

# Boundary Layer and Surface Verification of the High-Resolution Rapid Refresh, Version 3

ROBERT G. FOVELL AND ALEX GALLAGHER

*Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York*

(Manuscript received 22 June 2020, in final form 14 August 2020)

**ABSTRACT:** While numerical weather prediction models have made considerable progress regarding forecast skill, less attention has been paid to the planetary boundary layer. This study leverages High-Resolution Rapid Refresh (HRRR) forecasts on native levels, 1-s radiosonde data, and (primarily airport) surface observations across the conterminous United States. We construct temporally and spatially averaged composites of wind speed and potential temperature in the lowest 1 km for selected months to identify systematic errors in both forecasts and observations in this critical layer. We find near-surface temperature and wind speed predictions to be skillful, although wind biases were negatively correlated with observed speed and temperature biases revealed a robust relationship with station elevation. Above  $\approx 250$  m above ground level, below which radiosonde wind data were apparently contaminated by processing, biases were small for wind speed and potential temperature at the analysis time (which incorporates sonde data) but became substantial by the 24-h forecast. Wind biases were positive through the layer for both 0000 and 1200 UTC, and morning potential temperature profiles were marked by excessively steep lapse rates that persisted across seasons and (again) exaggerated at higher elevation sites. While the source or cause of these systematic errors are not fully understood, this analysis highlights areas for potential model improvement and the need for a continued and accessible archive of the data that make analyses like this possible.

**SIGNIFICANCE STATEMENT:** We employed high vertical resolution radiosonde and near-surface observations to evaluate an operational numerical weather prediction model, the High-Resolution Rapid Refresh (HRRR), focusing on the lowest 1 km above ground level. This layer of the atmosphere plays a critical role in overall forecast skill and uncertainty as processes within it must be parameterized. Our analysis critically assessed not only the model forecasts but also the observations, and we determined that radiosonde wind information to be contaminated below about 250 m above the ground. Our verification revealed consistent biases with respect to wind speed and also between temperature and elevation, thereby identifying several areas for model improvement and highlighting the value of high-resolution observations in the boundary layer.

**KEYWORDS:** Radiosonde observations; Surface observations; Forecast verification/skill; Numerical weather prediction/forecasting

## 1. Introduction

Forecast skill of operational numerical weather prediction (NWP) models has been increasing thanks to improvements on many fronts, not limited to initialization, data assimilation, ensemble approaches, model physical parameterizations, grid design, numerical methods, and even applications of artificial intelligence (e.g., Bytheway et al. 2017; Loeser et al. 2017; Caron and Steenburgh 2020; Beck et al. 2020; Burke et al. 2020). It is particularly straightforward to assess forecast skill against near-surface measures such as the 10-m wind speed and 2-m temperature and humidity owing to the substantial availability of observations, at least over land, thanks in part to the proliferation of mesonets (cf. Mahmood et al. 2017). However, we have remained largely unable to assess how good forecasts of winds or temperature are in the lower troposphere just above the heights of temperature shelters and anemometers. By any measure, the data available for verification in the planetary boundary layer (PBL), a highly parameterized region that can serve as a significant source of forecast error, is much more sparse in time and/or space, more difficult to obtain and manipulate, and can also be subject to contamination (e.g.,

Daniel et al. 1999; Walters et al. 2014). Furthermore, comparisons often focus on integrated or averaged metrics such as 500-hPa height or liquid water path rather than the vertical structure of the PBL.

This study emerged from a need to understand how good forecasts of winds and temperatures in the PBL are and whether they possessed systematic and predictable biases. Our area of interest is the conterminous United States (CONUS), and we have selected a set of observations, consisting of radiosondes from the Radiosonde Replacement System (RRS), and an operational model, the High-Resolution Rapid Refresh (HRRR), on which to focus. The HRRR is related to the Rapid Refresh (RAP) model described in Benjamin et al. (2016) and is based on the Weather Research and Forecasting (WRF) Model's Advanced Research WRF (ARW) core (Skamarock et al. 2008). The purpose of this analysis is to assess the skill of HRRR in the PBL, identify areas where further improvements in the modeling system such as parameterizations can and should be made, and also to assess limitations of the radiosonde observations employed in the verification. This will be accomplished by examining longer temporal and larger spatial periods that smooth away the influence of individual weather events in order to reveal systematic characteristics of the model forecasts that might be obscured at a single site and/or time.

*Corresponding author:* Robert G. Fovell, rfovell@albany.edu

DOI: 10.1175/WAF-D-20-0101.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

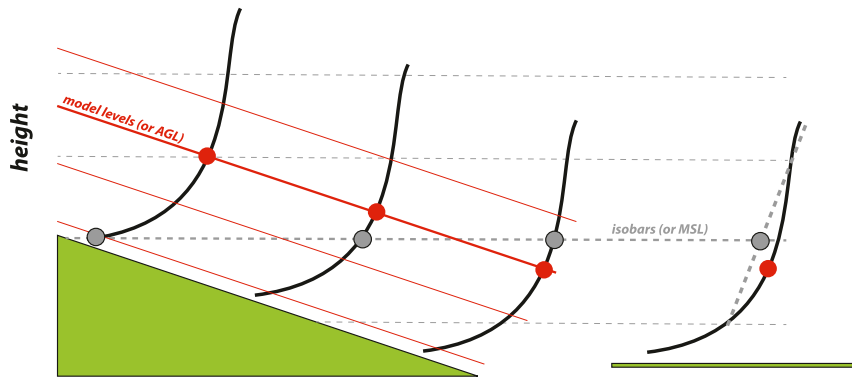


FIG. 1. Conceptual demonstration of how isobaric (dashed gray lines) and terrain-following (solid red) model levels can differ with respect to the composite wind profiles (shown at right) they create in the presence of topography (green). The red and gray dots help identify information that is aggregated into the composite profiles at right.

It is well recognized that all NWP models are imperfect for many reasons, including issues with initialization and assimilation, model physics, and even the observational data used to inform and verify the model. It was anticipated that the HRRR would be quite skillful compared to its peers owing to its relatively high horizontal resolution (3 km), rapid cycling, and advanced assimilation. As a popular model, the HRRR has undergone substantial verification in previous studies, although a large portion of these utilized a small number of specialized observations limited in geographic area or time (e.g., Olson et al. 2019b; Pichugina et al. 2019; Lee et al. 2019), focused on only a specific subset of atmospheric conditions or meteorological phenomenon (e.g., Pichugina et al. 2017; Pinto et al. 2015), or both. Others, such as Ikeda et al. (2013) and Griffin et al. (2017), have leveraged standardized observations over long periods of time to infer overall model biases with respect to a specific atmospheric variable (cold season precipitation and cloud cover, respectively). Similar attention has not yet been given to the PBL and especially its vertical structure.

The specific task here is to construct composite soundings from stations across the CONUS having different elevations above mean sea level (MSL). The RRS system consists of Global Positioning System (GPS) antennae and radiosonde packages, precision digital barometers, and software capable of capturing observations every second (National Weather Service 2018), corresponding to roughly 5–6-m vertical resolution. Although available starting in 2005, and seemingly ideal for PBL studies (Love and Geller 2012), it is not clear that RRS data have been utilized much in its very high-resolution form. Processed RRS data are most easily obtainable from sources such as NCAR’s Research Data Archive (as PREPBUFR files) and the Integrated Global Radiosonde Archive (IGRA) at the National Centers for Environmental Information (NCEI) in formats that facilitate analyses on isobaric surfaces, especially at the standard reporting levels (1000, 925, 850 hPa, etc.). These work rather well for temperature but less so for boundary layer winds, which are much more significantly influenced by distance above ground level (AGL). Indeed, composite wind profiles constructed in isobaric coordinates are

subjected to considerable vertical smearing when ground elevation varies among the sites, as well as potential information loss near the surface, as illustrated in Fig. 1.

Both the retention and accuracy of near-surface information in the model and observations are deemed important. An important consideration is that the HRRR outputs are made available on the model’s native vertical levels, which helps preserve simulated vertical structures near the surface. Most operational model products are only available on isobaric surfaces only, which entails the smearing as discussed above. Furthermore, we are utilizing both more complete and less processed versions of the RRS data to perform a more precise analysis of the winds near the surface in a coordinate system defined by height AGL or native levels, which are closely related in models employing a terrain-following sigma coordinate system like WRF-ARW.

NCEI’s RRS archive makes available binary files from which the 1-s radiosonde data can be extracted at various levels of processing. Unfortunately, the archive is incomplete and has been degrading in recent years; for our study period, observations were obtainable for only about three-quarters of the sites (see Fig. 2a). Even worse, regular archiving of these data ceased altogether during summer 2019, as this effort was apparently deemed to be redundant to the IGRA archive.<sup>1</sup> The IGRA data presently available are not complete in that useful information obtainable from the less-processed, higher-resolution, and timestamped observations are not included. Some of the data evaluation and analysis performed in this study would not be possible with IGRA data alone.

The structure of this paper is as follows. In section 2, we provide an extensive discussion of our data and methods, which

<sup>1</sup>“NCEI plans to discontinue DSI-6301, US and US operated station radiosonde data, on 31 May 2019. All Radiosonde data traditionally provided in DSI-6301 are now available in NCEI’s Integrated Global Radiosonde Archive (IGRA) version 2 data set.” From <https://www.ncdc.noaa.gov/data-access/weather-balloon-data>, retrieved 6 February 2020.

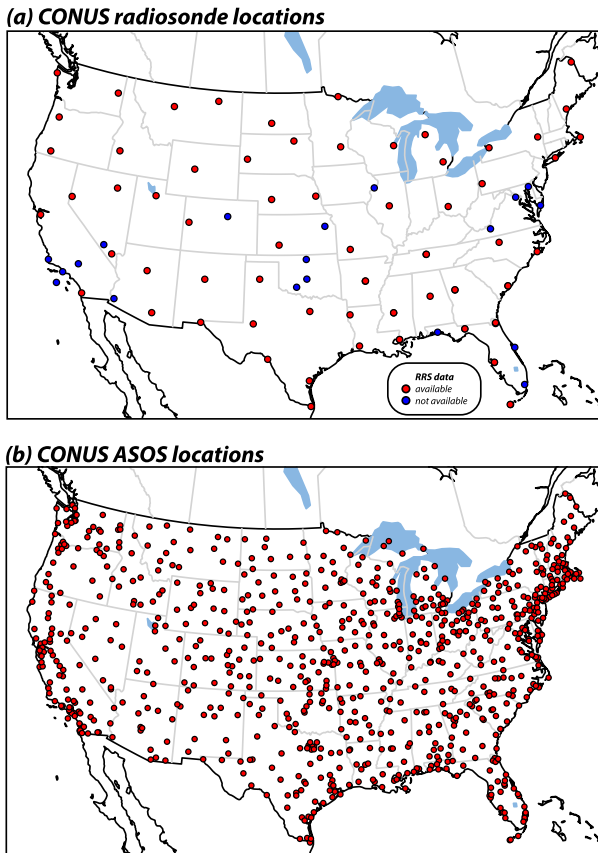


FIG. 2. CONUS maps of (a) radiosonde launch sites, with coloring indicating whether less-processed data are (at least partially) available (red) or unavailable (blue) in the NCEI RRS archive for the period between January and June 2019, inclusive, and (b) ASOS surface observation sites available for April 2019.

is important as it reveals the strengths and weaknesses of the RRS dataset as used to verify near-surface winds in particular. Section 3 presents our results, consisting of comparisons of HRRR analyses and forecasts against not only RRS observations but also near-surface information from the Automated Surface Observing System (ASOS) network. Finally, section 4 presents our discussion and conclusions.

## 2. Data and methods

The study period is 15 December 2018–15 July 2019, corresponding to the overlap period between HRRR model outputs and RRS data. Our particular focus will be on the month of April 2019 with other time periods examined as needed to demonstrate robustness of these results. Comparisons of the observational and model products described in this section were confirmed with and/or made using Model Evaluation Tools software (Bullock et al. 2017) maintained by the Developmental Testbed Center.

### a. HRRR model outputs

The HRRR model products employed herein were generated by version 3 of the ARW-based system (HRRRv3), which

used the Mellor–Yamada–Nakanishi–Niino Level 2.5 (MYNN2) PBL scheme (Nakanishi and Niino 2004), the Rapid Update Cycle land surface model (Smirnova et al. 2016), and the Rapid Radiative Transfer Model for General Circulation Models (RRTMG) radiation package (Iacono et al. 2008), among other model physics. The MYNN2 parameterization has been considerably refined in recent years (Olson et al. 2019a). HRRR simulations were initialized with RAP as described in James and Benjamin (2017) and Jankov et al. (2019). A new cycle was launched hourly, with 0000 and 1200 UTC runs integrated for 36 h, while the other runs were 18 h. James and Benjamin (2017) report that the RAP/HRRR assimilation system incorporated twice-daily radiosondes as well as 2200–2500 surface observations.

The model used 50 unequally spaced vertical levels for scalars and horizontal velocity components with a model top of 20 hPa. As noted above, model outputs were made available on the native model levels (“wrfnat” files) as well as 41 isobaric levels (“wrfprs” files) and NCEP’s NOMADS made both file types available for several days after a model run completed. After this time, the University of Utah maintains an archive of the isobaric level data (Horel and Blaylock 2019), but only for the analysis time (forecast hour 0). No publicly accessible archive of the native model level outputs is known. For this study, we employed wrfnat files verifying at 0000 and 1200 UTC that we collected daily since mid-December 2018.

### b. ASOS observations

There are about 850 ASOS stations in the CONUS (Fig. 2b) for which 1-min data are available from NCEI. These stations are generally (but not exclusively) located at airports<sup>2</sup> and measure winds (with sonic anemometers) and temperatures (with aspirated hygrothermometers) nominally at 10 and 2 m AGL, respectively. With surface observations, the term “wind” is used herein synonymously with “sustained wind,” representing an average of anemometer samples over a specified (for ASOS, 2 min) period with the largest sample during that interval representing the gust. The 1-min NCEI archive permits us to compare more precisely against on-the-hour observations that are not generally available via Meteorological Assimilation Data Ingest System (MADIS) and gives us access to the most complete gust observations. Unfortunately, the Meteorological Terminal Aviation Routine Weather Report (METAR) format employed by MADIS seriously compromises the gust record (cf. Gallagher 2016; Fovell and Gallagher 2018). From either source, the resolution of ASOS wind and gust reports is  $0.5144 \text{ m s}^{-1}$  (1 kt).

### c. RRS observations

The RRS archive at NCEI provides not only the processed wind and temperature data used in subsequent data products (such as the aforementioned PREPBUFR and IGRA datasets) but also the less refined information described below. Use of the terms “raw,” “smoothed,” “corrected,” and “processed”

<sup>2</sup> Notable exceptions include KYNC (Central Park, NY), KCQT (Downtown Los Angeles), and KDGP (Guadalupe Pass, TX).

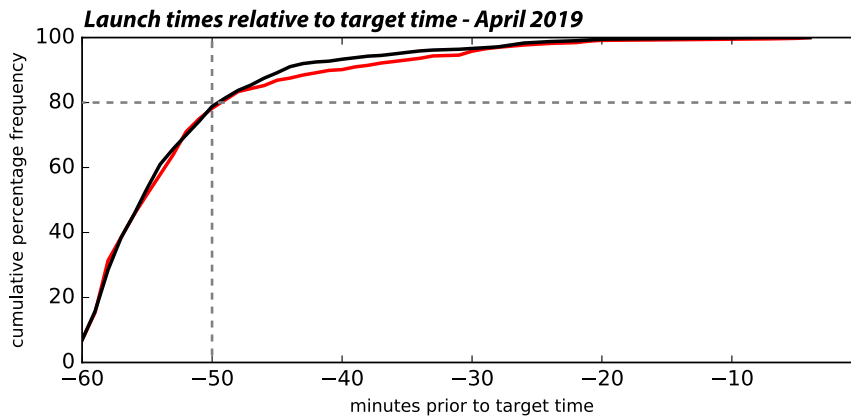


FIG. 3. Cumulative distribution function of balloon launch times relative to the nominal target times of 0000 UTC (red) and 1200 UTC (black) for radiosonde releases made at the sites shown in Fig. 2a for the month of April 2019. The dashed gray lines demonstrate that approximately 80% of all balloon launches occurred by 50 min prior to the nominal times.

herein is consistent with RRS documentation (National Weather Service 2003) and the RRS Workstation Software (RWS) manual (National Weather Service 2010). Software provided by NCEI was used to extract the following products from the archived data:

- The 3GPSu data provide raw, timestamped, 1-s horizontal wind ( $U$  and  $V$ ) components, along with GPS latitude, longitude, and geometric height MSL for the balloon location.
- The 4GPSs product provides smoothed versions of the 3GPSu data for the same timestamps. According to National Weather Service (2003) the 4GPSs data “may represent an average of several values over a period sample (vendor specific), thus allowing for the removal of random and systematic instrument noise. [These] data [are] smoothed in the SPS [sounding processing system] to remove the effects of erratic and pendular motion.” Smoothing algorithms are vendor-specific and proprietary.
- The 2rPTU dataset provides raw temperature, pressure, and moisture information at 1-s intervals, with timestamps that are shifted from those in 3GPSu/4GPSs by a fraction of a second.
- The 5pPTU product is the processed information, including corrected pressure, corrected and smoothed pressure, smoothed uncorrected and corrected temperature and moisture, and derived geopotential height for the same timestamps as indicated in 2rPTU. During this step, the initial ( $T_0$ ) observation of the smoothed and uncorrected temperature is replaced with a contemporaneous surface observation taken nearby (National Weather Service 2010). After that replacement, proprietary corrections are applied, which include adjustments for radiation.
- The 6pGPS wind dataset is processed by RWS and “arrived at by applying normalization, correction, smoothing, outlier removal, and data plausibility checks to the raw GPS ‘smoothed wind’ [4GPSs] data provided by the radiosonde” (National Weather Service 2003). As with temperature, the  $T_0$  observation is replaced with a contemporaneous surface

measurement. This is then connected to the radiosonde observation sequence via linear interpolation (National Weather Service 2010), with the consequences discussed below. Timestamps have been reconciled with those in the 5pPTU dataset.

- The 7Lvls product joins the processed wind and PTU information, providing timestamped “mandatory, significant, and other special winds and PTU levels” (National Weather Service 2003) at irregular intervals arrived at as described in National Weather Service (2010), including wind speed and direction, geopotential and geometric heights MSL, and dewpoint and relative humidity. Data for standard isobaric levels are created via interpolation if necessary. Balloon latitude and longitude information, however, is missing.

The processed IGRA and PREPBUFR observations are subsampled from the 7Lvls information, so many levels are missing as well as all observation record timestamps and geometric heights.

#### d. RRS data issues and handling

Our analysis made use of the processed, 1-s 5pPTU temperature, pressure and geopotential height information and the 6pGPS wind components, from which wind speed was computed. (The 7Lvls information was not used owing to the irregular time intervals among records.) However, the 3GPSu wind components and geometric heights were also retained for each record, for the reasons discussed presently. The 3GPSu and 6pGPS/5pPTU datasets were merged by record number and timestamps were monitored for anomalies.

Utilizing radiosonde data requires addressing a number of obstacles and limitations. First, it is important to recognize that nearly all balloon releases in the CONUS occur *before* the nominal observations times of 0000 and 1200 UTC (cf. Coniglio et al. 2013; Evans et al. 2018). In April 2019, for instance, about 80% of launches had occurred by 50 min prior to the nominal times (Fig. 3), which meant that the instrument packages were transiting the boundary layer earlier than might

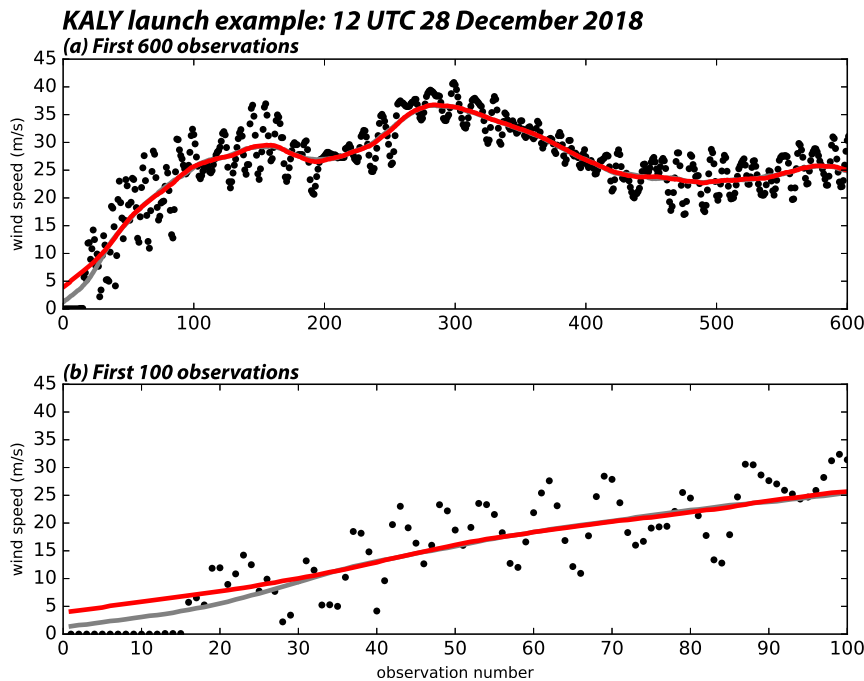


FIG. 4. Radiosonde wind speed observations from an example launch at Albany, NY (KALY), for 1200 UTC 28 Dec 2018 showing the first (a) 600 and (b) 100 observations, representing 10 min and 100 s of record, respectively. Shown are the 3GPSu raw wind values (black dots), 4GPSs smoothed observations (gray line), and 6pGPS processed observations (red line).

be anticipated. This means 23-h forecasts from 0000 to 1200 UTC may represent more faithful comparisons to the observations than their 24-h counterparts. This is particularly relevant to temperature near the ground, which may be changing rapidly around release times.

Other issues can directly and negatively affect the usefulness of RRS data recorded near the surface. Although not shown, geometric and geopotential MSL height estimates provided in the processed RRS can disagree by sizable amounts.<sup>3</sup> We adopted the pressure-derived height estimates for our analysis owing to concerns regarding GPS estimates close to the surface (Dirksen et al. 2014), although analyses repeated using geometric heights revealed no important differences in basic results or fundamental conclusions (not shown). Additionally, the RRS observation metadata supplied MSL-relative release point and barometer elevations but we need AGL heights, which could be a problem if the release point is not actually at the local surface. We determined the surface MSL height at each reported release point via using Google Earth and compared them to the height specified in the metadata. In some obvious cases, such as Albany, New York (KALY, 72518); Tucson, Arizona (KTWC, 72274); and Tallahassee, Florida

(KTAE, 72214), the launches clearly took place from tall rooftops so even the first record represented a sizable distance above the actual ground. The release points at most sites, typically associated with domes mounted on low-rise structures, appeared to be several meters above the local terrain, representing height adjustments between 0 and 3 m. These adjustments were implemented but ultimately determined to be unimportant owing to the most serious issue described next.

We unfortunately conclude that a combination of ostensible prerelease information, surface data replacement, smoothing, and linear interpolation precludes the use of observations within about 250 m of the launch height, at least for wind speeds. The aforementioned pendular motion is readily apparent in the raw wind (3GPSu) observations, as illustrated in Fig. 4a, representing a nominal 1200 UTC rooftop release from KALY from December 2018. While the variation of the raw values about the smoothed (4GPSs) and processed (6pGPS) wind speed data is large—up to about  $10 \text{ m s}^{-1}$ —one might expect further aggregation of wind data across sites and times could mitigate residual pendular and erratic effects in the processed winds. However, in some cases, including the present one, the raw observations suggest that ostensible prerelease information has contaminated the near-surface winds. Note that for the first 15 records (= 15 s), the raw wind speeds were precisely zero (Fig. 4b), and the GPS position was also constant during this interval (not shown), suggesting the balloon had not yet been launched. Yet, smoothing within the SPS introduced spurious positive readings into the 4 GPSs series,

<sup>3</sup> For April 2019, over all launches, the mean difference between geometric and geopotential heights in the lowest 1 km AGL was close to zero but the 95th percentile of absolute height differences was  $\approx 16$  m.



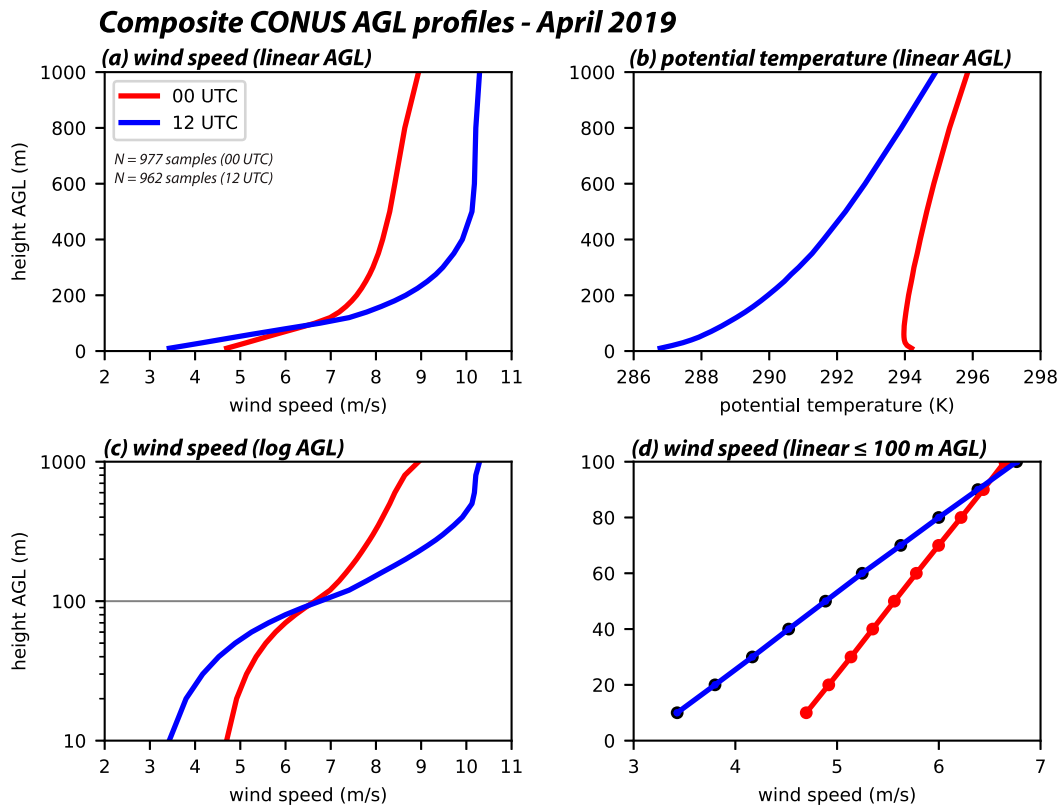


FIG. 5. Composite profiles averaged on fixed AGL heights (10-m interval) and over 0000 UTC (red) and 1200 UTC (blue) releases for available RRS sites for the month of April 2019, showing processed (a) wind speed (6pGPS); (b) potential temperature (5pPTU); (c) wind speed as in (a), but plotted on a log height axis; and (d) wind speed as in (a), but for the lowest 100 m AGL. Dots in (d) indicate the AGL heights on which the profiles are averaged.

shown in gray. As described above, the 6pGPS (red curve) values were subsequently forged by replacing the first (T0) record with the nearby surface observation and linearly interpolating back to the 4GPSs data, possibly with further smoothing applied. In the present example, the 6pGPS and 4GPSs wind speeds come into near agreement after about 35 s, at which time the balloon had ascended roughly 100 m from the launch point (or about 120 m AGL at this rooftop site).

With respect to wind information, this interpolation establishes a false linear profile in a layer in which a stability-modulated logarithmic profile is anticipated. Figure 5 presents composite profiles of processed wind speed and potential temperature (6pGPS and 5pPTU), aggregated on fixed AGL heights over available RRS sites for the month of April 2019. Apparent prelaunch data had already been identified and removed, resulting in a small reduction of the observation count at many of the sites. Furthermore, to avoid profile distortion, only releases that had valid wind and temperature data for all height levels between 10 and 1000 m AGL were considered, so only 56 of the 60 RRS sites contributed to a total of about 970 samples at each level for each time.

At first glance, the profiles are reasonable. As expected, there was more vertical shear as well as weaker winds near the surface at 1200 UTC compared to 0000 UTC (Fig. 5a) and a

slightly superadiabatic layer survived in the afternoon/evening composite (Fig. 5b) while the morning profile was strongly stable. However, note that the wind profiles were precisely linear below 100 m AGL at both times (Figs. 5c,d). A linear profile appeared at nearly every station as revealed in Fig. 6, which presents temporally averaged processed wind profiles for individual sites (with the station composite superposed) as a function of observation record number. A strong tendency is present for the profiles to be linear below a sharp change in slope<sup>4</sup> at about record number 23, or 22 s after nominal launch, at which time the sondes were about 130 m above their launch points based on average altitude measures. At 0000 UTC, the near-surface slope not only varied among the stations but also was well correlated ( $r \approx 0.8$ ) with longitude, thereby reflecting local time within the diurnal cycle (not shown). This suggests some potentially useful information may be present, but it remains that the linearity is forced and spurious.

RWS documentation (National Weather Service 2010) states that the “first good block at or after Surface will contain

<sup>4</sup> These include all 6pGPS observations, including those affected by suspected prelaunch values. Exclusion of those suspect observations had little material effect on this plot (not shown).

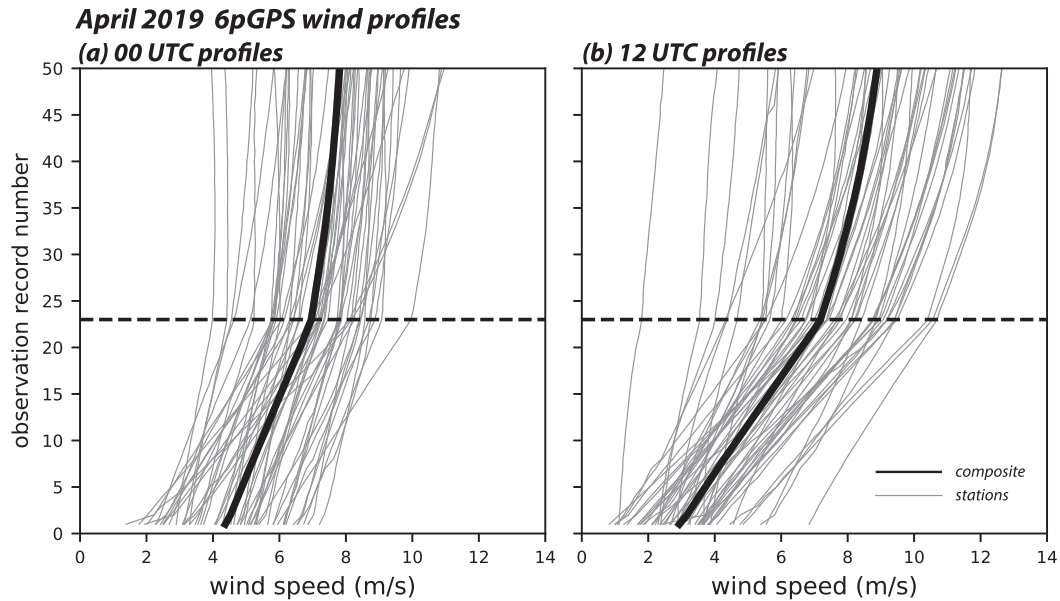


FIG. 6. Processed wind speed profiles (6pGPS) for the first 50 observational records (each separated by one second) from each radiosonde site, temporally averaged over the month of April 2019 (gray lines) for (a) 0000 UTC and (b) 1200 UTC, along with the 51station composite (black line). The dashed black line indicates observation number 23, highlighting the typical record at which the observed linear slope changed.

45-or-more smoothed  $U/V$  points.” Given the standard, centered smoothing window is 45 points, the first observation that could be beyond the forced linearity layer would be at or near record number 46, which based on average ascent corresponds to approximately 250–260 m AGL. Based on temporally and spatially averaged conditions, this falls between HRRR model levels 4 and 5, so we will pay more attention to model-observation comparisons farther aloft. It is possible that given access to the raw (3GPSu) observations, a more careful reconstruction of surface layer winds could be pursued, particularly with respect to removing prelaunch data, as long as the substantial pendular motion is also addressed [perhaps as in Wade et al. (2018)].

RWS documentation is less detailed regarding temperature handling near the ground, other than to indicate the T0 observation is also replaced by an externally supplied value and smoothing involved fewer (9) data points. Inspection of raw and processed temperature records suggest that near-surface readings may be less corrupted than for winds, but information in the lowest levels will be treated with suspicion anyway.

#### e. Merging HRRR and RRS information

Our analysis commences with HRRR model information being interpolated to the observation sites. Next, for RRS data, the observations were interpolated to the HRRR scalar/horizontal wind model levels at each individual site and for each launch, and then aggregated across both sites and times to create composites at the 0000 and 1200 UTC target times. The composites were then plotted against heights representing the temporal and spatial means for each model level. As the HRRR employs a mass coordinate that is affected by temperature, the local heights of model levels vary across space and also with time at individual locations. For April 2019,

the first HRRR model level resided at a mean local altitude of 10.5 m (standard deviation  $\sigma = 0.3$  m), and eight levels were positioned within the lowest 1 km AGL. Once released, radiosonde balloons can drift away horizontally with the wind, which is a concern because displacements of tens of kilometers, representing many grid points, can occur by the time the upper troposphere is reached (Coniglio 2012). Drift has been ignored as our interest is primarily in the lowest 1 km AGL, which can be traversed in about 3 min.

Although the HRRR employs relatively high resolution, there are systematic discrepancies between model and actual site altitudes. The true site elevation, as determined by the launch point, averaged about 10 m higher than the corresponding model location, and ranged between +100 and –60 m but with 46 of the 60 sites having absolute errors < 14 m. As can be anticipated, MSL elevation discrepancies are particularly obvious in temperature and pressure comparisons, which is one reason why changes in the former with forecast time is emphasized. No important differences were noted when sites with larger (>14 m) altitude discrepancies were retained or excluded, so by default all sites were employed in our analyses.

Two issues need to be addressed at the outset. First, *the radiosonde data we are examining are assimilated into the HRRR analyses* for the 0000 and 1200 UTC cycles we are examining. Thus, for the 0-h forecast, the ostensible forecast bias should be small because the observations and analyses are not truly independent. A superior manner of assessing the analysis quality would be to employ soundings not included in or withheld from the assimilation system (e.g., Coniglio 2012). Second, radiosonde observations themselves naturally have errors. Benjamin et al. (2004), for example, presumed errors for pre-RRS soundings of 0.5 K for temperature [in agreement

with, e.g., Zeng et al. (2019) for RRS] and  $2.5\text{--}4\text{ m s}^{-1}$  for wind speed, and the preceding discussion could be used to argue for still larger inaccuracies in the latter. As a consequence of these two issues, note that our primary goal is to compare 24-h forecasts to their corresponding analyses, valid for the same times, and to composite over both space and time, in an attempt to extract systematic issues that may (and do) exist. This still involves the implicit assumption that the analyses are more accurate than the forecasts, which cannot be conclusively proven without independent information. However, it can also reveal modeling system issues, which may be due at least in part to the physical parameterizations employed.

### 3. Results

In this section, we present an examination of HRRR analyses and forecasts from the 0000 and 1200 UTC cycles, commencing with a comparison to surface observations from the ASOS network. Although 0000 and 1200 UTC HRRRv3 cycles extended to forecast hour 36, we will focus on the first 24-h period. Our current interest is in forecasts of wind speed and temperature (potential temperature for the radiosondes).

#### a. Surface verification

Figure 7 presents comparisons of ASOS observed and HRRR forecasted 10-m wind speeds and 2-m temperatures for the 0000 and 1200 UTC cycle runs, not only averaged across the ASOS network but also aggregated over all of April 2019, thereby smoothing away weather events and leaving behind the diurnal cycles of these variables smeared over four time zones. As noted above, instantaneous hourly wind and temperature forecasts were compared to top-of-the-hour observations, but we note that using mean observations over a  $\pm 6$ -min window instead (not shown) altered neither our results nor conclusions. We required stations to have at least 90% data availability for the month, resulting in 733 stations being retained in the April wind assessment. Owing to the outsized influence of altitude, the temperature analysis also required the HRRR absolute elevation discrepancy to be less than 80 m, which left 630 locations. For each hour, the mean and standard deviation of  $\approx 19700$  samples are shown, the latter being necessarily large owing to the temporal and areal extent of the observations.

Evidently, the HRRR has substantial skill in reproducing the diurnal variations. Additionally, calm observations, here defined as values  $< 1\text{ kt}$  owing to ASOS' precision, were roughly equally common in the observations (0.6%) and the forecasts (1.5%).<sup>5</sup> However, for both 0000 and 1200 UTC cycles, the HRRR had a negative forecast wind bias—defined as forecast minus observation—of about  $-0.5\text{ m s}^{-1}$  at the analysis time (0-h forecast) although most of that disappeared over the first 24 h (Figs. 7a,b). For temperature, the 1200 UTC cycle started with a negative bias of about  $-0.4\text{ K}$  that vanished even more rapidly (Fig. 7d). The magnitude and evolution of these errors for other months examined were comparable (not shown).

<sup>5</sup> Had ASOS METAR reports been used instead, calm observations would have been more frequent, owing to the METAR reporting rules (cf. Fovell and Gallagher 2018).

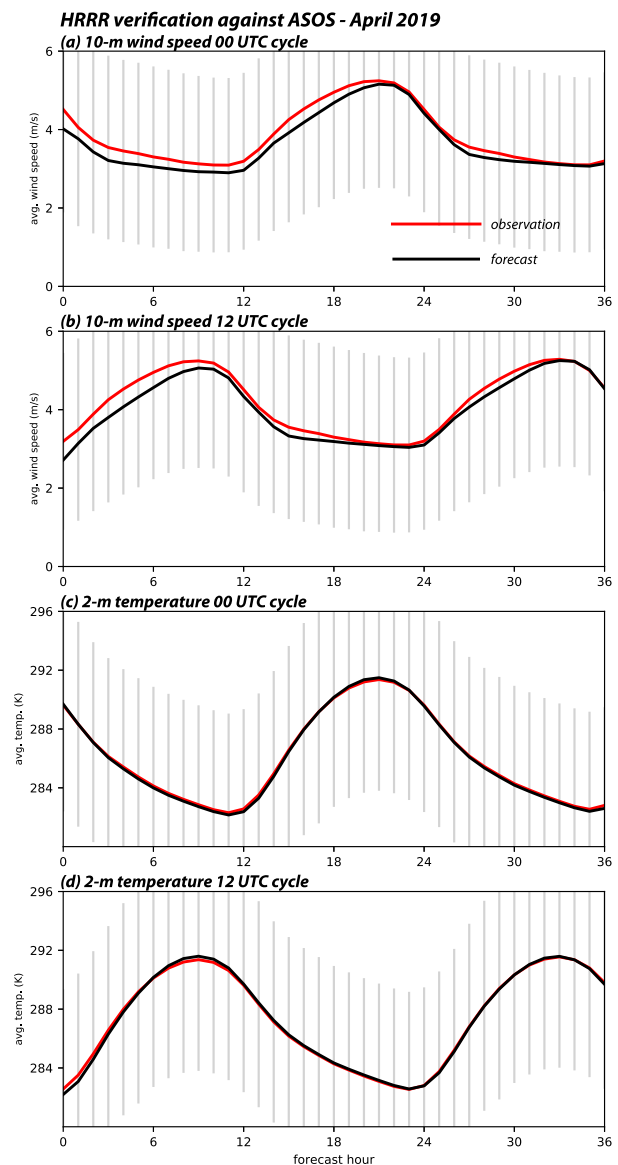


FIG. 7. ASOS observations (red) and HRRR forecasts (black), averaged spatially across the ASOS network and temporally over the month of April 2019, of (a),(b) 10-m wind speeds and (c),(d) 2-m temperatures for every hour of the (a),(c) 0000 UTC and (b),(d) 1200 UTC forecast cycles. The vertical gray bars denote the standard deviation of the averaged observations.

The analysis of individual stations commences with Fig. 8, plots of forecasted versus observed winds averaged over the month of April 2019 on which each dot is a station. Incorporating all forecast hours, including the analysis time, yielded about 1060 forecast–observation comparisons per retained site for this month.<sup>6</sup> Aggregating over all hours

<sup>6</sup> As the 36-h runs overlapped in time, most observations were employed in more than one verification.



**HRRR 10-m wind speed verification against ASOS - April 2019**

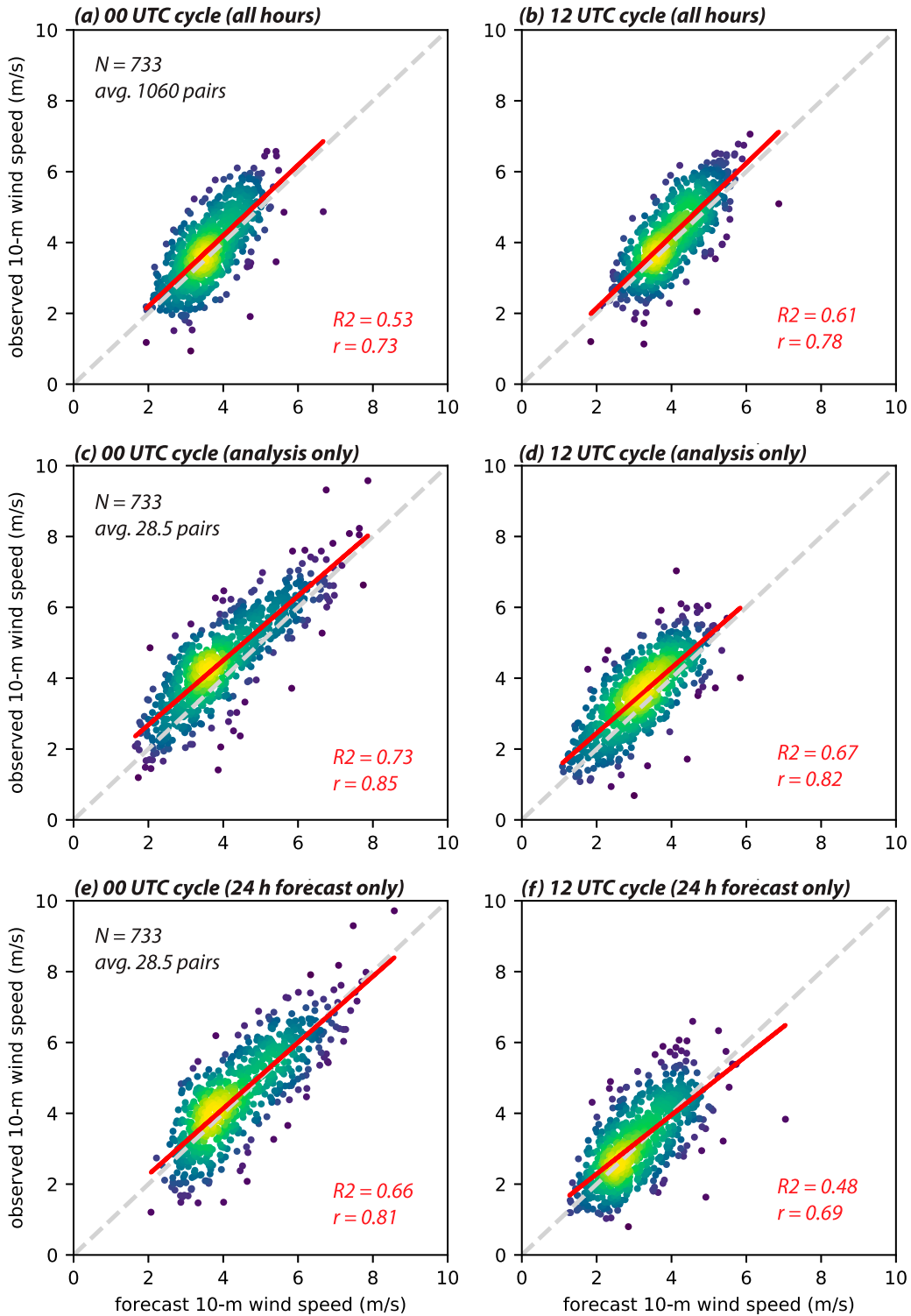


FIG. 8. Forecasted vs observed wind speeds averaged over April 2019 for (a) all 0000 UTC cycle forecast hours, (b) all 1200 UTC cycle forecast hours, (c) 0000 UTC analysis times, (d) 1200 UTC analysis times, (e) 24-h forecasts from 0000 UTC cycles, and (f) 12-h forecast from 1200 UTC cycles. Each dot is an individual station color coded by the density of neighboring points (cooler colors are less dense, warmer colors are more dense). Also shown are the linear regression fits (solid red lines) and one-to-one correspondences (dashed gray lines).

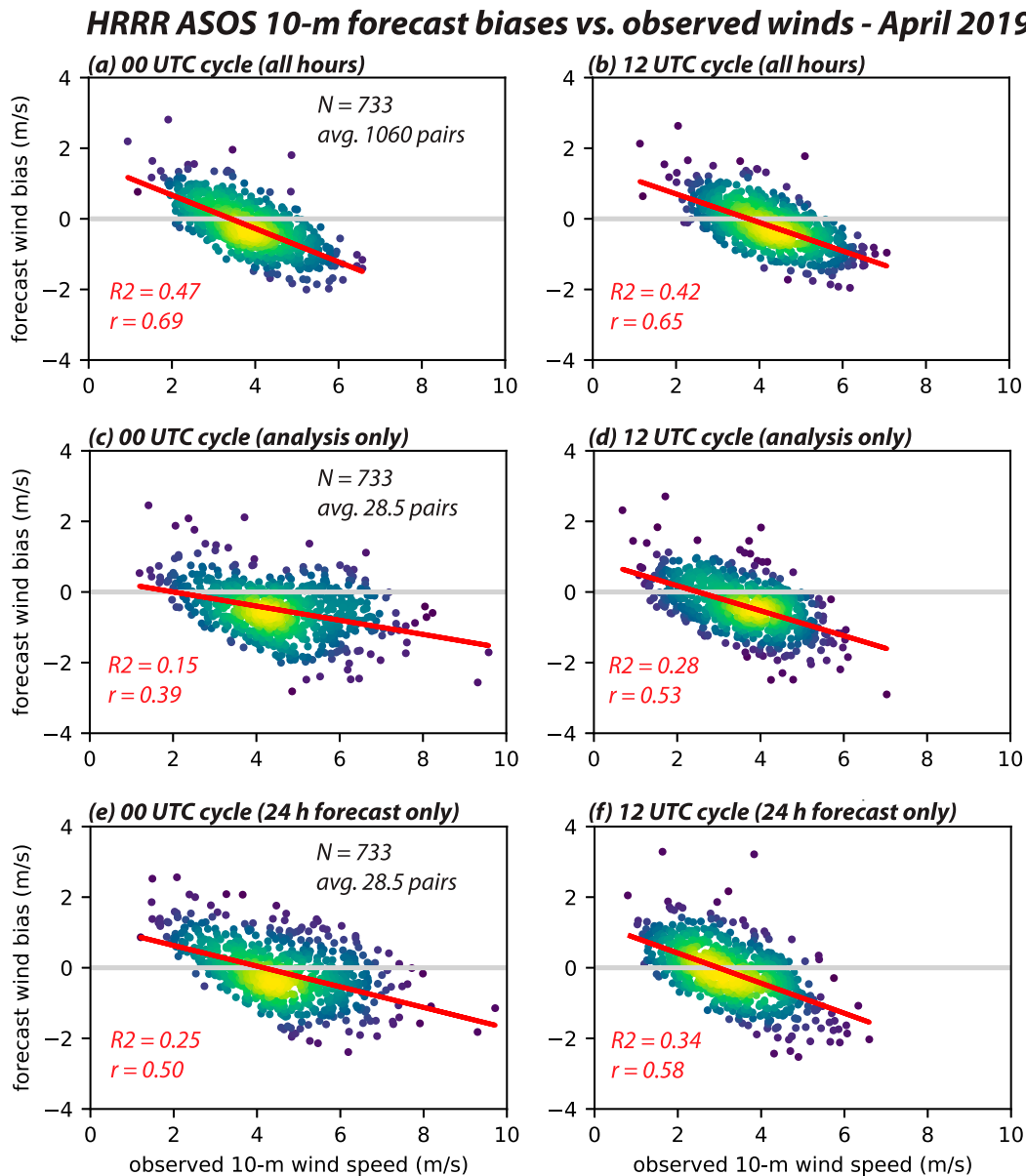


FIG. 9. As in Fig. 8, but comparing temporal averages of observed wind speed to forecast bias for ASOS stations. Panels are representative of the same forecast cycles and forecast hours as in Fig. 8, and dots still represent individual stations color coded by the density of nearby points. Zero bias lines are shown in light gray.

(Figs. 8a,b), there is overall skill (coefficient of variation  $R^2 \approx 0.53$ ) and a small ( $-0.19 \text{ m s}^{-1}$ ) net bias but also some substantial scatter about the 1:1 correspondence line (dashed).<sup>7</sup> In previous work focusing in forecast verification during Southern California “Santa Ana” wind events (cf. Cao and Fovell 2016; Fovell and Cao 2017; Cao and Fovell 2018; Fovell and Gallagher 2018), forecast wind bias was found to be

<sup>7</sup> Some stations possess issues, such as with their land-use assignments, but these are not influencing the results or conclusions.

negatively correlated with average observed wind speed, meaning that sites characterized by slower observed winds were systematically more likely to be overpredicted while windier sites were underestimated. Comparable results were obtained in this study (Figs. 9a,b). Also similar to Cao and Fovell (2018), albeit not shown here, the forecast bias was uncorrelated with average forecast wind speed but was positively associated with gustiness in the form of the gust factor (gust divided by sustained wind), which was interpreted by Cao and Fovell (2018) as conveying information about site exposure.

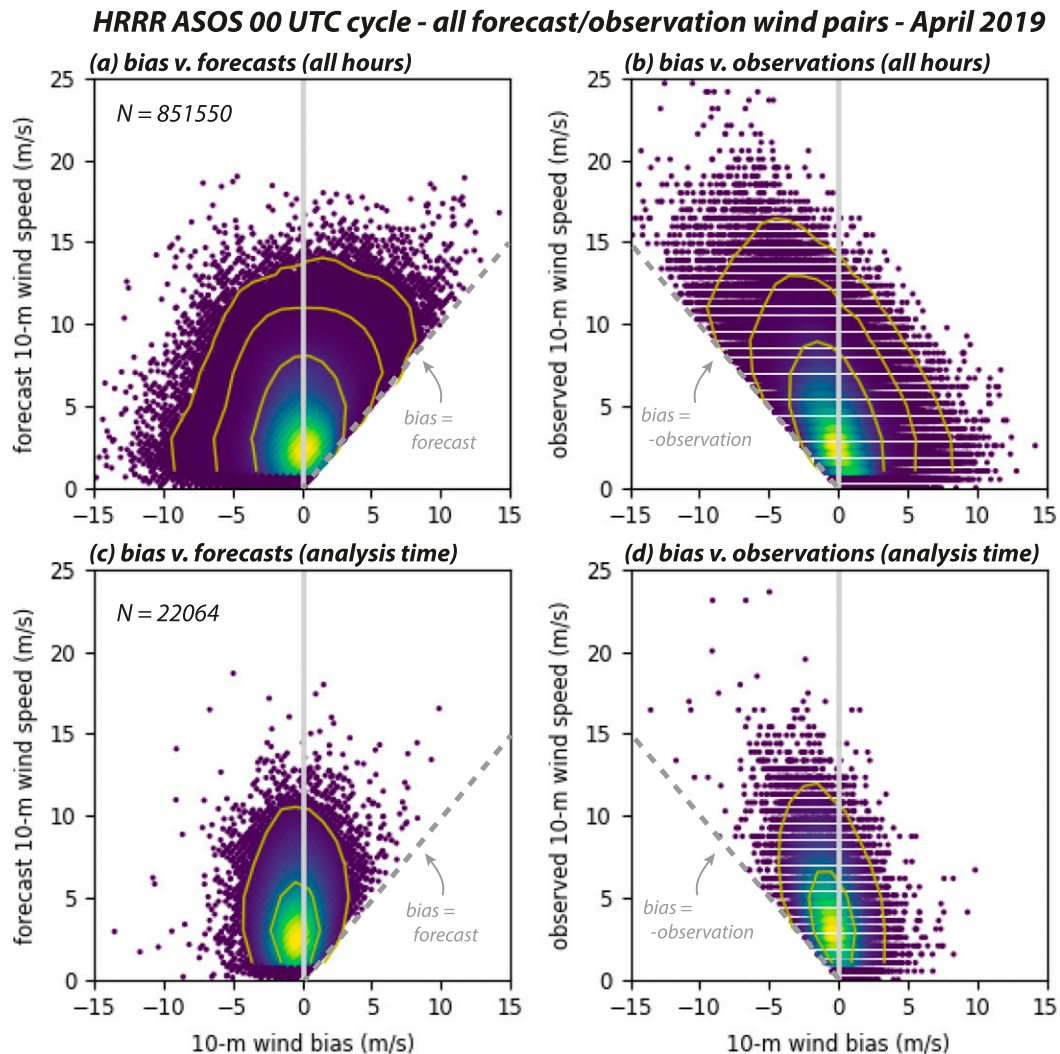


FIG. 10. Forecast wind bias vs (a),(c) forecasted 10-m wind speed and (b),(d) observed 10-m wind speed for individual forecast observation pairings during April 2019 from (top) all forecast hours or (bottom) only for the analysis times. As in Figs. 8 and 9, the color coding indicates point neighborhood density. Additionally, frequency contours of 100, 1000, and 10 000 have been plotted and smoothed, introducing some artifacts. The diagonal dashed lines represent bounds represented by zero (calm) observations in (a) and zero (calm) forecasts in (b). The resolution of ASOS wind speed observations is  $0.51 \text{ m s}^{-1}$  (1 kt).

Figures 8 and 9 also examine skill and bias at the analysis and 24-h forecast times, comparisons that necessarily involved far fewer samples. As noted above, new HRRR runs are initialized with previously generated forecasts combined with data assimilation incorporating many, many surface observations. The imperfect relationship between observed and modeled ASOS winds at the analysis time (Figs. 8c,d) may reflect biases that were inherited from previous forecasts and suppressed, but not completely eliminated, via assimilation (cf. the  $R^2$  values). Another potentially important factor is the influence of non-ASOS information in the assimilation. Note the average station bias was markedly more *negative* (at  $-0.5 \text{ m s}^{-1}$ ) at the analysis time, a likely consequence of the assimilation and consistent with the slow bias seen at

hour 0 in Figs. 7a and 7b. As in the ASOS analysis, the 24 h forecast time scatterplots (Figs. 8e,f) reveal the negative bias disappeared with time.

The tendency for slower wind stations to be overpredicted may reflect, in part, an inherent characteristic of the wind speed bias ( $b = f - o$ ), where  $f$  and  $o$  are the forecasted and observed values, respectively. Because wind speeds are necessarily nonnegative, bias is capped by  $f$  when the observed wind is calm and bounded by  $-o$  as the forecasted wind approaches zero. In particular, the latter may force positive biases in less windy instances. These hard constraints can be seen in Figs. 10a and 10b, which incorporate every forecast–observation pair (over 850 000 examples) from the HRRR 0000 UTC cycle runs from April 2019 and present wind speed bias plotted against

### HRRR 10-m wind station ranks - April 2019

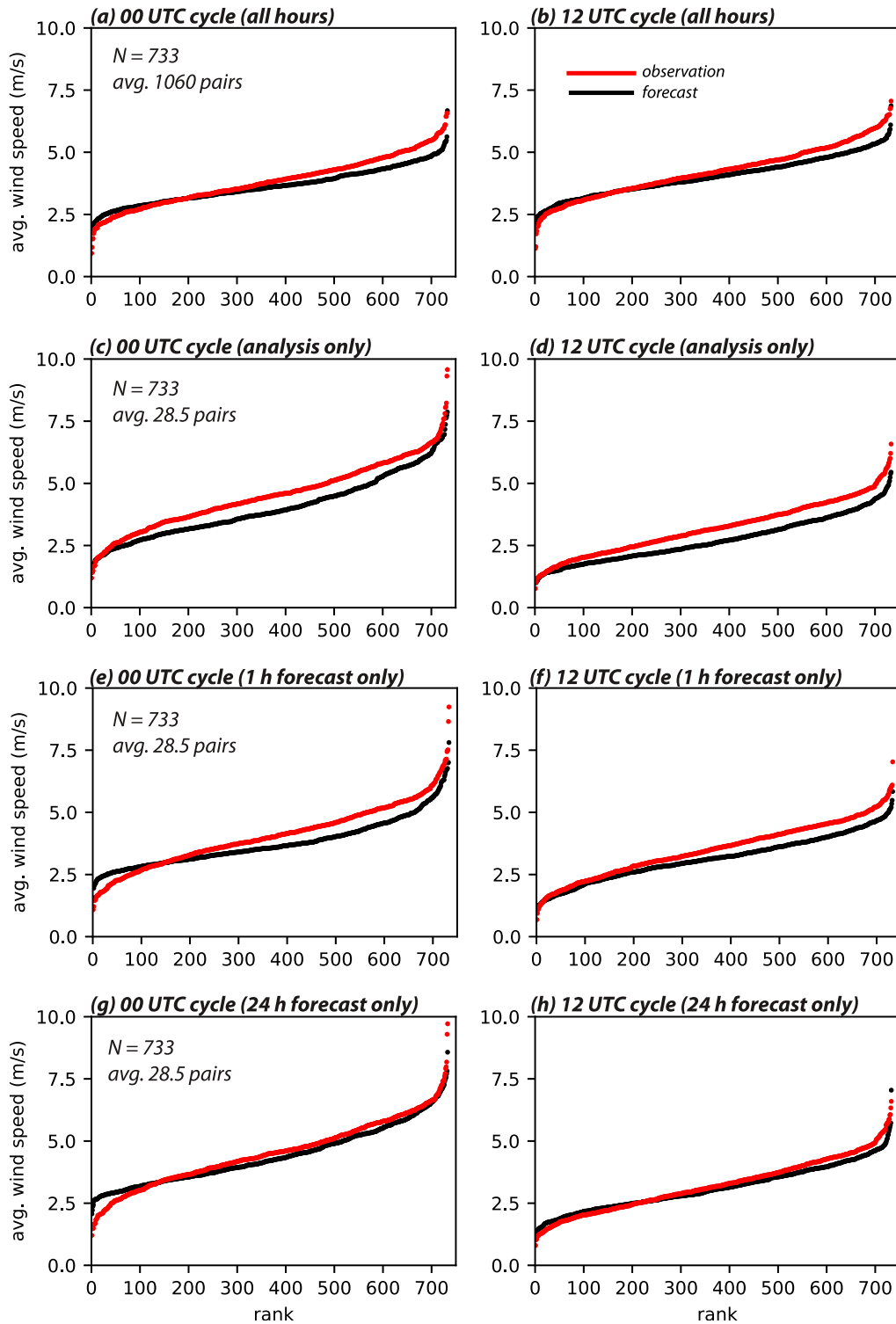


FIG. 11. ASOS stations ranked separately according to their temporally averaged observed (red) and forecasted (black) wind speeds for April 2019 for (left) 0000 UTC and (right) 1200 UTC forecast cycles using (a),(b) all forecast hours; (c),(d) analysis times only; (e),(f) the first forecast hour; and (g),(h) forecast hour 24.

### HRRR 2-m temperature verification against ASOS - April 2019

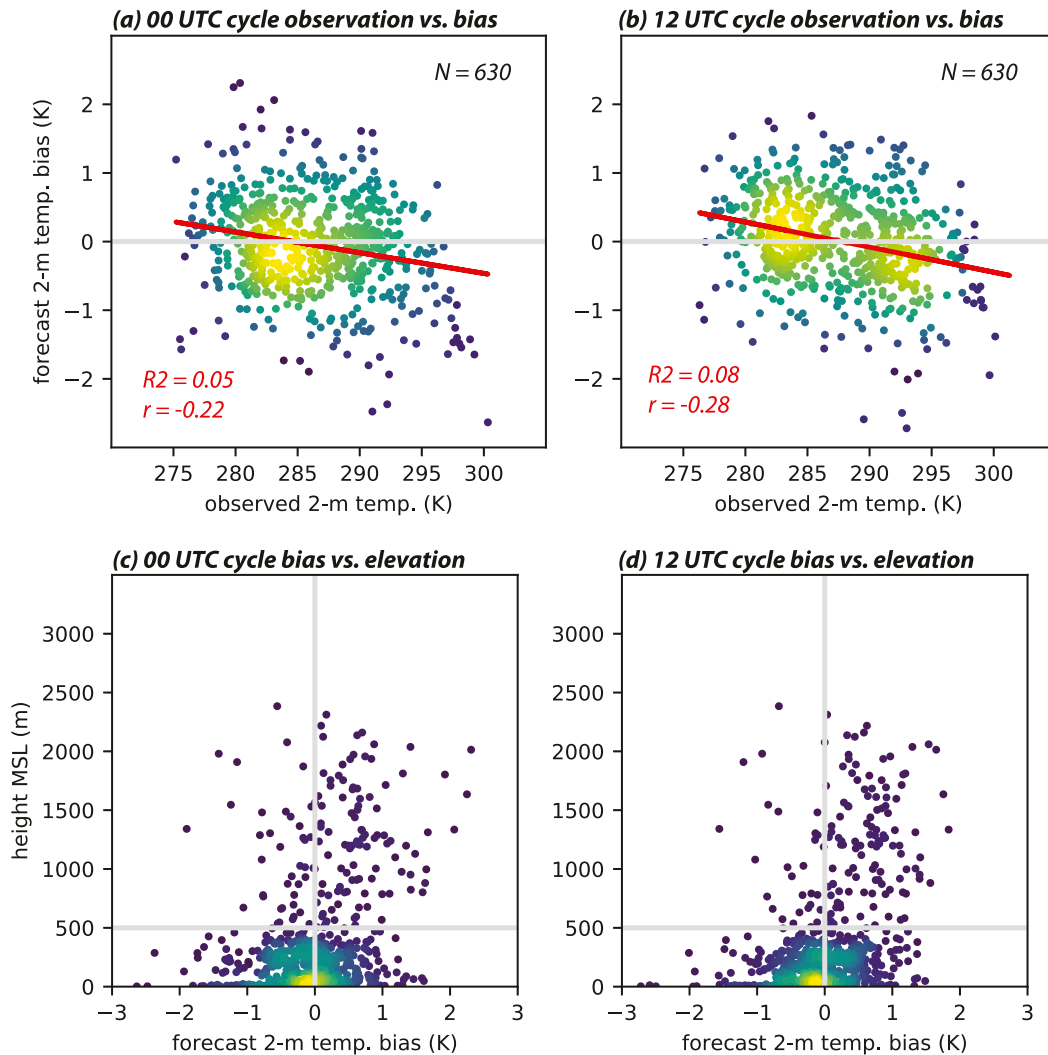


FIG. 12. As in Fig. 9, but comparing observed 2-m temperatures and forecast biases for the (a) 0000 UTC and (b) 1200 UTC forecast cycles. Again, each dot is a station averaged over April 2019 and color coded by nearby point density and least squares fits are shown (red lines). (c),(d) Forecast bias and station elevation are compared for the 0000 and 1200 UTC cycles, respectively, also color coded by neighborhood density.

the forecasted and observed values, color coded and contoured by local point density.<sup>8</sup> Despite these constraints, the plots reveal that it is not the forecasts that are biased, as the contours in Fig. 10a help emphasize the most likely error for a given forecast of nearly any speed was zero. Note further that the distribution of forecast winds was nearly random about a mean bias value of zero and that although a sector was missing owing to the  $b < f$  constraint, its impact on the distribution of biases versus forecasted wind speeds was small owing to the very high concentration of bias values near zero.

<sup>8</sup>The finite resolution ( $0.5144 \text{ m s}^{-1}$ ) of the ASOS observations is readily seen in Fig. 10b. Coarsening the forecasts to the same resolution had virtually no impact on these results.

In pointed contrast, the distribution of observations versus biases (Fig. 10b) tilted strongly toward the left (i.e., negative bias) with observed wind speed. For less windy events, the mean bias was indeed positive (averaging  $0.85 \text{ m s}^{-1}$  for  $o < 2.5 \text{ m s}^{-1}$ ) as suggested by this figure, but this quickly shifted to underprediction as the measured winds increased. Observed winds exceeded  $2.5$  and  $5 \text{ m s}^{-1}$  67% and 29% of the time, respectively, and mean biases in those situations were  $-0.70$  and  $-1.38 \text{ m s}^{-1}$  when averaged over all forecast hours. Although this result is not completely understood, and may reflect something inherent in sonic anemometers, it remains that the higher the observed wind speed was, the more likely the corresponding HRRR forecast was too slow and this should be recognized when using HRRR-forecasted winds. These findings also held when only the analysis time pairs were considered (Figs. 10c,d).



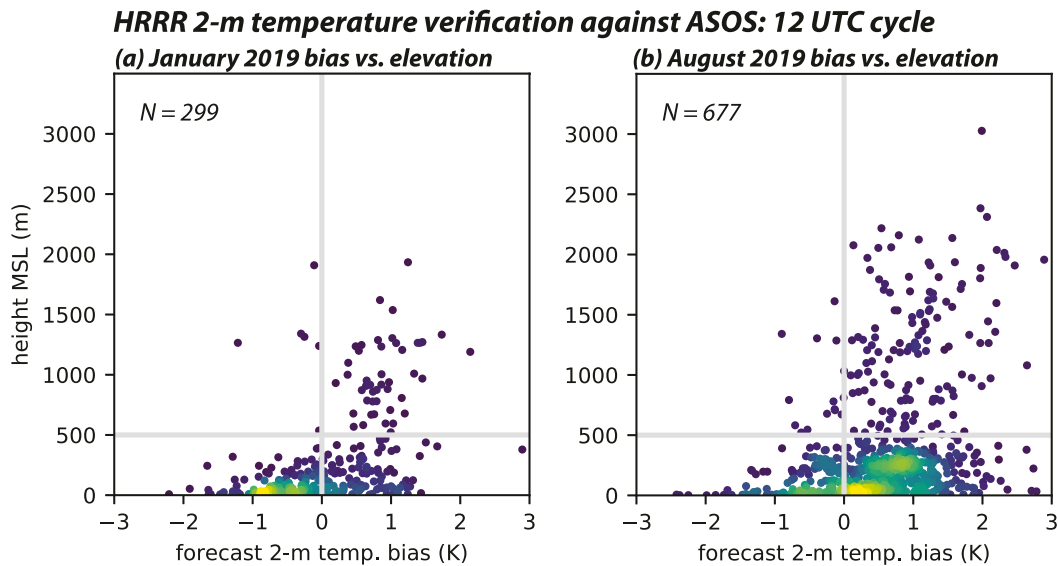


FIG. 13. As in Figs. 12c and 12d, but for January and August 2019, respectively. Fewer stations satisfied the data completeness requirement in January owing to missing observations.

Another way of considering the station-aggregated data is presented in Fig. 11. Forecasts and observations utilized in Fig. 8 were sorted *separately*, ranked by increasing magnitude, and then superimposed, permitting a visual comparison of relative wind speed distributions. Aggregated over all hours (Figs. 11a,b), the forecast winds were clearly faster than the observations among the lower rankings, but were slower beyond ranks 195 and 173 in the 0000 and 1200 UTC runs, representing about 73% and 76% of total stations, respectively. This is why the net bias was negative, albeit by a small amount, when averaged over all possible pairs. Interestingly, at the analysis time, only 11 and 1 ranks had faster forecasted than observed winds at 0000 and 1200 UTC (Figs. 11c,d), establishing a marked gap between the two distributions, and yet a sizable fraction of that discrepancy disappeared within the very first forecast hour (Figs. 11e,f). This may again suggest the initial negative bias was a consequence of the data assimilation, but the supplied information was not retained very long, especially in the 0000 UTC runs that were more likely to commence with a convective boundary layer. By the 24-h forecast time (Figs. 11g,h), the forecasted and observed wind speed distributions were very similar, apart from the very lowest ranks in the 0000 UTC cycle.

Finally, with respect to 2-m temperature, Fig. 12 presents average observed temperature plotted against mean forecast bias for April 2019. Again, each dot is a station. At both 0000 and 1200 UTC, temperature biases were negatively but quite weakly (correlations  $r = -0.22$  and  $-0.28$ , respectively) associated with observed temperature (Figs. 12a,b). Elevation error had essentially no association ( $r \approx 0.08$ ) with temperature bias (not shown), in part because the sites with the largest elevation discrepancies had already been removed. Additionally, the altitude discrepancies that remained were not correlated with elevation itself (also not shown).

When plotted against station elevation above MSL, however, a more striking association emerged (Figs. 12c,d). Only 23% of the

ASOS stations in this dataset resided above 500 m MSL but, for that subset, about 75% exhibited positive forecast biases while only 41% did so below that altitude. (Note again that sites with large elevation errors have already been excluded.) Furthermore, the 1200 UTC cycle runs from January and August in Fig. 13 demonstrate that the high altitude site temperature bias is neither just a spring season issue nor a simple function of snow cover, as might be surmised. For those two months, 89% of stations above 500 m MSL exhibited positive biases, and while the average bias for stations below 500 m shifted among the three months examined, it remains that the positive bias seen at higher altitude sites persisted across seasons. This issue may reflect problems with the land surface characterization, including (and not limited to) albedo, vegetation, slope effects, radiation and/or cloud cover, and/or issues with resolution (such as unresolved valleys and terrain flows) owing to differences between actual and modeled terrain, and deserves closer inspection.

#### b. Boundary layer verification

Wind speed profiles averaged across 60 radiosonde sites for April 2019 are presented in Fig. 14. The gray horizontal dashed lines indicate the level below which the aforementioned near-surface data manipulation renders the observed wind data suspect. Owing to early releases relative to the target times as discussed above, the forecast at 23 h (dashed green) may provide a more valid comparison to the observed profile (red) than that at 24 h (solid green). While the same valid times (consisting of 24 launches each for 0000 and 1200 UTC) were used to construct the observation and model profiles, the number of samples employed in profile construction varied somewhat with height owing to rooftop releases and occasional missing or invalid data. Above the gray line, the average number of comparisons was 1360 and 1387 at 0000 and 1200 UTC, respectively. The lowest model level, being close to 10 m AGL, utilized about 856 samples.

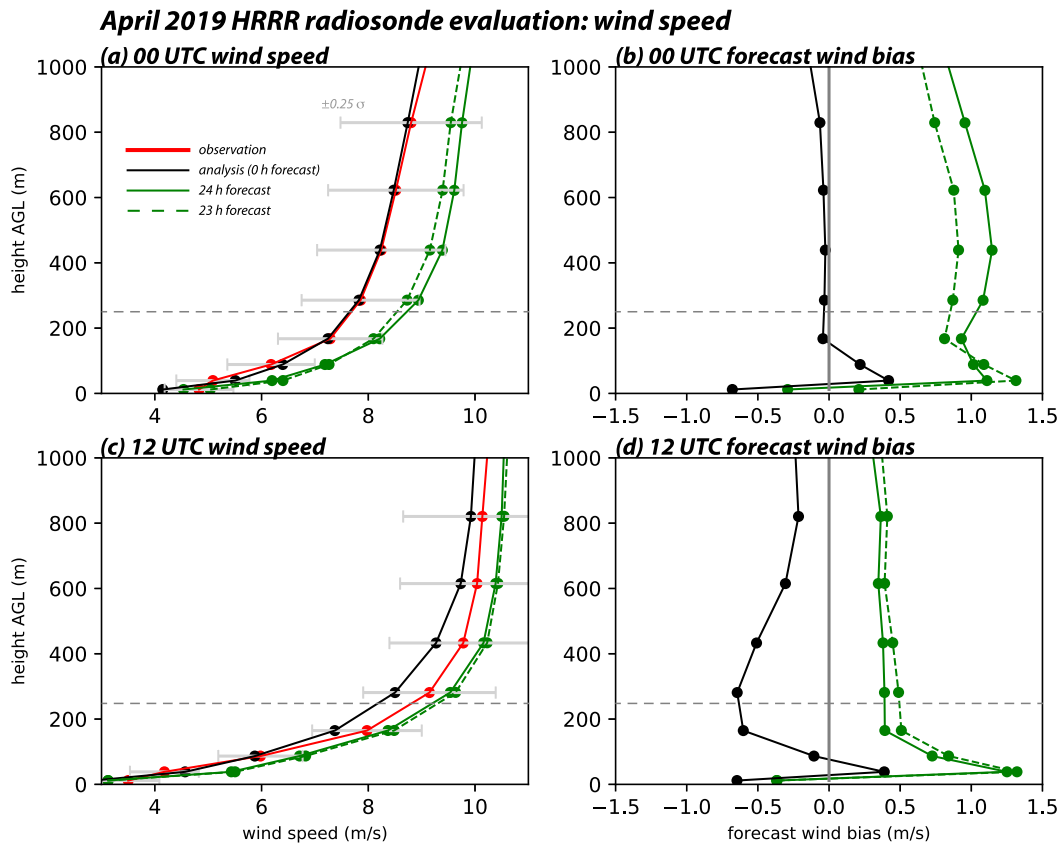


FIG. 14. April composite wind speed profiles and forecast wind speed biases aggregated among available RRS sites. The observation (red), analysis (black), and 23- and 24-h forecasts (dashed and solid green, respectively) profiles were constructed from the same April 2019 subsets to maximize comparability. This subset consisted of the 50 launches between 0000 UTC 2 Apr and 1200 UTC 27 Apr, inclusive. The gray horizontal bars on the observation profiles represent only  $\pm 0.25\sigma$  to permit more detail to be displayed.

At 0000 UTC, the observed and model analysis time composite wind profiles were nearly identical, except close to the ground in the contaminated layer. As already discussed, reasonable comparability should be anticipated because the radiosonde information was assimilated into the analysis. That said, 0-h forecast winds were weaker than the observations at every height shown at 1200 UTC, with a negative bias as large in magnitude as  $-0.64 \text{ m s}^{-1}$  at model level 5 (about 280 m AGL). As with the ASOS data, this could be a consequence of bias inherited from the first guess or some other aspect of the assimilation system. These aspects of the 0000 and 1200 UTC wind profiles were also found in both colder and warmer months (e.g., Fig. 15).

In the day-ahead (23- and 24-h) forecasts, however, there was a decided shift in the wind profiles toward higher speeds, and this occurred at all model levels shown. At 0000 UTC, the initially unbiased winds above the contaminated layer became positively biased by as much as  $\approx 1 \text{ m s}^{-1}$  at model level 6 (about 440 m AGL). For 1200 UTC, the original negative bias evolved into positive biases, shifting by  $1.1 \text{ m s}^{-1}$  at model level 5 (about 280 m AGL). Similar substantial shifts also occurred in January and July (Fig. 15) and other months available for

examination (not shown). To the degree that the observations and analysis time profiles can be construed as “truth,” these shifts toward higher wind speeds in the 23- and 24-h forecasts represent errors that need to be resolved.

We note that a consistent feature of the contaminated layer is the very sharp bias increase between the lowest two HRRR model levels (roughly 10–39 m AGL) followed immediately by a sizable reversal. This “kinked” bias profile appeared both morning and afternoon/evening, at all forecast hours, and for all months examined. The sharp shift between the first two HRRR levels appears to be caused by the aforementioned RWS processing that imposes a linear wind profile within the (more likely logarithmic) surface layer. That said, recalling that the first radiosonde observation is replaced with a contemporaneous surface observation, presumably from an anemometer mounted at 10 m AGL at most sites, observations interpolated to the first HRRR level (also being close to 10 m) may be relatively less corrupted than those within the forced linear layer immediately above. Indeed, at the analysis time, the wind speed bias at this level was about  $-0.65 \text{ m s}^{-1}$  at both 0000 and 1200 UTC, which is consistent with the initiation time bias ( $-0.5 \text{ m s}^{-1}$ ) found for the (much more extensive) ASOS

**HRRR radiosonde wind speed evaluation - January and June 2019**

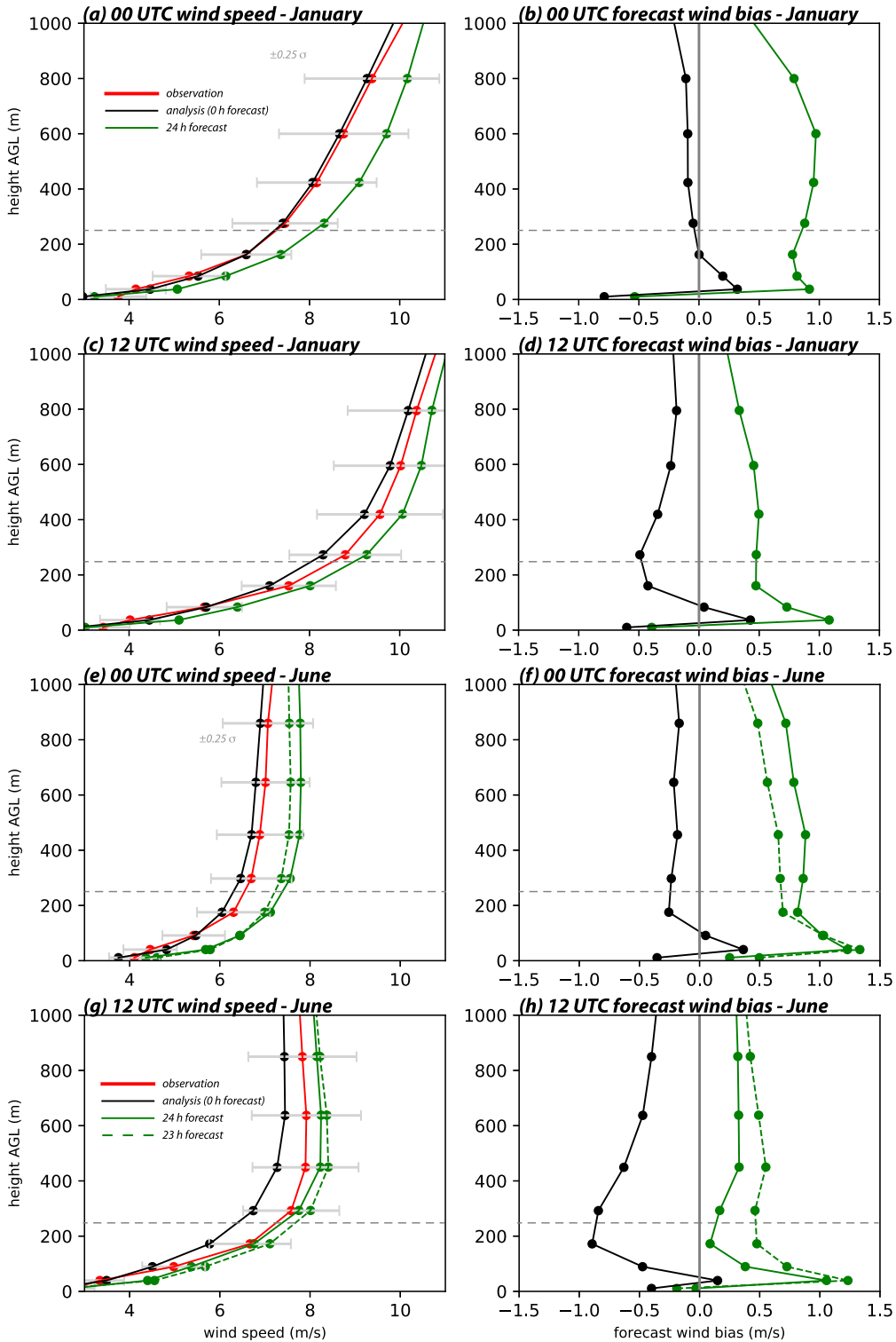


FIG. 15. As in Fig. 14, but for (a)–(d) January and (e)–(h) June 2019, respectively.

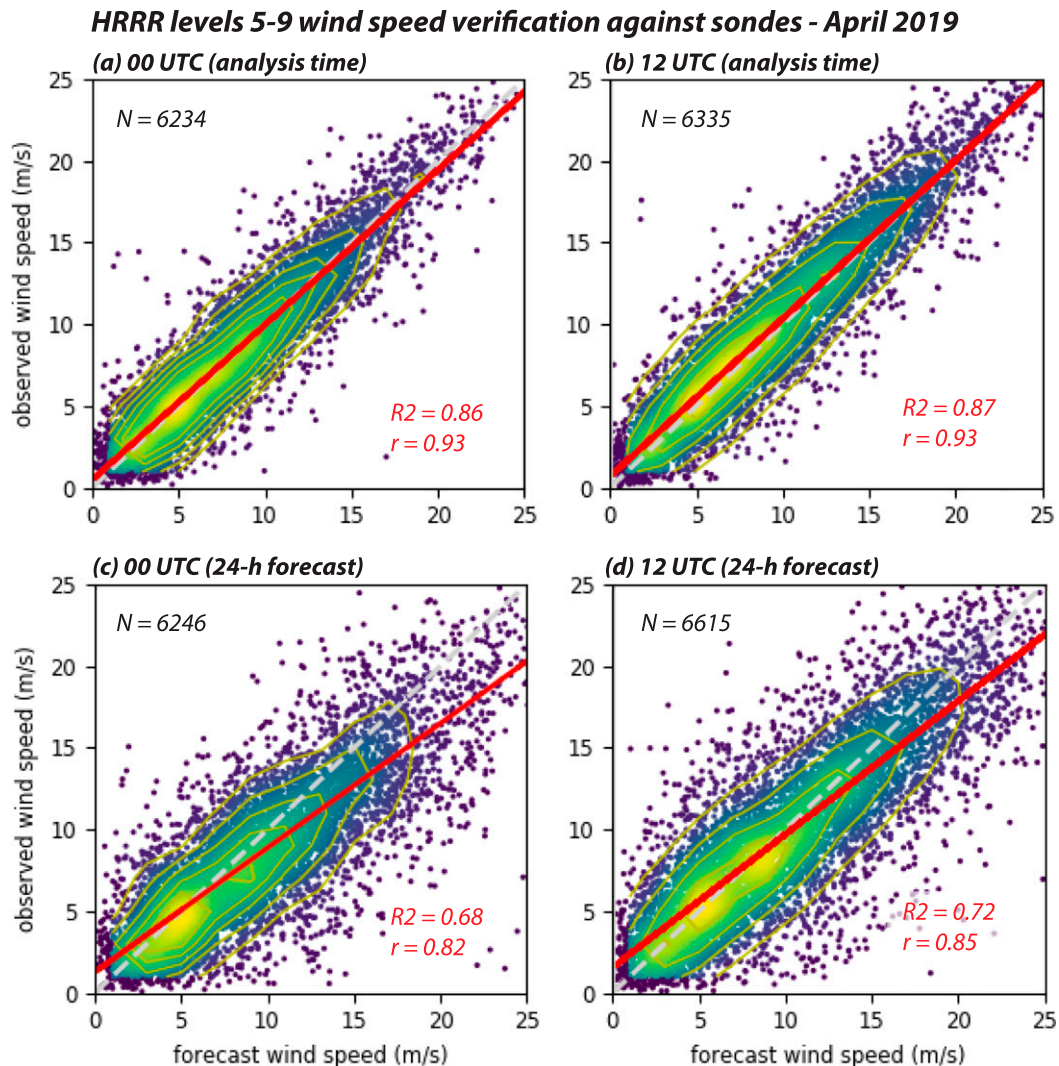


FIG. 16. Scatterplot of all April 2019 observation–forecast pairs available for HRRR levels 5–9, inclusive, for the (a),(c) 0000 UTC and (b),(d) 1200 UTC cycles at (top) the analysis time and (bottom) forecast hour 24. Similar to previous figures, dots are color coded by nearby point density and are additionally contoured every 50 samples. Also shown are least squares fits (solid red lines) and the one-to-one correspondences (dashed gray). Plots for forecast hour 23 (not shown) are very similar to hour 24.

network (Figs. 7a,b). Also as in the ASOS analysis, the biases at the lowest model level shifted toward zero in the day-ahead forecasts.

Figure 16 plots all forecast–observation pairs from HRRR model levels 5–9 inclusive (roughly 280–1060 m AGL) for April 2019 for all stations and times, and Fig. 17 matches those values against their respective forecast wind biases. These data were employed in the construction of Fig. 14’s wind profiles and may be more revealing than averaging by station owing to the relatively small number of available launch sites and times. The analysis time plots reveal a faithful (yet again imperfect) relationship between the modeled and observed winds (Figs. 16a,b). For 0000 UTC, the average bias was not only nearly zero (as anticipated from Fig. 14a) but also neither a function of the forecasted nor observed wind, despite the

rather obvious hard constraints discussed earlier. The small negative bias (about  $-0.4 \text{ m s}^{-1}$  averaged through the layer) among the 1200 UTC analyses, consistent with the black profile in Fig. 14d, is more identifiable in Figs. 17c and 17d, and was nearly independent of forecasted and observed wind speed.

The situation had again changed significantly by the 24-h forecast time (Figs. 16c,d and 17e–h), however. A clear positive bias developed, especially at 0000 UTC (Fig. 16c) that, in contrast with the situation seen with the surface wind observations, is more associated with the *forecasted* than observed values. This is indicated by the rightward tilt of the data points and contours on Figs. 17e and 17g compared to Figs. 17f and 17h. Relatively larger wind speeds, especially those between 6 and  $12 \text{ m s}^{-1}$  for which there are many examples, were more



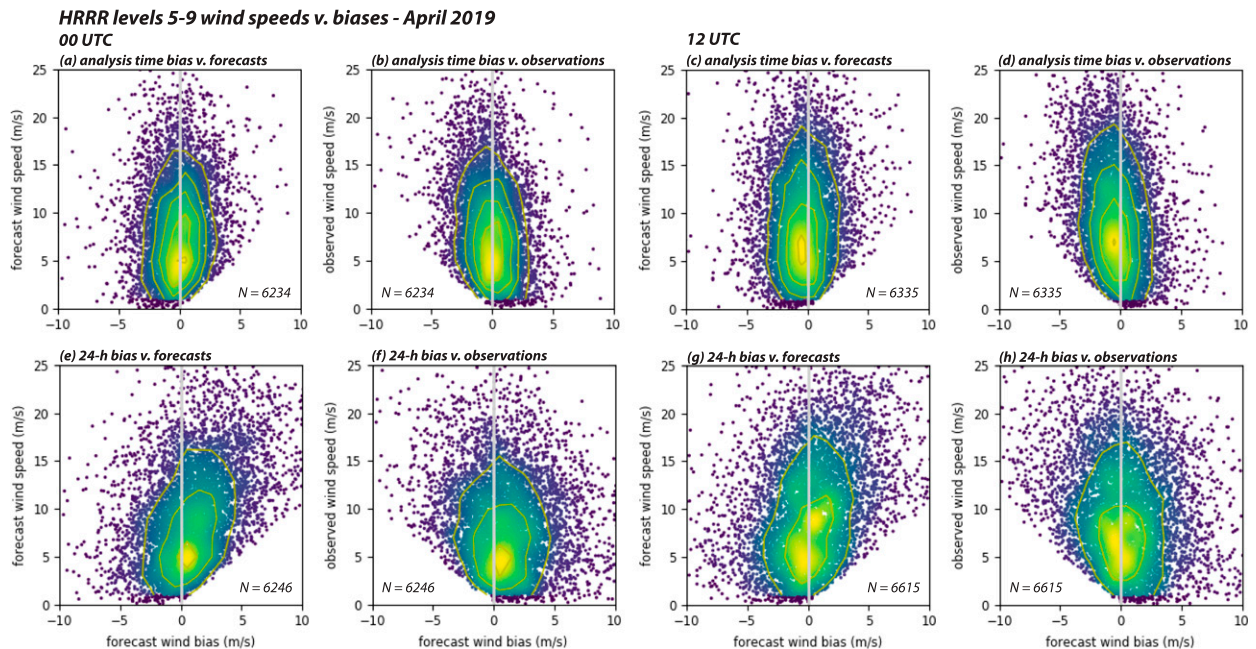


FIG. 17. As in Fig. 10, but for observation–forecast pairs of wind speeds on HRRR levels 5–9, inclusive. Shown are the (a),(b),(e),(f) 0000 UTC and (c),(d),(g),(h) 1200 UTC forecast cycles with comparisons being forecast wind bias vs forecasted [in (a), (c), (e), (g)] and observed [in (b), (d), (f), (h)] wind speeds. Only analysis time pairings are incorporated in (a)–(d) and (e)–(h) use only forecast hour 24 pairs. Similar to previous figures, dots are color coded by nearby point density and are additionally contoured every 50 samples. Plots for forecast hour 23 (not shown) are very similar to hour 24.

likely to be overpredicted. (Plots for the 23-h forecasts are similar, and thus not shown.)

Next, we turn to the temperature analysis for April (Fig. 18) and January and June (Fig. 19) 2019. We elected to examine potential temperature  $\theta$  as it is somewhat less sensitive to elevation above sea level. Unlike wind speed,  $\theta$  (like temperature) varies more across the CONUS than at a particular site, a characteristic that makes plots of forecasts versus observations (not shown) appear very skillful, with  $R^2 > 0.96$  even for 1-day forecasts. In contrast to the wind evaluation, we are not identifying a contaminated zone as it is less clear that temperature information near the surface was substantially compromised during processing, but emphasis will be placed on how these profiles changed over time anyway. Clearly, the relatively larger distinction (compared to wind speed) between forecast hours 23 and 24 occurred because temperature was evolving more quickly around the launch times (see also Fig. 7), especially around 0000 UTC. This makes the analyses less straightforward.

At the analysis time (Figs. 18b,d), April's potential temperature biases were small at and above model level 3 (about 88 m AGL) at 0000 UTC and a modest negative bias (averaging  $-0.32$  K) was apparent at 1200 UTC. Again, only minor discrepancies were expected as these data were assimilated. Whether that bias became more or less negative in the afternoon/evening, however, depended on the selected forecast hour owing to the comparatively quickly changing conditions. The apparent biases in the 0000 UTC analyses were similar in January and June to what was seen in

April (Figs. 19b,f), although how that bias evolved to the 23- or 24-h forecast ostensibly varied among the months.

For 1200 UTC, both the 23- and 24-h forecasts showed a combination of temperature increases beneath decreases, both relative to the analysis time. This indicates that the vertical potential temperature lapse rate was markedly less stable in the day-ahead forecast compared to the initial time (Fig. 18d). An essentially similar phenomenon occurred in the cooler and warmer months (Figs. 19d,h). While the sign of the  $\theta$  bias was dependent upon the level and month examined, the consistent result was that the site-composite mean 1200 UTC temperature at model level 3 shifted more than it did at level 6, and did so in a positive direction. As a consequence, the reduced stability present in the 1-day forecast compared to the analysis time was primarily due to shifts in forecasted temperatures closer to the surface.

It was seen in the 2-m temperature evaluation for the ASOS network that there was a pronounced tendency for higher elevation ( $>500$  m MSL) stations to develop positive forecast biases by the 24-h lead time (Figs. 12c,d and 13). A similar yet more complex result was found for HRRR model levels near the surface (Fig. 20), and this in particular helps to shed light on the lapse rate steepening seen at 1200 UTC (Fig. 18d). The present comparison only involves forecasts and also differences in predictions at hours 0 (analysis time) and 24 and for HRRR model levels 3 and 6 (about 88 and 441 m AGL, respectively) for the same set of valid times, sidestepping issues concerning launch times and potentially flawed initial conditions. For our sonde subset, 20 of the 60 launch sites resided above 500 m MSL.



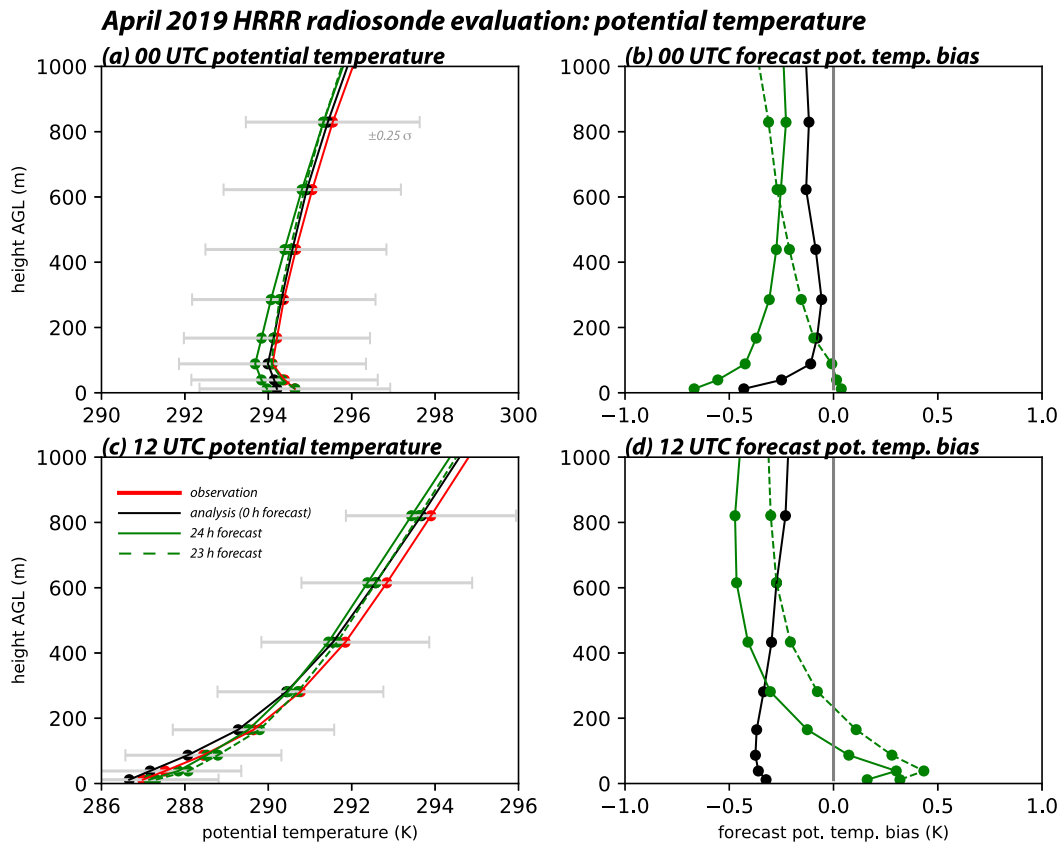


FIG. 18. As in Fig. 14, but for potential temperature and forecast bias averaged over April 2019.

At 0000 UTC, potential temperature changes between the 24- and 0-h forecasts were similar for both model levels, although they also tended to be positive for most higher elevation sites and be generally negative closer to sea level (Figs. 20a,b). As a consequence, the lapse rate shifts (Fig. 20c) were relatively modest, consistent with Figs. 18a and 18b. The results for January and June 2019 were similar (Figs. 21a–c,g–i), with the exception that the  $\theta$  shift at higher altitude sites were less consistent in the latter.

The situation was distinctly different at 1200 UTC (bottom row), however. At the third level, 19 of the 20 elevated locations had  $\theta$  increases for April (averaging to +1.3 K), while those closer to sea level were more evenly split (Fig. 20d). At level 6 (Fig. 20e), in contrast, temperatures generally increased at higher elevation sites and decreased at lower ones. These shifts combined to make lapse rates (Fig. 20f) for most of the sites (51 of 60) steeper by an average of  $1.7 \text{ K km}^{-1}$  (and by  $2.8 \text{ K km}^{-1}$  for sites above 500 m MSL) over the month, which are significant shifts that may reflect different problems depending on site altitude.

Results for January and June 2019 were strongly similar (Figs. 21d–f,j–l). As was the case with 2-m temperature, the identified errors were not correlated with the discrepancy between model and observation elevation (not shown). Taken together, these results confirm the finding that there is something systematically flawed about the way higher elevation

stations in particular are handled that is not confined to the 2-m temperature, and may also suggest issues with handling of the stable boundary layer. These findings also deserve closer examination in future work.

#### 4. Discussion and summary

Accurate forecasts of atmospheric conditions—especially wind speed and temperature—at the surface and through the PBL are critical for many industries including agriculture, aviation, wind energy, and fire weather (e.g., Zaitchik et al. 2005; Gultepe et al. 2019; Olson et al. 2019b; Erickson et al. 2018). In this study, we evaluated forecasts made by the 0000 and 1200 UTC cycles of the HRRRv3 model out to 24 h. For our verification, we utilized high-resolution (1-s) radiosonde (RRS) data, 1-min ASOS surface observations, and also HRRR model output on native model levels to preserve vertical structure near the surface. Our aim was to assess the current skill of HRRR forecasts, identify areas for potential improvement, and review the limitations of observations utilized in the verification process.

Our surface analysis of composited observations from hundreds of CONUS ASOS sites demonstrated that the HRRR skillfully reproduced the diurnal variations of both 10-m wind speed and 2-m temperature, albeit with nonzero biases in both variables ( $-0.5 \text{ m s}^{-1}$  and  $-0.4 \text{ K}$ , respectively) at the analysis time that disappeared as the forecast progressed. Consistent

**HRRR radiosonde potential temperature evaluation - January and June 2019**

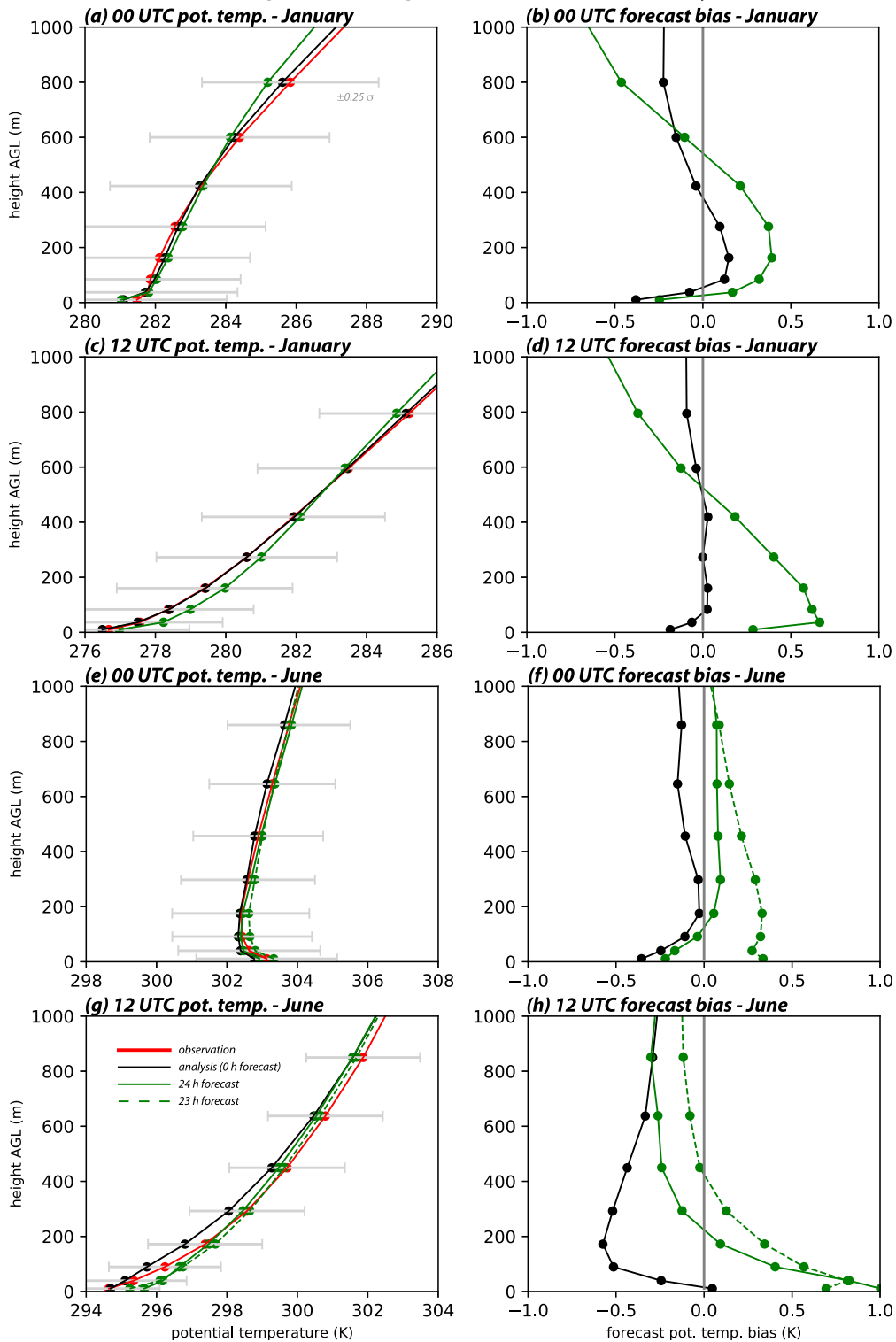


FIG. 19. As in Fig. 15, but for potential temperature for January and June 2019.

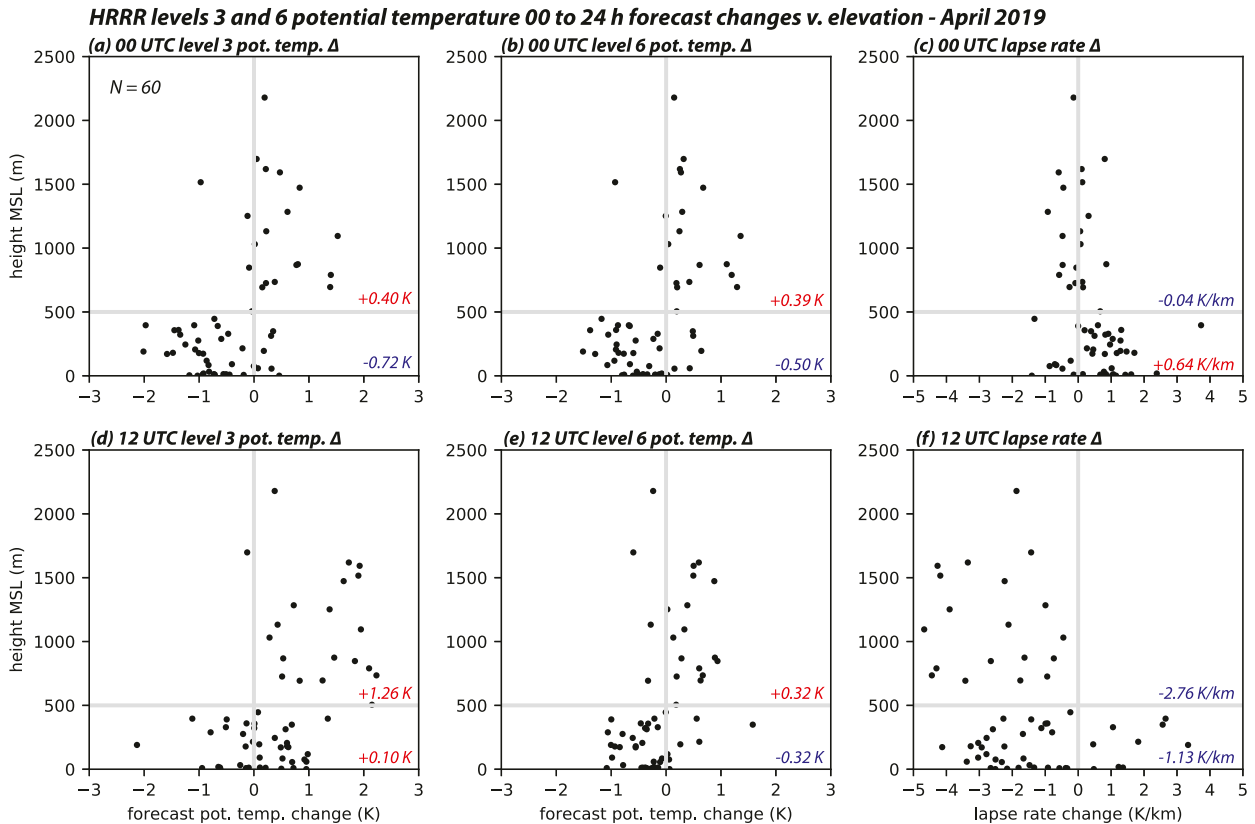


FIG. 20. Change ( $\Delta$ ) in potential temperature from the analysis time to the 24-h forecast vs station elevation for the (a),(b) 0000 and (d),(e) 1200 UTC forecast cycles at the (left) third and (center) sixth HRRR model levels. Also shown is the change in lapse rate (calculated between HRRR levels 3 and 6) from the analysis time to the 24-h forecast for the (c) 0000 and (f) 1200 UTC forecast cycles. Solid gray lines denote the zero-change line (vertical) and 500-m elevation (horizontal). Each dot is a station.

with Cao and Fovell (2018), HRRR systematically over-predicted the winds at sites reporting slower winds with under-predictions at faster wind locations, potentially as a result of unresolvable site exposure. Further analysis demonstrated that the wind speed bias was more a function of the observations than the forecasts, for reasons not fully appreciated. Independently ranked distributions of forecasted and observed mean station wind speeds emphasized that the model under-specified the wind at windier locations and suggested the negative bias seen at the analysis time (perhaps inherited from the assimilation process) quickly disappeared with time. Last, the 2-m temperature errors were found largely unrelated to both forecasted and observed values but were found to have a robust relationship with station elevation, where the majority (75%) of locations above 500 m MSL were characterized by positive temperature biases. This pattern was found to be persistent across seasons and indicates a model deficiency in handling temperature at high elevations that should be investigated further.

For the boundary layer verification, we interpolated the RRS' processed temperature (5pPTU) and wind (6pGPS) datasets at 1-s resolution to the AGL heights of the HRRR native model levels at each site and time, and then spatially and temporally averaged both forecasts and observations to create

vertical profiles representing the nominal times of 0000 and 1200 UTC. Regarding wind speed, our analysis indicated that the first 45 records after launch were contaminated during radiosonde processing so values below about 260 m AGL were discounted. Above that level, wind profiles for all available months show similar patterns of relatively small biases at the analysis time. This result was expected as the model analyses have assimilated the radiosonde information.

However, wind speed biases swung to become positive by the 23- and 24-h forecasts, increasing by approximately  $1 \text{ m s}^{-1}$  at HRRR model levels 5–6 (about 280–440 m AGL). This pattern is evident at both 0000 and 1200 UTC and also across seasons. Notably, the relationship between forecast bias and average forecasted or observed wind speed between model levels 5–9, inclusive, was reversed compared to the 10-m wind in that the bias was nearly random with respect to observed wind but had a positive relationship with forecasted wind speed, with a greater probability of overprediction when the forecasts were relatively higher. This result is not understood.

Regarding potential temperature, its relatively rapid evolution near the surface made forecast errors around the 0000 UTC target time less straightforward to interpret. However, a much more consistent pattern was found in the vertical bias structure around 1200 UTC. While analysis time

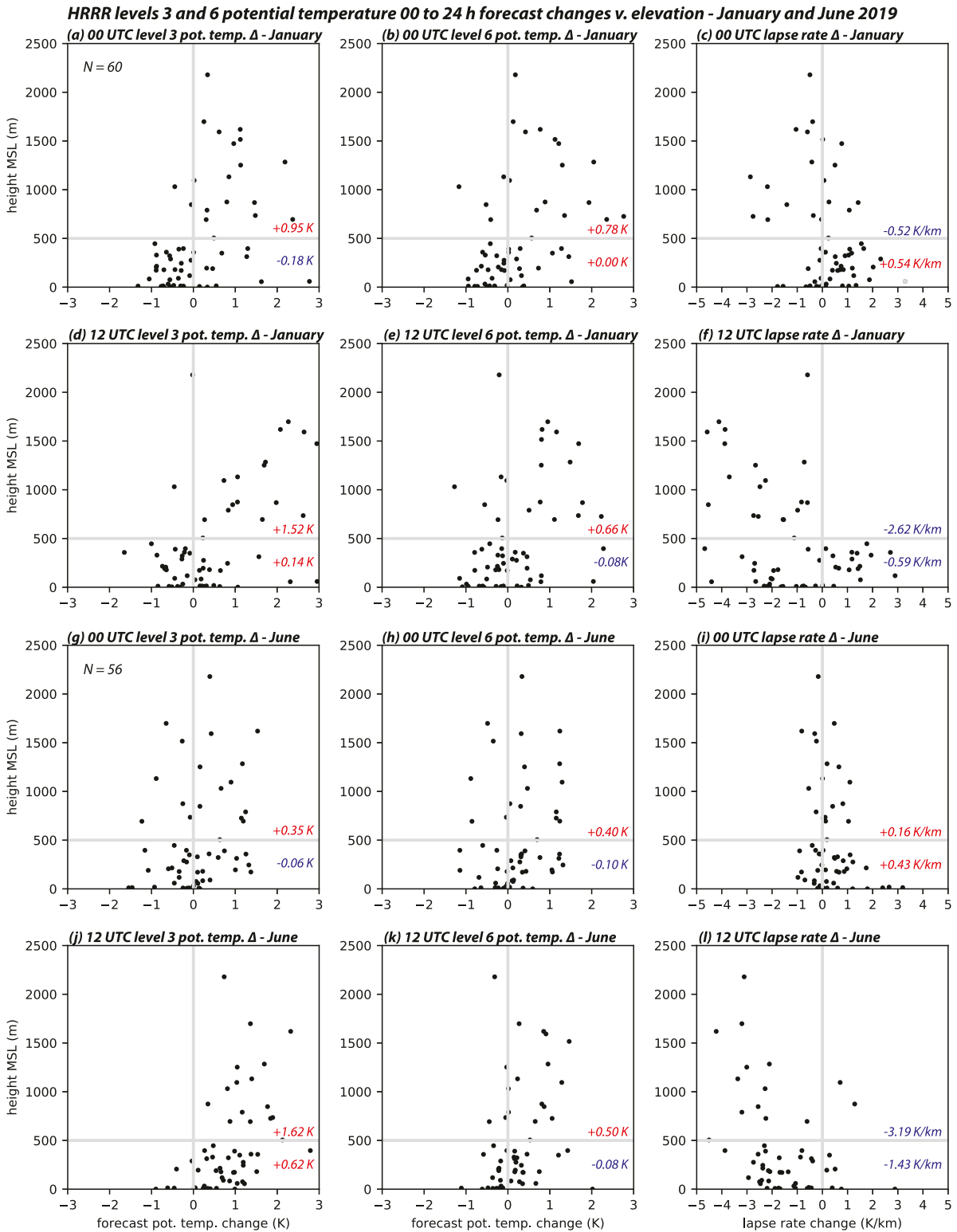


FIG. 21. As in Fig. 20, but for (a)–(f) January and (g)–(l) June 2019, respectively.

bias values varied among seasons, the next-day forecast revealed a persistent trend in which the vertical lapse rate between HRRR model levels 3 and 6 (about 88–441 m AGL) had become markedly steeper, by 1.3, 1.7, and 2.0 K km<sup>-1</sup> in January, April, and June 2019, respectively. For higher elevation sites, the steepening was driven by positive temperature biases at level 3 that exceeded those farther aloft, while for those closer to sea level it was generally warming residing beneath cooling. As with surface temperature, this analysis may indicate a mishandling of temperature at higher elevations and possibly also a more general misrepresentation of the stable boundary layer.

In summary, our analysis has shown that the HRRRv3 is very skillful, at least to 24 h, but also possesses systematic biases in its forecasts of surface and near-surface conditions that merit further investigation. These errors potentially stem from a variety of sources including (and not limited to) data assimilation, model physics assumptions and tunings, how higher elevation locations are being represented, and issues relating to the observations themselves, and may change as the model is refined. Additionally it should be noted that this verification benefited from access to both the higher-resolution and less processed RRS radiosonde data and the HRRR analyses and forecasts on native model levels. At this writing, the former are no longer being archived at NCEI and there is no known publicly available repository for the latter.

*Acknowledgments.* The authors thank three anonymous reviewers for their constructive comments. This research was sponsored by National Science Foundation Grant 1450195.

*Data availability statement.* This study used HRRR model outputs on native model coordinates that were generated by the National Oceanic and Atmospheric Administration's (NOAA) Earth System Research Laboratory that are temporarily made available in real time and were archived at the University of Albany, SUNY. There is no known publicly available archive for forecasts and analyses in that format. High-resolution radiosonde data were obtained from the NOAA archive at ftp.ncdc.noaa.gov/pub/data/ua/rrs-data/bufr and 1-min ASOS data are available at ftp.ncdc.noaa.gov/pub/data/asos-onemin. RRS decoding software was obtained from ftp.ncdc.noaa.gov/pub/data/ua/software.

## REFERENCES

- Beck, J., and Coauthors, 2020: An evaluation of a hybrid, terrain-following vertical coordinate in the WRF-based RAP and HRRR models. *Wea. Forecasting*, **35**, 1081–1096, <https://doi.org/10.1175/WAF-D-19-0146.1>.
- Benjamin, S. G., and Coauthors, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518, [https://doi.org/10.1175/1520-0493\(2004\)132<0495:AHACTR>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2).
- , and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Bullock, R., and Coauthors, 2017: Model Evaluation Tools version 6.1 (METv6.1): Users guide. Developmental Testbed Center, 400 pp.
- Burke, A., N. Snook, D. J. Gagne II, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning-based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, <https://doi.org/10.1175/WAF-D-19-0105.1>.
- Bytheway, J. L., C. D. Kummerow, and C. Alexander, 2017: A features-based assessment of the evolution of warm season precipitation forecasts from the HRRR model over three years of development. *Wea. Forecasting*, **32**, 1841–1856, <https://doi.org/10.1175/WAF-D-17-0050.1>.
- Cao, Y., and R. G. Fovell, 2016: Downslope windstorms of San Diego County. Part I: A case study. *Mon. Wea. Rev.*, **144**, 529–552, <https://doi.org/10.1175/MWR-D-15-0147.1>.
- , and —, 2018: Downslope windstorms of San Diego County. Part II: Physics ensemble analyses and gust forecasting. *Wea. Forecasting*, **33**, 539–559, <https://doi.org/10.1175/WAF-D-17-0177.1>.
- Caron, M., and W. J. Steenburgh, 2020: Evaluation of recent NCEP operational model upgrades for cool-season precipitation forecasting over the western conterminous United States. *Wea. Forecasting*, **35**, 857–877, <https://doi.org/10.1175/WAF-D-19-0182.1>.
- Coniglio, M. C., 2012: Verification of RUC0–1-h forecasts and SPC mesoscale analyses using VORTEX2 soundings. *Wea. Forecasting*, **27**, 667–683, <https://doi.org/10.1175/WAF-D-11-00096.1>.
- , J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, <https://doi.org/10.1175/WAF-D-12-00103.1>.
- Daniel, C. J., R. W. Arritt, and C. J. Anderson, 1999: Accuracy of 404-MHz radar profilers for detection of low-level jets over the central United States. *J. Appl. Meteor.*, **38**, 1391–1396, [https://doi.org/10.1175/1520-0450\(1999\)038<1391:AOMRPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<1391:AOMRPF>2.0.CO;2).
- Dirksen, R. J., M. Sommer, F. J. Immler, D. F. Hurst, R. Kivi, and H. Vömel, 2014: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmos. Meas. Tech.*, **7**, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>.
- Erickson, M. J., B. A. Colle, and J. J. Charney, 2018: Evaluation and postprocessing of ensemble fire weather predictions over the Northeast United States. *J. Appl. Meteor. Climatol.*, **57**, 1135–1153, <https://doi.org/10.1175/JAMC-D-17-0180.1>.
- Evans, C., S. J. Weiss, I. L. Jirak, A. R. Dean, and D. S. Nevius, 2018: An evaluation of paired regional/convection-allowing forecast vertical thermodynamic profiles in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1547–1566, <https://doi.org/10.1175/WAF-D-18-0124.1>.
- Fovell, R. G., and Y. Cao, 2017: The Santa Ana winds of Southern California: Winds, gusts, and the 2007 Witch fire. *Wind Struct.*, **24**, 529–564.
- , and A. Gallagher, 2018: Winds and gusts during the Thomas fire. *Fire*, **1**, 47, <https://doi.org/10.3390/fire1030047>.
- Gallagher, A. A., 2016: The network average gust factor, its measurement and environmental controls, and role in gust forecasting. M.S. thesis, Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, 139 pp.
- Griffin, S. M., J. A. Otkin, C. M. Rozoff, J. M. Sieglaff, L. M. Counce, C. R. Alexander, T. L. Jensen, and J. K. Wolff, 2017: Seasonal analysis of cloud objects in the High-Resolution Rapid Refresh (HRRR) model using object-based verification. *J. Appl. Meteor. Climatol.*, **56**, 2317–2334, <https://doi.org/10.1175/JAMC-D-17-0004.1>.



- Gultepe, I., and Coauthors, 2019: A review of high impact weather for aviation meteorology. *Pure Appl. Geophys.*, **176**, 1869–1921, <https://doi.org/10.1007/s00024-019-02168-6>.
- Horel, J., and B. Blaylock, 2019: Archive of the High-Resolution Rapid Refresh model. University of Utah Center for High Performance Computing, <https://doi.org/10.7278/S5JQ0Z5B>.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating High-Resolution Rapid Refresh model. *Wea. Forecasting*, **28**, 921–939, <https://doi.org/10.1175/WAF-D-12-00085.1>.
- James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble-variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, <https://doi.org/10.1175/MWR-D-16-0398.1>.
- Jankov, I., J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, <https://doi.org/10.1175/MWR-D-18-0092.1>.
- Lee, T. R., M. Buban, D. D. Turner, T. P. Meyers, and C. B. Baker, 2019: Evaluation of the High-Resolution Rapid Refresh (HRRR) model using near-surface meteorological and flux observations from northern Alabama. *Wea. Forecasting*, **34**, 635–663, <https://doi.org/10.1175/WAF-D-18-0184.1>.
- Loeser, C. F., M. A. Herrera, and I. Szunyogh, 2017: An assessment of the performance of the operational global ensemble forecast systems in predicting the forecast uncertainty. *Wea. Forecasting*, **32**, 149–164, <https://doi.org/10.1175/WAF-D-16-0126.1>.
- Love, P. T., and M. A. Geller, 2012: Research using high (and higher) resolution radiosonde data. *Eos, Trans. Amer. Geophys. Union*, **93**, 337–338, <https://doi.org/10.1029/2012EO350001>.
- Mahmood, R., and Coauthors, 2017: Mesonets: Mesoscale weather and climate observations for the United States. *Bull. Amer. Meteor. Soc.*, **98**, 1349–1361, <https://doi.org/10.1175/BAMS-D-15-00258.1>.
- Nakanishi, M., and H. Niino, 2004: An improved Mellor-Yamada Level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31, <https://doi.org/10.1023/B:BOUN.0000020164.04146.98>.
- National Weather Service, 2003: Data products and data archiving. National Oceanic and Atmospheric Administration/National Weather Service, 17 pp., [ftp://ftp.ncdc.noaa.gov/pub/data/ua/docs/RRS-data-archiving\\_Req9rev16.pdf](ftp://ftp.ncdc.noaa.gov/pub/data/ua/docs/RRS-data-archiving_Req9rev16.pdf).
- , 2010: RRS workstation software: Build 2 data processing description. NOAA Service Field Systems Operations Center, 72 pp.
- , 2018: Radiosonde Replacement System (RRS) workstation user guide for RWS software version 3.4.0.2. NOAA Tech. Memo., National Oceanic and Atmospheric Administration/National Weather Service Field Systems Operations Center, 334 pp.
- Olson, J. B., J. S. Kenyon, W. A. Angevine, J. M. Brown, M. Pagowski, and K. Suselj, 2019a: A description of the MYNN-EDMF scheme and the coupling to other components in WRF-ARW. NOAA Tech. Memo. OAR GSD-61, National Oceanic and Atmospheric Administration, Office of Oceanic and Atmospheric Research, 42 pp., <https://doi.org/10.25923/n9wm-be49>.
- , and Coauthors, 2019b: Improving wind energy forecasting through numerical weather prediction model development. *Bull. Amer. Meteor. Soc.*, **100**, 2201–2220, <https://doi.org/10.1175/BAMS-D-18-0040.1>.
- Pichugina, Y. L., and Coauthors, 2017: Assessment of NWP forecast models in simulating offshore winds through the lower boundary layer by measurements from a ship-based scanning Doppler lidar. *Mon. Wea. Rev.*, **145**, 4277–4301, <https://doi.org/10.1175/MWR-D-16-0442.1>.
- , and Coauthors, 2019: Spatial variability of winds and HRRR–NCEP model error statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia River Basin. *J. Appl. Meteor. Climatol.*, **58**, 1633–1656, <https://doi.org/10.1175/JAMC-D-18-0244.1>.
- Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, <https://doi.org/10.1175/WAF-D-14-00118.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle land surface model (RUC LSM) available in the Weather Research and Forecasting (WRF) model. *Mon. Wea. Rev.*, **144**, 1851–1865, <https://doi.org/10.1175/MWR-D-15-0198.1>.
- Wade, A. R., M. C. Coniglio, and C. L. Ziegler, 2018: Comparison of near- and far-field supercell inflow environments using radiosonde observations. *Mon. Wea. Rev.*, **146**, 2403–2415, <https://doi.org/10.1175/MWR-D-17-0276.1>.
- Walters, C. K., J. A. Winkler, S. Husseini, R. Keeling, J. Nikolic, and S. Zhong, 2014: Low-level jets in the North American Regional Reanalysis (NARR): A comparison with rawinsonde observations. *J. Appl. Meteor. Climatol.*, **53**, 2093–2113, <https://doi.org/10.1175/JAMC-D-13-0364.1>.
- Zaitchik, B. F., J. Evans, and R. B. Smith, 2005: MODIS-derived boundary conditions for a mesoscale climate model: Application to irrigated agriculture in the Euphrates Basin. *Mon. Wea. Rev.*, **133**, 1727–1743, <https://doi.org/10.1175/MWR2947.1>.
- Zeng, Z., S. Sokolovskiy, W. S. Schreiner, and D. Hunt, 2019: Representation of vertical atmospheric structures by radio occultation observations in the upper troposphere and lower stratosphere: Comparison to high-resolution radiosonde profiles. *J. Atmos. Oceanic Technol.*, **36**, 655–670, <https://doi.org/10.1175/JTECH-D-18-0105.1>.