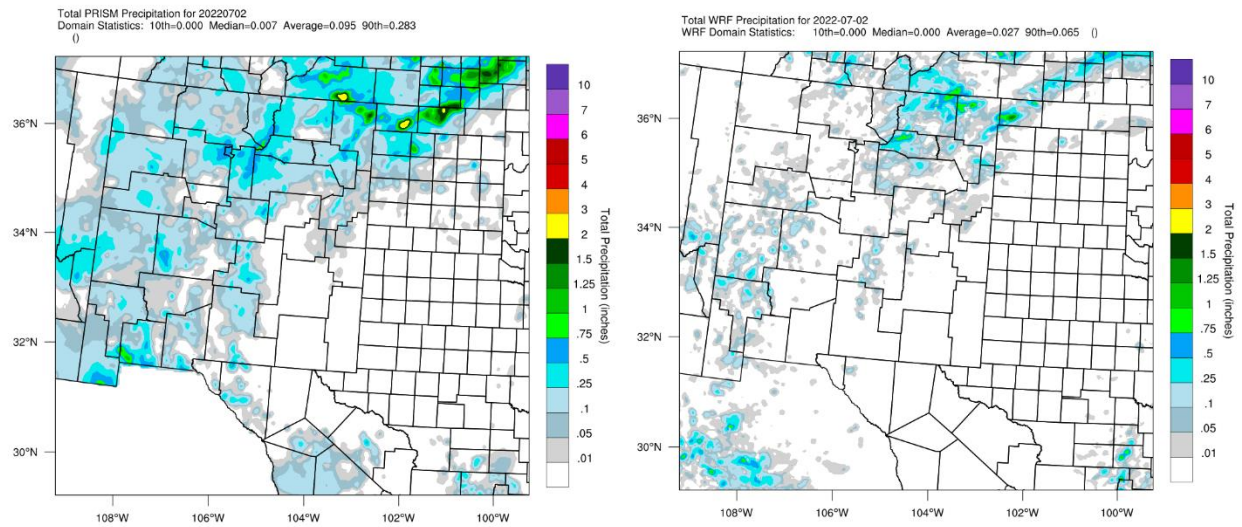
On July 1 (Figure 20) both PRISM and WRF show precipitation over the region, following the pattern of WRF simulating less precipitation than what PRISM indicates. PRISM showed much of New Mexico to have received up to half an inch, compared with the simulated amounts under .25" and dry for many places. The pattern of heavier precipitation areas, which are also simulated in WRF, indicates that WRF captured the mesoscale atmospheric dynamics well, although again it missed the mark by half an inch or more in the northeastern corner of the domain. The model was also too dry in southwestern New Mexico and western Texas. Wind directions were replicated well but speed had greater error, while temperature and humidity had a warm and dry bias on this day.

On July 2 (Figure 21), there is a similar pattern as the previous day, with comparable model performance. WRF again did not show as widespread coverage of precipitation and instead indicates more spotty convection-oriented patterns, while PRISM is more widespread with greater totals in the wettest areas. However, the pattern in WRF may be more reasonable (though likely not accurate in time and space) as PRISM spatially interpolates between rain gauge data, which are lacking in rural areas, and thus may overstate the spatial extent of rainfall. While winds were well replicated on this day, temperature was slightly over predicted, and humidity was under predicted.

Figure 22 through Figure 28 compare precipitation patterns between PRISM and WRF for a series of days with high ozone in mid-July. The July 8-10 period showed a similar tendency toward overly dry WRF simulation. Compared to New Mexico surface station observations, WRF simulations had a warm bias between 1 and 3 K and a consistent 1.5 g/kg low bias for simulated humidity, while wind speed was well modeled. On July 8 (Figure 22) WRF shows less overall accumulation throughout the domain, failing to simulate the broader areas of precipitation. Where WRF did simulate rain happened to be the dry areas on the PRISM plot, and the model was mostly dry for the areas where convective activity fueled higher totals including in northern Texas and west-central New Mexico. July 8 was also not impacted by smoke, unlike the rest of the July exceedance days with light density. The dry bias and suppressed precipitation was most apparent on July 9 (Figure 23). July 9 saw widespread totals up to a half inch throughout much of New Mexico, which was almost completely missed by WRF in New Mexico. For the Texas portion of the Permian Basin, WRF modeled precipitation patterns better, but still missed at least .1" over much of western Texas. On July 10 (Figure 24) WRF had similar performance to July 8 and had significant errors in the extent of precipitation.

The simulated precipitation patterns for July 11 (Figure 25) match observations well in extent in the northern part of the domain, although WRF did not catch much of the light rain in southwestern New Mexico. Throughout, WRF totals were generally lower, and the sparse convective activity appears to have been resolved but limited in accumulation. July 13 (Figure 26) and July 14 (Figure 27) also follow the general pattern of under-simulated precipitation within the observed footprint of the rain event. Later on July 15 (Figure 28) WRF simulated very little precipitation in New Mexico, although PRISM showed some convective activity in northwestern New Mexico associated with up to a half inch of precipitation.

The mid-August episode of three very high ozone days qualitatively matched the precipitation patterns well, although simulated precipitation for August 12 (Figure 29) and particularly August 14 (Figure 31) lacked the extent of PRISM observations. August 13 (Figure 30) shows strong agreement in extent and magnitude, although the simulated day contains examples of WRF's spurious convection in the southeastern portion of the domain. The overall agreement between WRF and PRISM on precipitation patterns is generally the best during this period, compared to the other ozone exceedance days. The slight warm and dry biases were also present during this period.



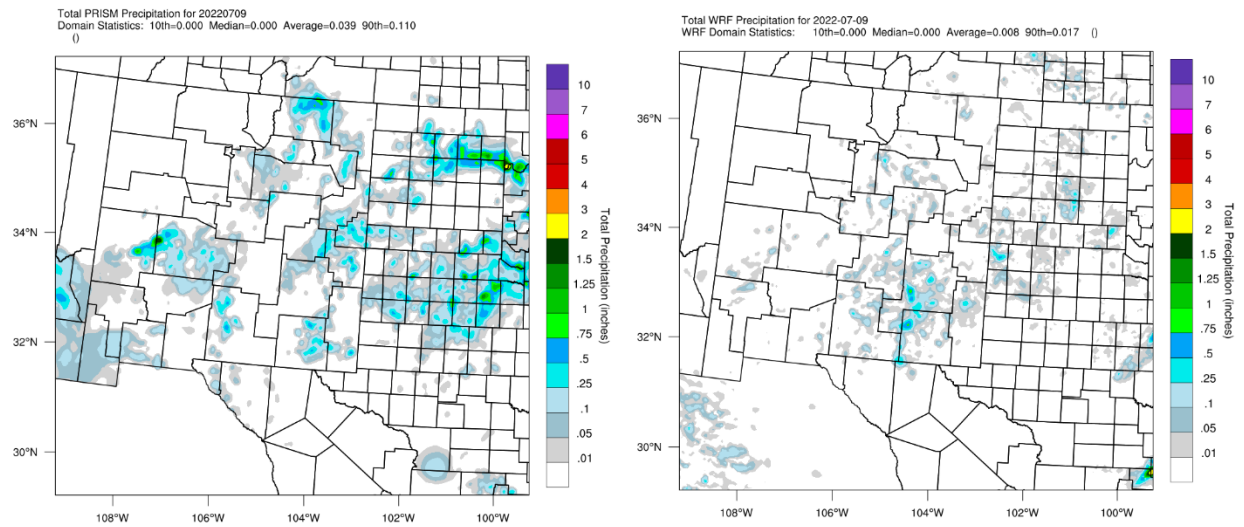


Figure 22. Daily precipitation patterns from PRISM based on observations (left) and modeled by EPA WRF (right) for the 24-hour period from July 8 at 6 AM ending July 9, 2022 at 6 AM MDT.

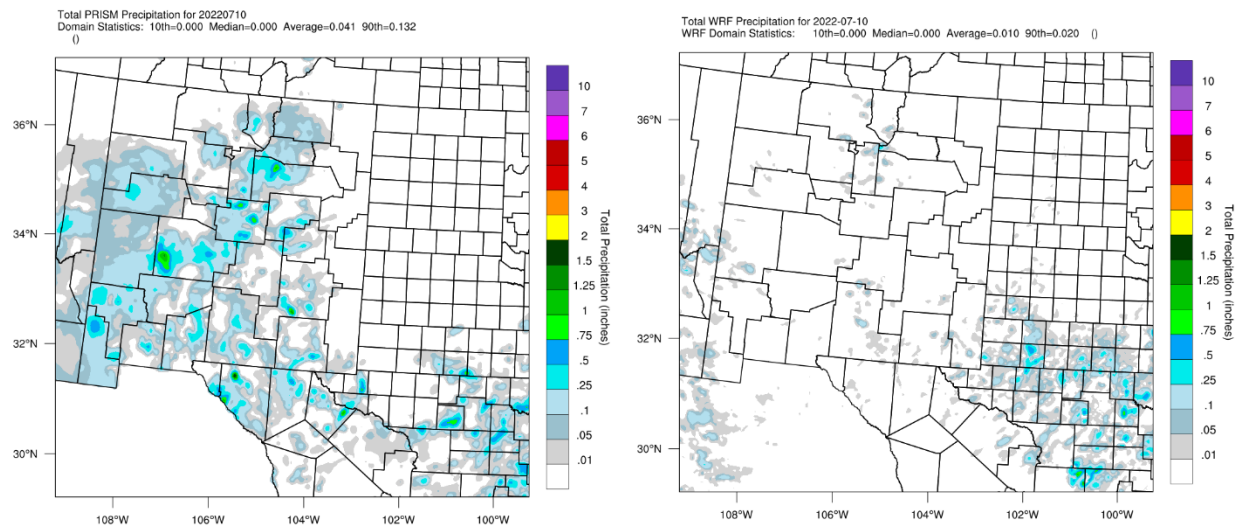
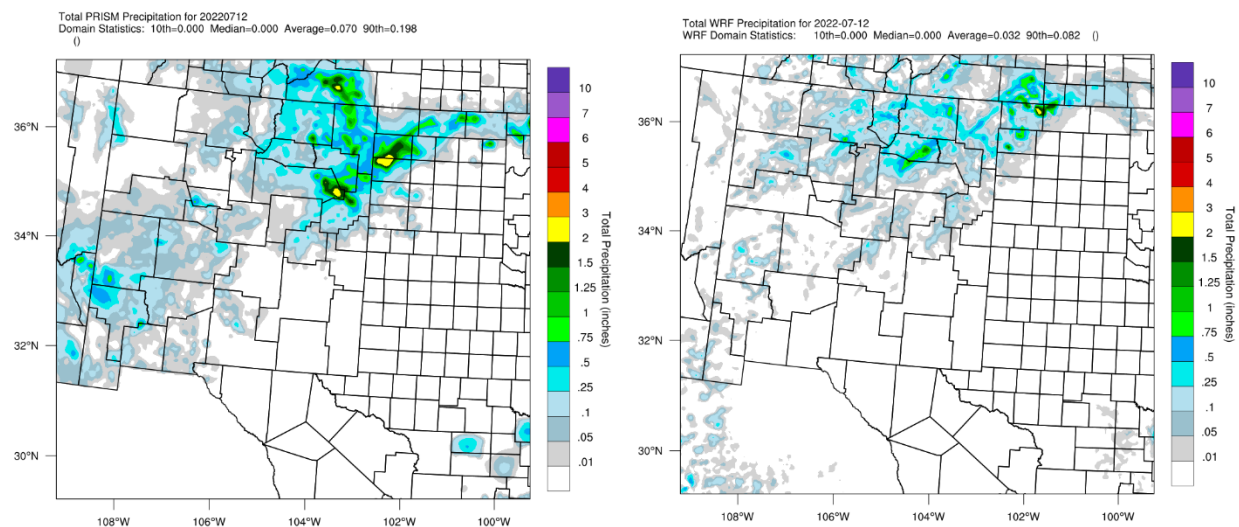
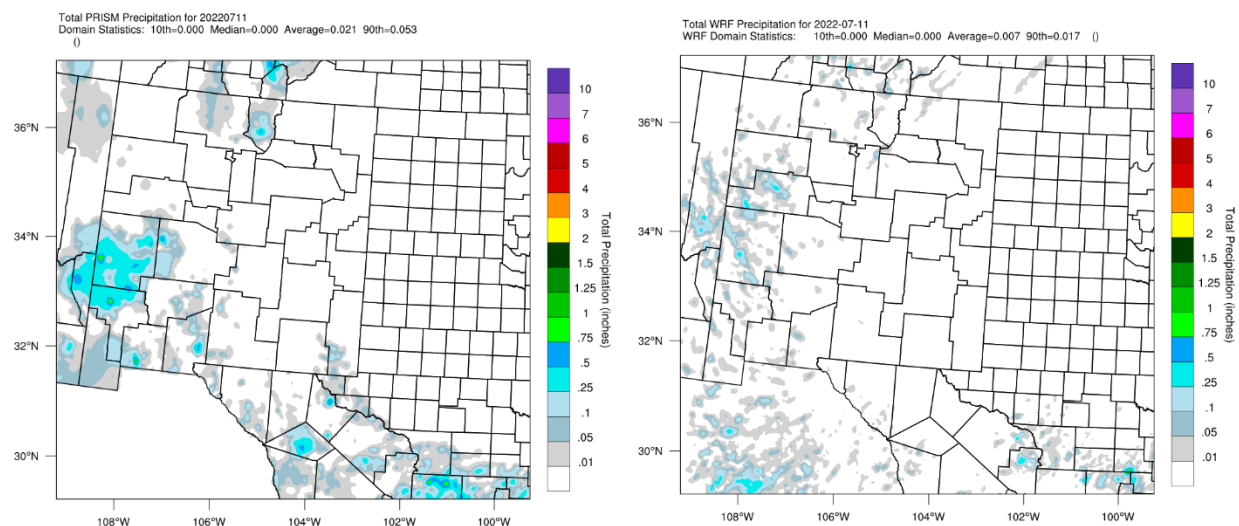


Figure 23. Daily precipitation patterns from PRISM based on observations (left) and modeled by EPA WRF (right) for the 24-hour period from July 9 at 6 AM ending July 10, 2022 at 6 AM MDT.



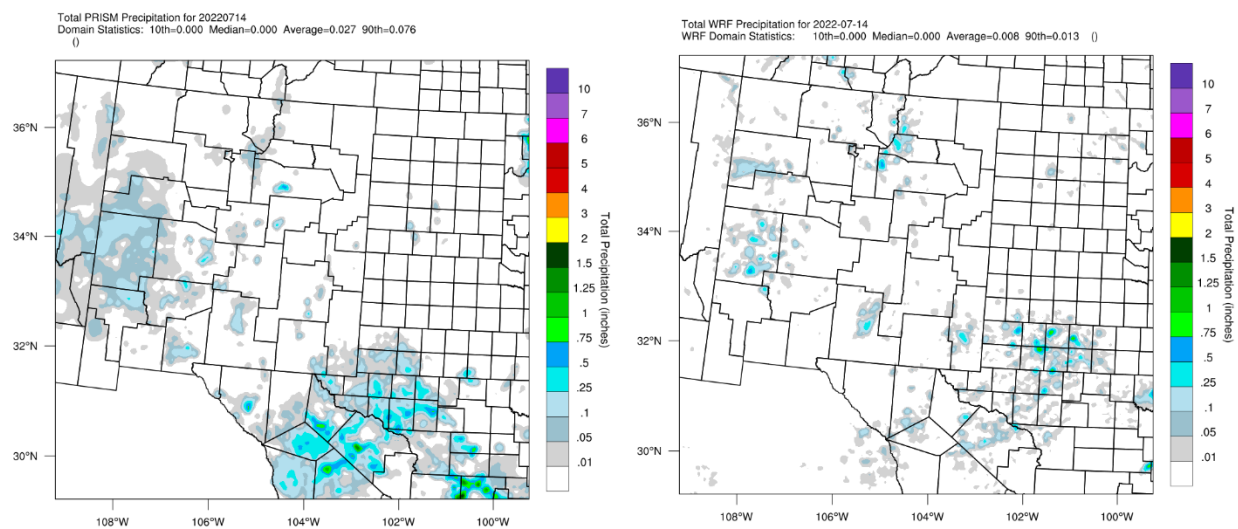


Figure 26. Daily precipitation patterns from PRISM based on observations (left) and modeled by EPA WRF (right) for the 24-hour period from July 13 at 6 AM ending July 14, 2022 at 6 AM MDT.

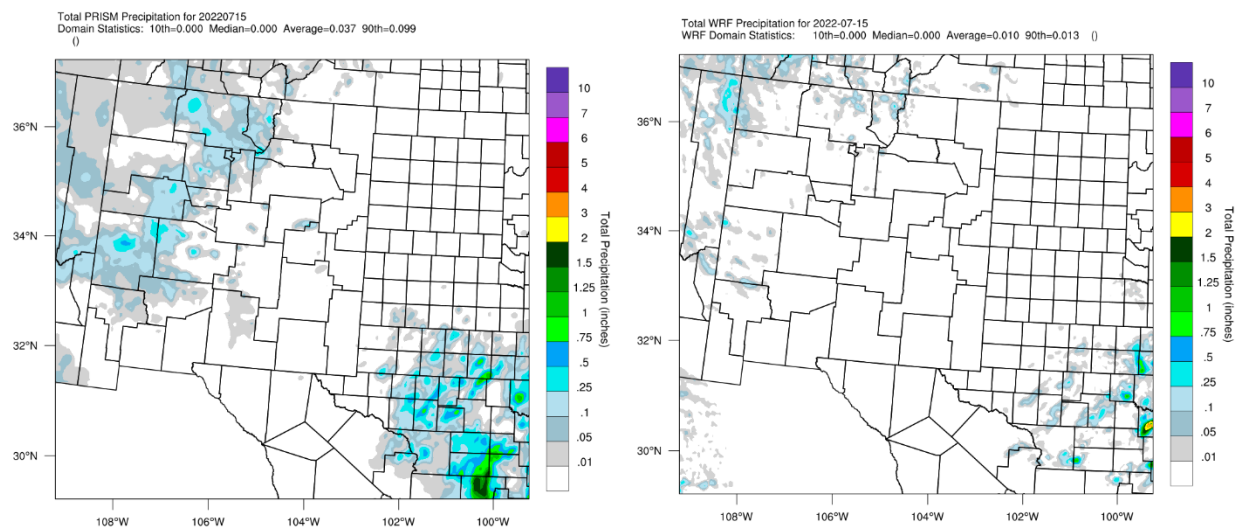


Figure 27. Daily precipitation patterns from PRISM based on observations (left) and modeled by EPA WRF (right) for the 24-hour period from July 14 at 6 AM ending July 15, 2022 at 6 AM MDT.

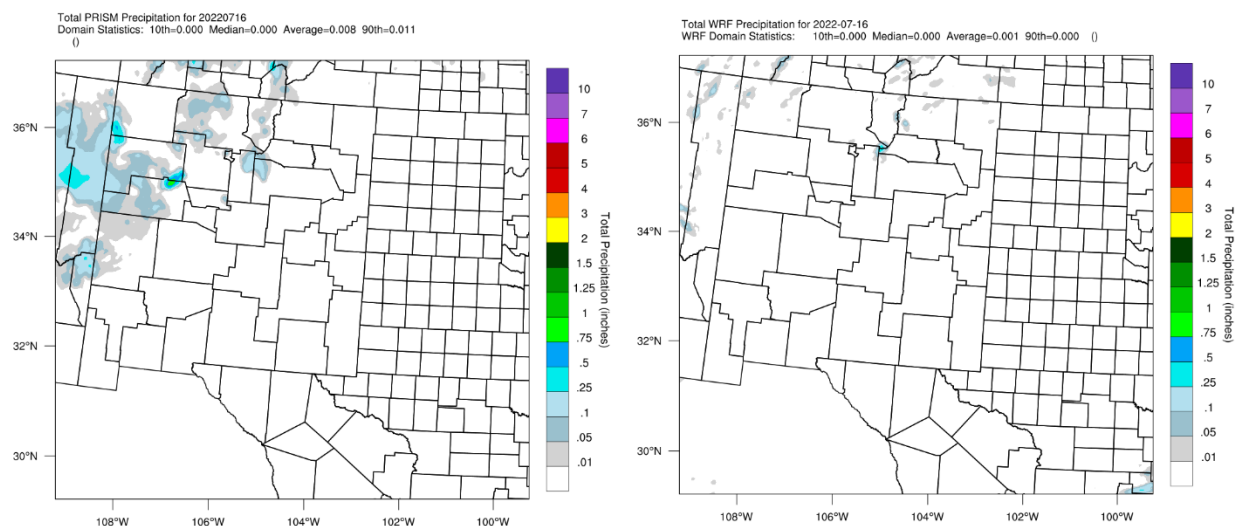


Figure 28. Daily precipitation patterns from PRISM based on observations (left) and modeled by EPA WRF (right) for the 24-hour period from July 15 at 6 AM ending July 16, 2022 at 6 AM MDT.

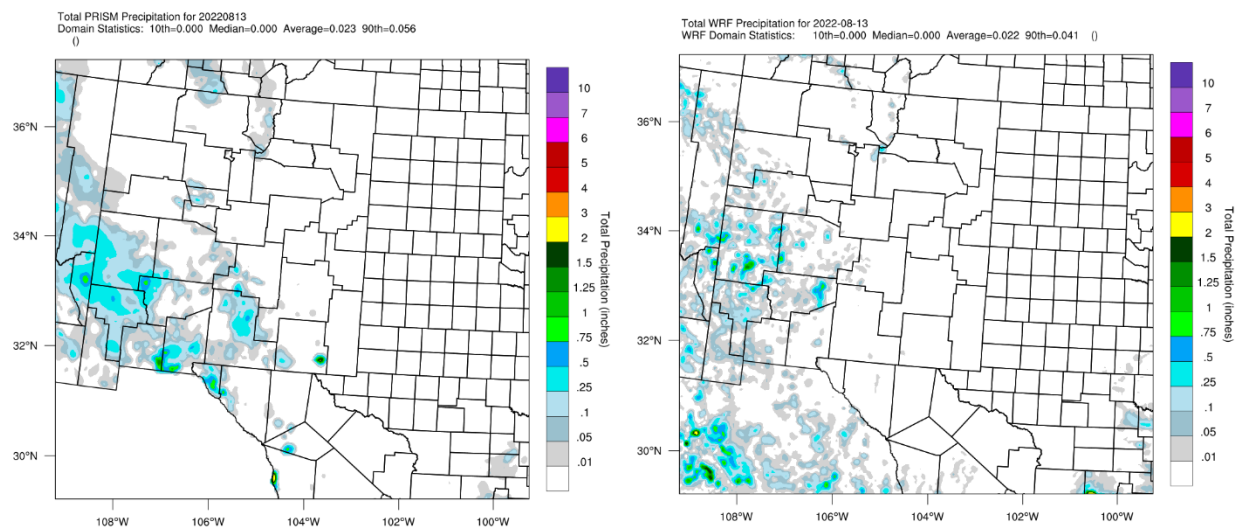
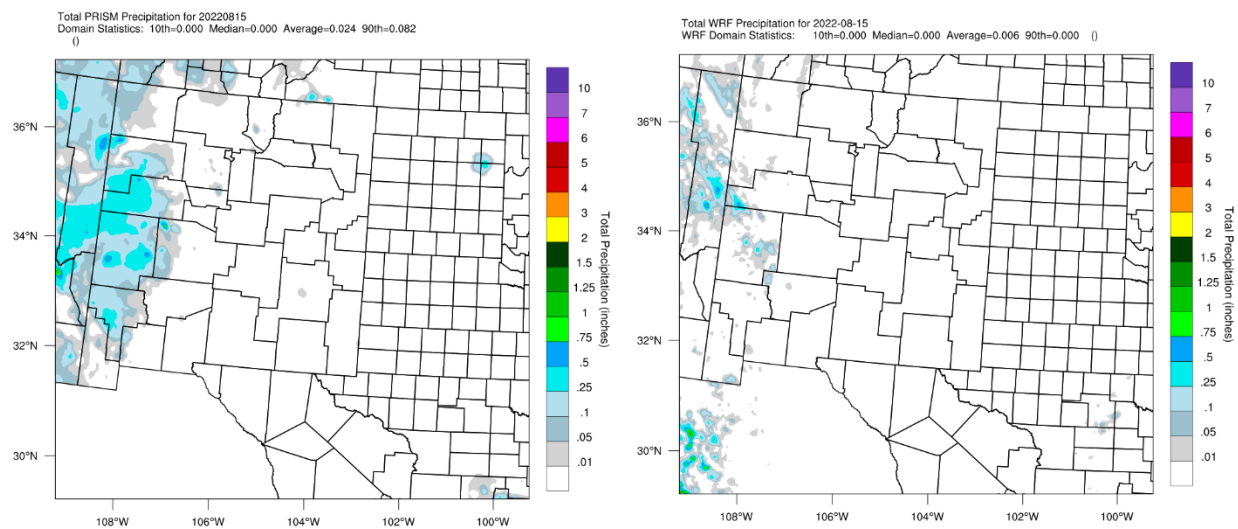
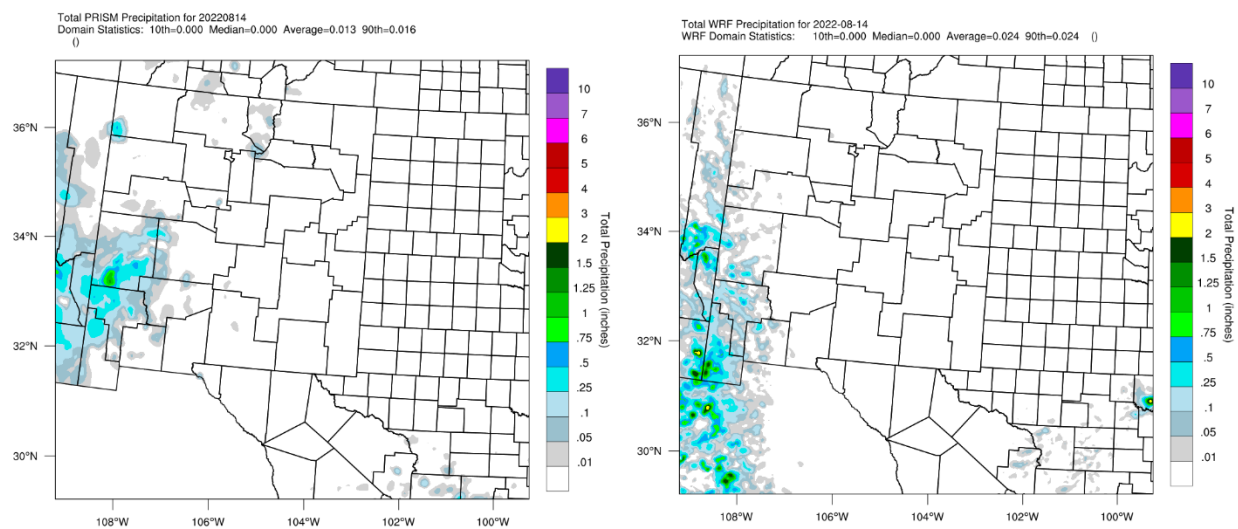


Figure 29. Daily precipitation patterns from PRISM based on observations (left) and modeled by EPA WRF (right) for the 24-hour period from August 12 at 6 AM ending August 13, 2022 at 6 AM MDT.



Phenomenological Evaluation

EPA's Modeling Guidance (EPA, 2018) recommends a phenomenological or an event-based meteorological evaluation as part of any air modeling study. For the phenomenological evaluation in this study, we evaluate how large-scale meteorological features in the lower atmosphere (above the influence of terrain) are simulated in WRF relative to observational analyses. Figure 32 shows an example comparing a 500 hPa (~6 km above sea level) upper air analysis chart² against 500 hPa WRF fields from the 12 km grid on August 12, 2022 at 6 PM MDT. This day was characterized by the highest MDA8 ozone concentrations at the Carlsbad City monitoring site within the New Mexico Permian Basin, and light smoke density.

On August 12, both the analysis chart and WRF results show the dominance of high pressure, weak pressure gradients, and light winds over the western U.S. Placement of WRF height contours generally agree with those shown on the analysis chart, although the small areas of relatively higher pressure in northern Chihuahua and southern Sinaloa present in the analysis map are not captured by the model. Wind patterns are reproduced rather well, with differences between the model and observations confined to primarily calm areas. Temperatures indicate a maximum heat dome over the southern Plains extending into the desert Southwest.

Similar graphical comparisons for 7 other high ozone days are provided in Appendix A; consistently good performance in replicating synoptic scale conditions is evident among all high ozone days. The selected high ozone days saw New Mexico and the Permian Basin at the center of broad atmospheric ridging, characterized by relatively calm winds throughout the vertical profile and generally warmer temperatures. High dew point depressions are common across all days in the region, signifying absent cloud cover and low humidity aloft. WRF disagrees with the analysis map slightly on the spatial extent of the ridge's peak areas on July 11 and August 12, although the upper-level wind and temperature patterns for those days agree strongly.

² Available at: <https://www.spc.noaa.gov/obswx/maps/>

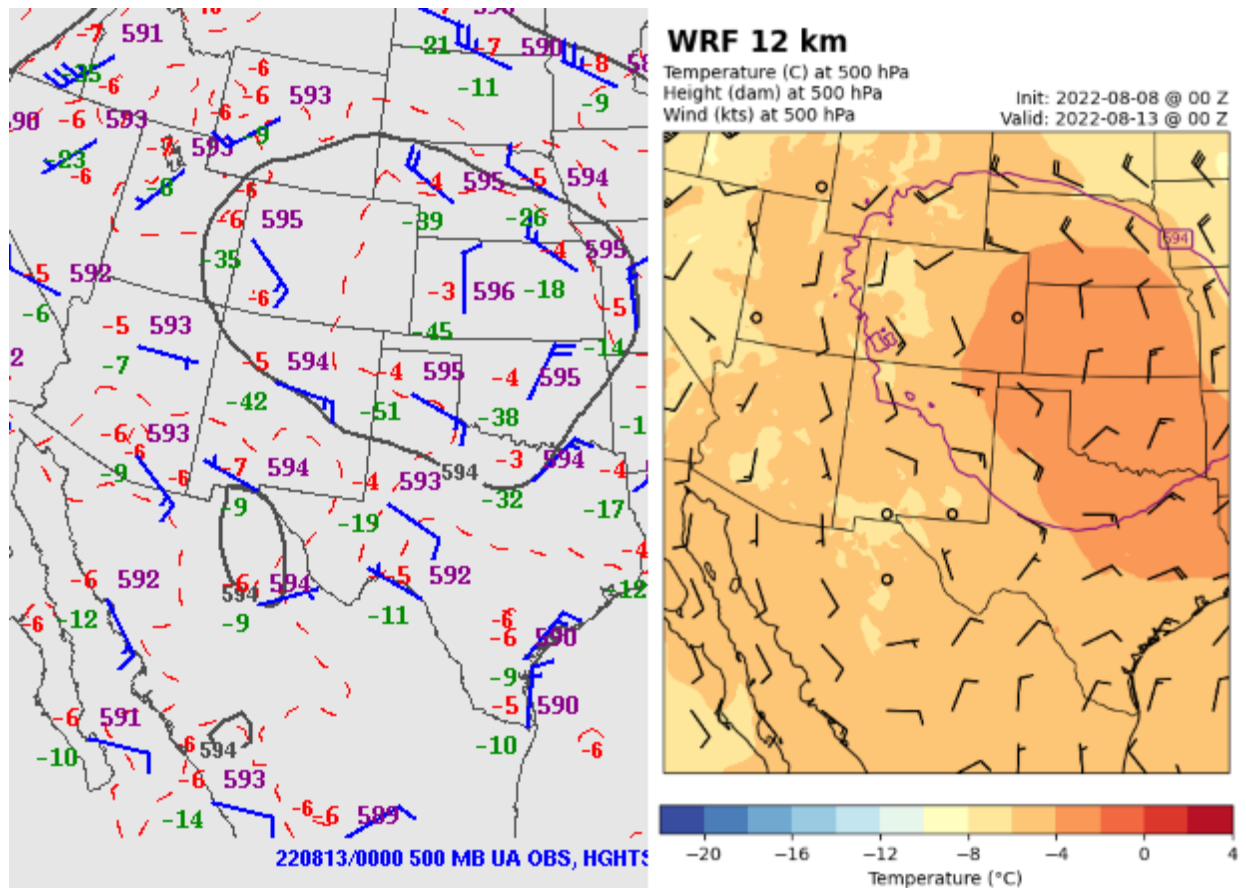


Figure 32. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on August 12, 2022 at 6 PM MDT.

REFERENCES

- Daley, C., M. Halbleib, J. Smith, W. Gibson, M. Doggett, G. Taylor, J. Curtis, P. Pasteris, 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Intl. J. Climate, DOI: 10.1002/joc.1688.
http://prism.oregonstate.edu/documents/Daly2008_PhysiographicMapping_IntJnClim.pdf.
- Emery, C., E. Tai, G. Yarwood, 2001. Enhanced Meteorological Modeling and Performance Evaluation for Two Texas Episodes. Prepared for the Texas Natural Resources Conservation Commission by ENVIRON International Corp., Novato, CA.
- EPA, 2018. Modeling Guidance for Demonstrating Air Quality Goals for Ozone, PM_{2.5}, and Regional Haze. Prepared by the US Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division (EPA 454/R-18-009, November 2018). https://www3.epa.gov/ttn/scram/guidance/guide/O3-PM-RH-Modeling_Guidance-2018.pdf.
- Kemball-Cook, S., Y. Jia, C. Emery, R. Morris, 2004. 2002 annual MM5 simulation to support WRAP CMAQ visibility modeling for the section 308 SIP/TIP. Prepared for The Western Regional Air Partnership by ENVIRON International Corp, Novato, CA.
- McNally, D., G. Schewe, J. Johnson, R. Morris. 2008. Evaluation of Preliminary MM5 Meteorological Model Simulation for the June-July 2006 Denver Ozone SIP Modeling Period Focused on Colorado. Prepared for the Denver Regional Air Quality Council (RAQC) by Alpine Geophysics, LLC and ENVIRON International Corporation (February 2008). http://www.ozoneaware.org/documents/MM5_Eval_DENSIP_Feb25_2008.pdf.
- Ramboll, 2021. New Mexico Ozone Attainment Initiative Photochemical Modeling Study – Final Air Quality Technical Support Document. May.
- Ramboll, 2025a. Work Plan: 2026 Ozone Source Apportionment Modeling for the Permian Basin using the 2022v1 Modeling Platform. Ramboll, Novato, California. April.
- Ramboll. 2025b. 2026 Ozone Source Apportionment Modeling for the Permian Basin Using the 2022v1 Modeling Platform – 2022v1 Model Evaluation and 2026v1 Source Apportionment Modeling – Draft. Ramboll, Novato, California. June.
- Ramboll. 2025c. Preliminary CAMx 2022 4 km Ozone Model Performance Evaluation Using New WRF 2022 4 km Meteorological Inputs. Ramboll, Novato, California. June.
- Skamarock, W.C., et al., 2019. A Description of the Advanced Research WRF Model Version 4. NCAR Technical Note TN-556+STR, Mesoscale and Microscale Meteorology Laboratory, National Center for Atmospheric Research, Boulder, CO. March.
<https://opensky.ucar.edu/islandora/object/opensky:2898>.

APPENDIX A

Comparisons of 500 hPa analyses and WRF 12 km simulation on 7 ozone exceedance days in summer 2022

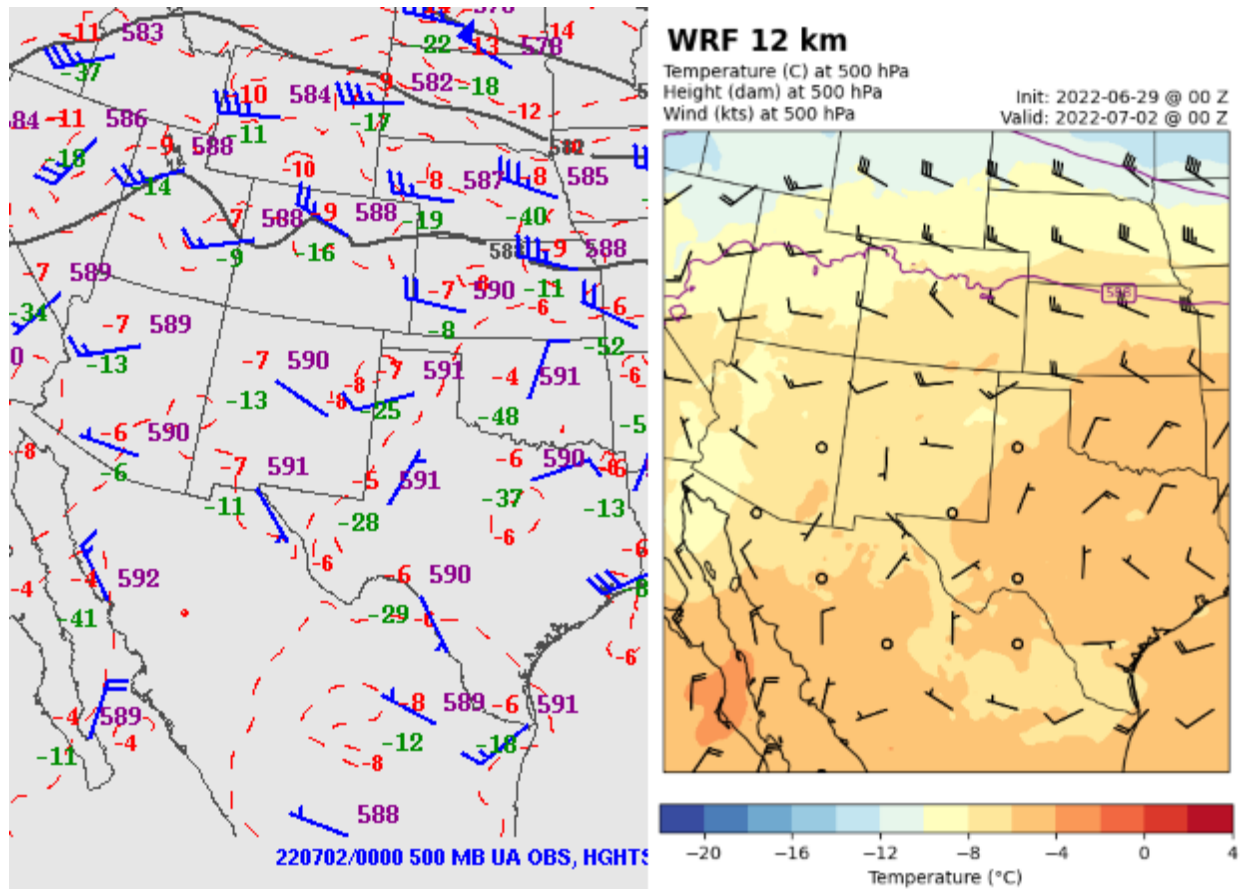


Figure A-1. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on July 1, 2022 at 6 PM MDT.

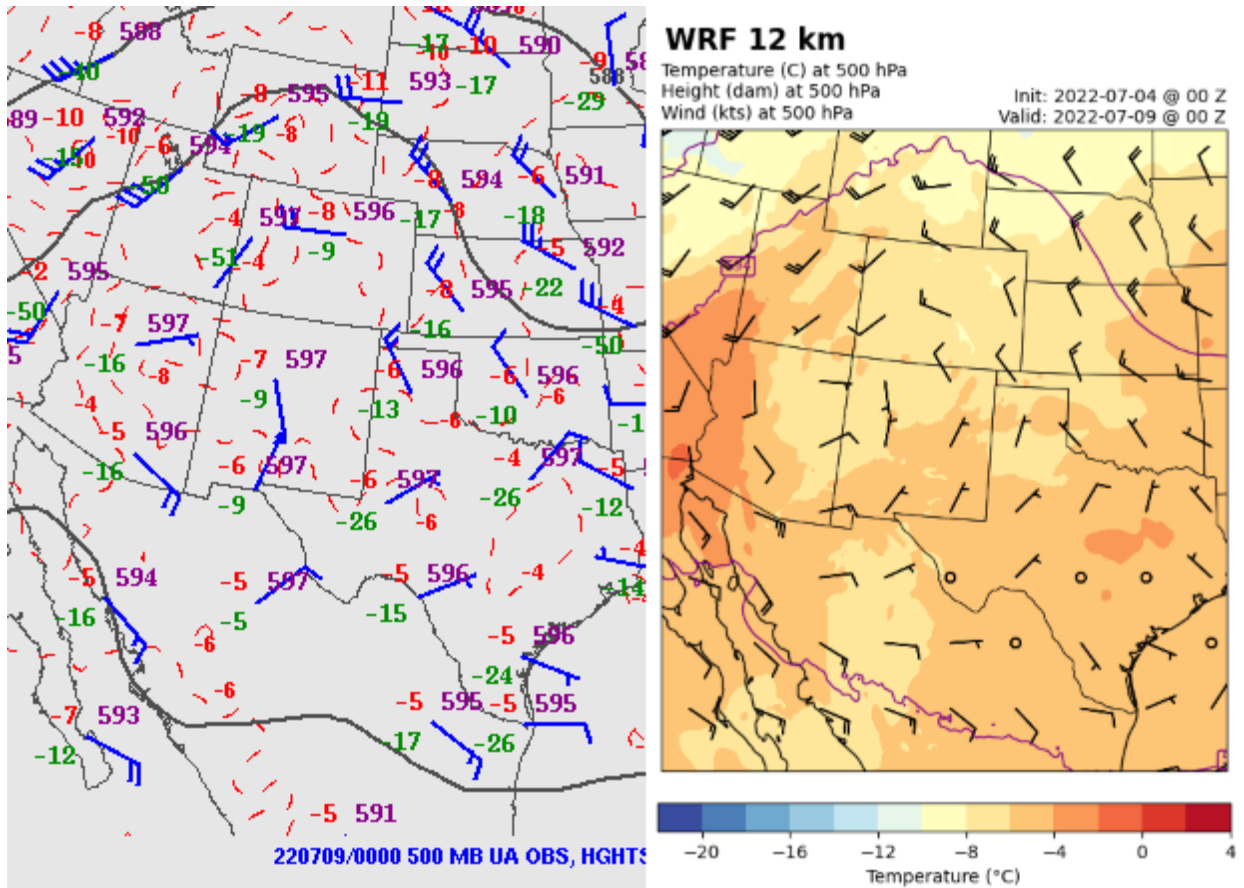


Figure A-2. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on July 8, 2022 at 6 PM MDT.

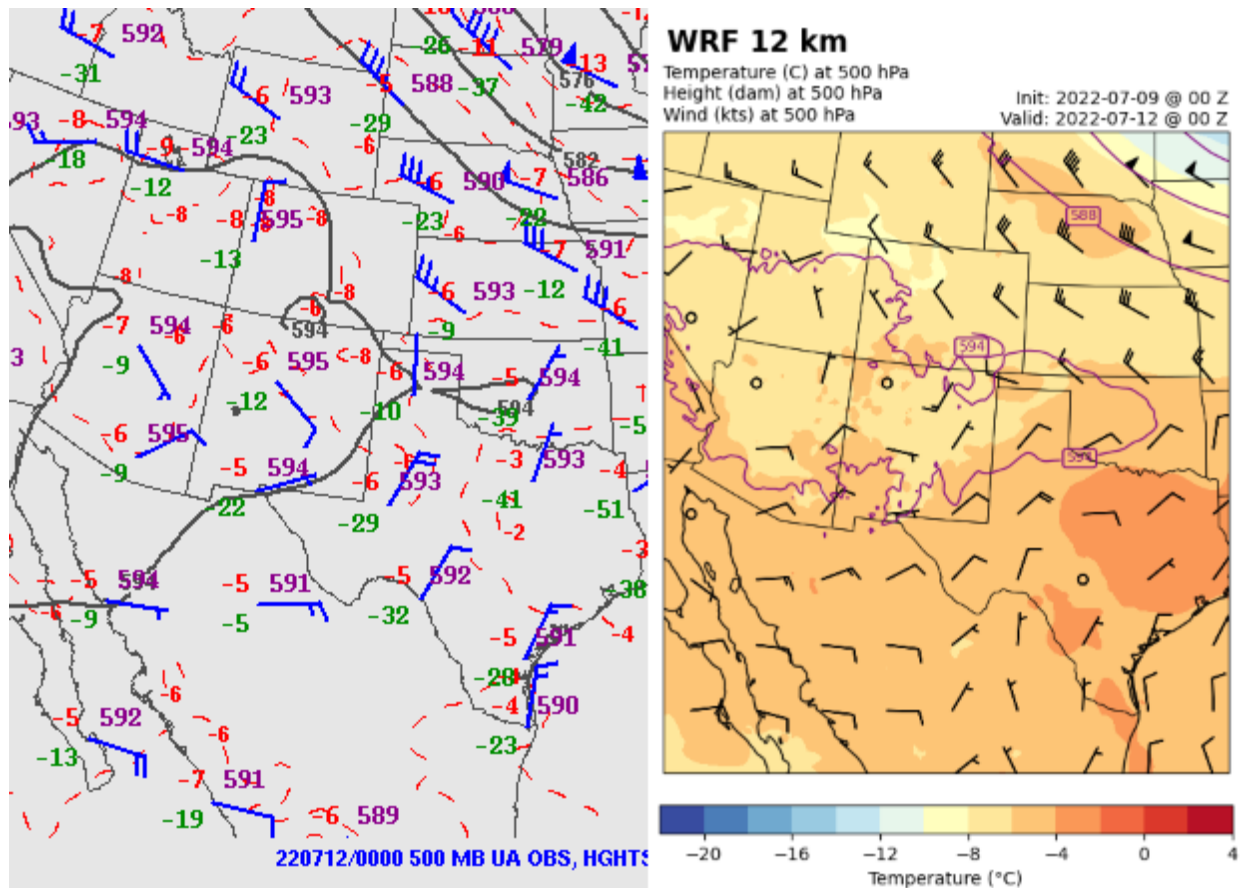


Figure A-3. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on July 11, 2022 at 6 PM MDT.

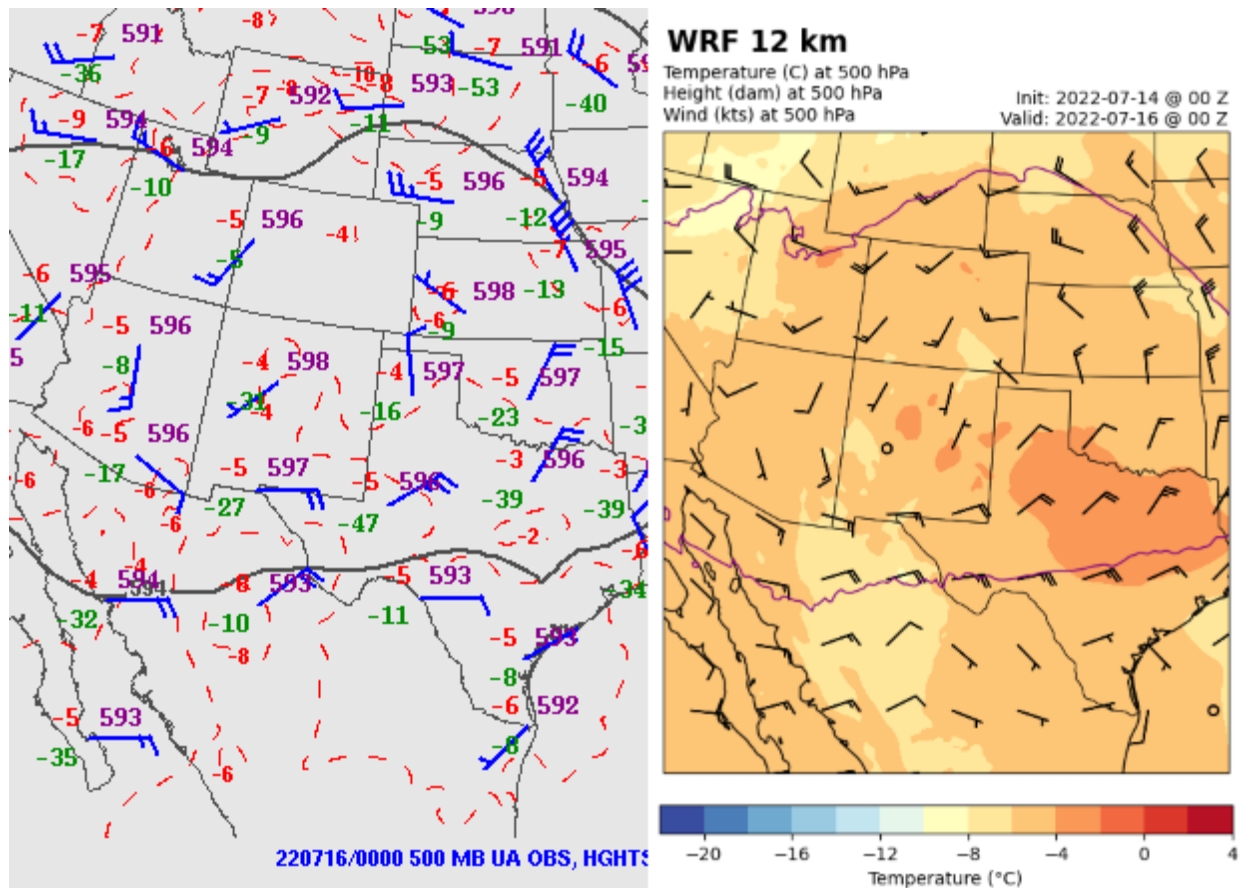


Figure A-4. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on July 15, 2022 at 6 PM MDT.

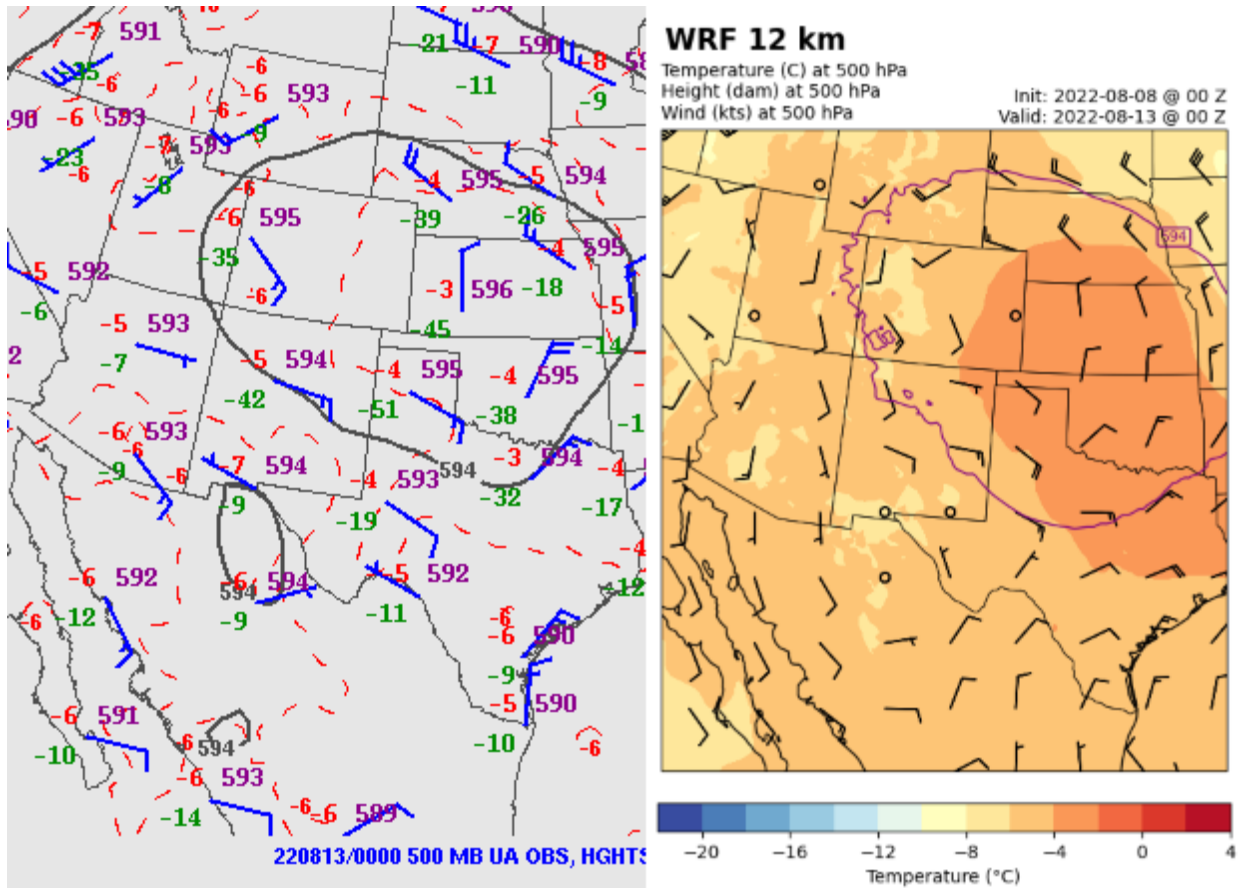


Figure A-5. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on August 12, 2022 at 6 PM MDT.

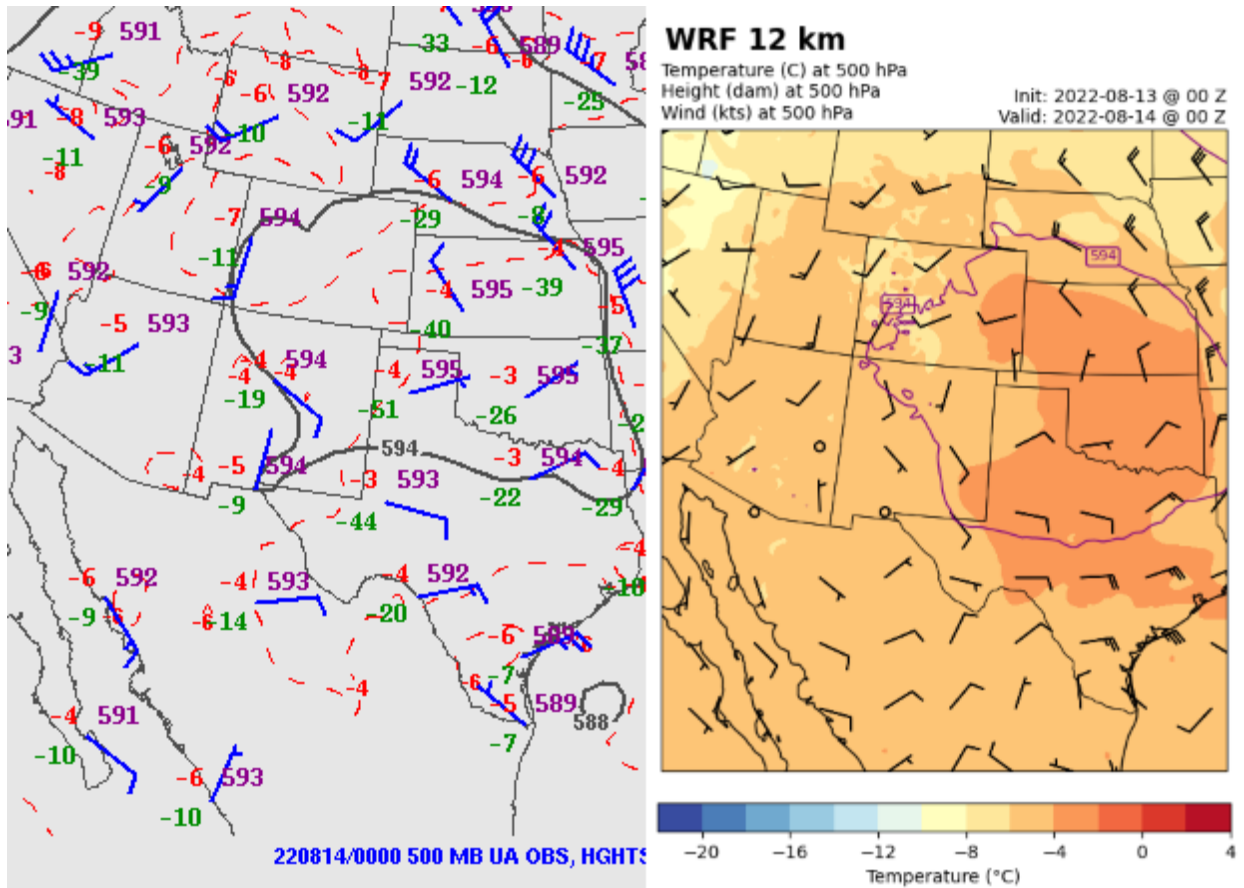


Figure A-6. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on August 13, 2022 at 6 PM MDT.

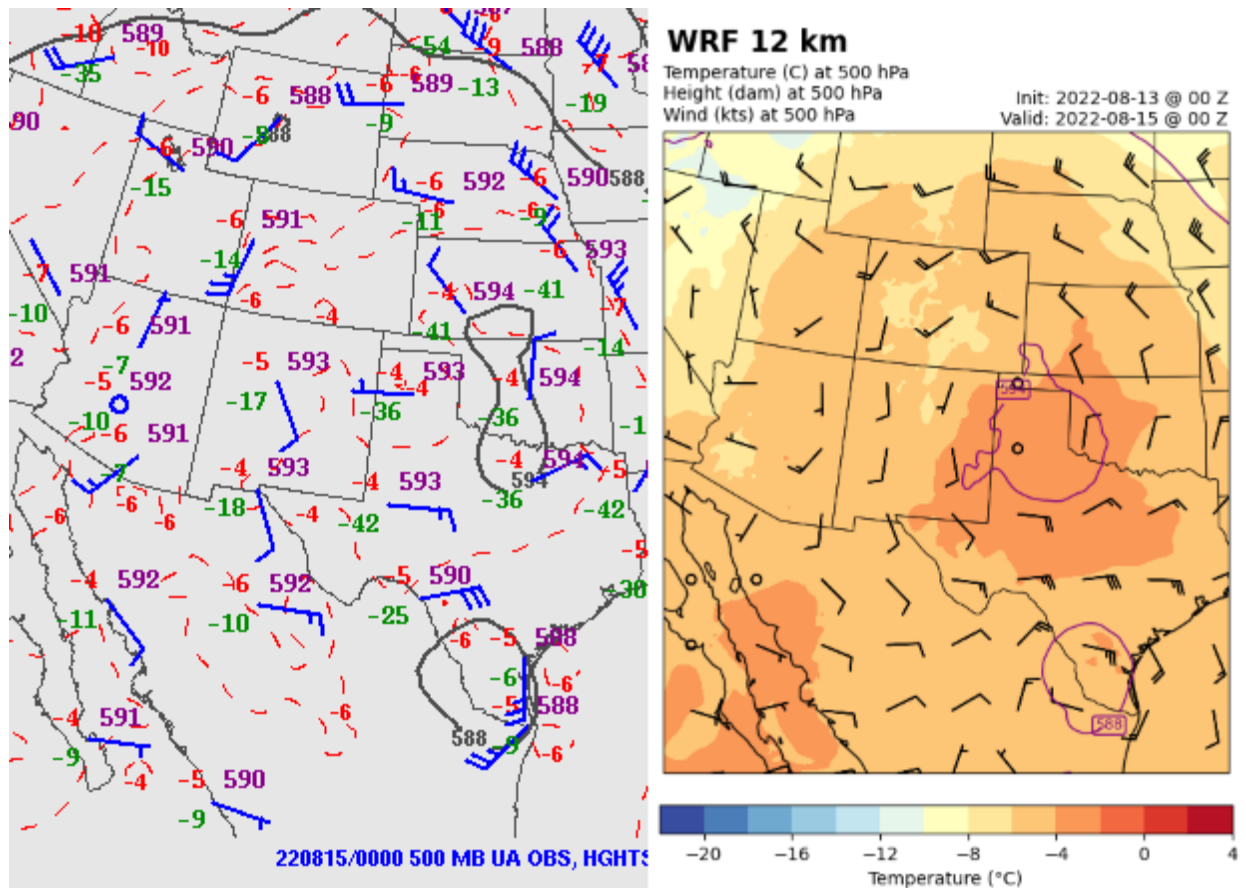


Figure A-7. 500 hPa upper air analysis chart (left; wind barbs, height contours in dark grey; temperature contours in dashed red) and 500 hPa WRF 12 km (right) height contours (purple), wind vectors and temperature (color shaded) on August 14, 2022 at 6 PM MDT.