An Evaluation of Surface Wind and Gust Forecasts from the High-Resolution Rapid Refresh Model

Robert G. Fovell and Alex Gallagher

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

Corresponding author: Robert G. Fovell, Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, NY, 12222, USA. Email: rfovell@albany.edu

ABSTRACT: We utilized high temporal resolution, near-surface observations of sustained winds and gusts from two networks, the primarily airport-based Automated Surface Observing System (ASOS) and the New York State Mesonet (NYSM), to evaluate forecasts from the operational High-Resolution Rapid Refresh (HRRR) model, versions 3 and 4. Consistent with past studies, we showed the model has a high degree of skill in reproducing the diurnal variation of network-averaged wind speed of ASOS stations, but also revealed several areas where improvements could be made. Forecasts were found to be underdispersive, deficient in both temporal and spatial variability, with significant errors occurring during local nighttime hours in all regions and in forested environments for all hours of the day. This explained why the model overpredicted the network-averaged wind in the NYSM because much of that network's stations are in forested areas. A simple gust parameterization was shown not only to have skill in predicting gusts in both networks but also to mitigate systemic biases found in the sustained wind forecasts.

SIGNIFICANCE STATEMENT: Many users depend on forecasts from operational models and need to know their strengths, weaknesses, and limitations. We examined generally high-quality near-surface observations of sustained winds and gusts from the nationwide Automated Surface Observing System (ASOS) and the New York State Mesonet (NYSM) and used them to evaluate forecasts from the previous (Version 3) and current (Version 4) operational High-Resolution Rapid Refresh (HRRR) model for a selected month. Evidence indicated that the wind forecasts are excellent yet imperfect and areas for further improvement remain. In particular, we showed there is a high degree of skill in representing the diurnal variation of sustained wind at ASOS stations but insufficient spatial and temporal forecast variability and overprediction at night everywhere, in forested areas at all times of day, and at NYSM sites in particular, which are more likely to be sited in the forest. Gusts are subgrid even at the fine grid spacing of the HRRR (3 km) and thus must be parameterized. Our simple gust algorithm corrected for some of these systemic biases, resulting in very good predictions of the maximum hourly gust.

1. Introduction

Accurate wind forecasts are important in a number of areas, including and not limited to wind energy (Piccardo and Solari 1998; Petersen et al. 1998), pollution transport (Arya 1999), and anticipation and mitigation of damage resulting from strong winds (Holmes et al. 2014). An example of the latter is the "Santa Ana" weather event (cf. Rolinski et al. 2019), a cool-season pattern of offshore flow in Southern California that is known to dramatically increase the risk of large wildfires (Westerling et al. 2004; Rolinski et al. 2016). Numerical modeling of Santa Ana events using the Weather Research and Forecasting (WRF) model's Advanced Research WRF (ARW) core (Skamarock et al. 2019) for the purposes of model verification and wind reconstruction (e.g., Cao and Fovell 2016; Fovell and Cao 2017; Cao and Fovell 2018; Fovell and Gallagher 2018) has revealed strengths and weaknesses of both the forecasts and the observations of the *sustained wind*, which in practice implies averaging over periods of time such as 2 or 10 min. At mesoscale grid spacings, short-period (e.g., 3-s) *gusts* are a subgrid-scale phenomenon, necessitating parameterization in all operational numerical weather prediction models at this writing. There have been many such parameterizations proposed (cf. Sheridan 2011), some being rather complex (Panofsky

et al. 1977; Nakamura et al. 1996; Brasseur 2001; Gray 2003; Stucki et al. 2016; Gutiérrez and Fovell 2018; Benjamin et al. 2021, to name a few).

Many users rely on wind predictions from operational models such as NOAA's operational High-Resolution Rapid Refresh (HRRR) (cf. Benjamin et al. 2016; Dowell and co authors 2022). HRRR is based on WRF-ARW and has 3 km horizontal grid spacing covering the conterminous United States (CONUS). A number of studies have focused on verification of HRRR forecast fields, including wind speed (cf. Olson et al. 2019b; Pichugina et al. 2019; Wilczak et al. 2019). In particular, Fovell and Gallagher (2020), hereafter FG20, presented a forecast verification of HRRR version 3's (HRRRV3 or V3) 00 and 12 UTC cycles, which were selected for their relatively long (36-h) forecast periods. (Although new HRRR cycles were launched hourly, only the 00 and 12 UTC model runs ran longer than 18 h in V3.) Also, while other select months were also examined, the primary focus was on April 2019 as a representative time period.

In addition to the boundary layer analysis that employed high-resolution radiosonde data, an evaluation of 2-m temperature and 10-m wind speed forecasts for $\approx$ 800 Automated Surface Observing System (ASOS) sites was conducted. These installations are typically, but not always, found at airports. FG20 demonstrated that the HRRRV3 produced skillful forecasts when averaged over the ASOS network although temperature biases were robustly related to station elevation and wind biases were negatively correlated with observed speed. The latter means that "sites characterized by slower observed winds were systematically more likely to be overpredicted while windier sites were underestimated" (FG20), consistent with the results of prior studies focusing specifically on Santa Ana events (cf. Cao and Fovell 2016; Fovell and Cao 2017; Cao and Fovell 2018; Fovell and Gallagher 2018).

In this work, FG20's evaluation of forecasts for ASOS stations was reconsidered from scratch and considerably extended and improved. As in FG20, we started with April 2019, but the specific emphasis is on hourly mean winds and maximum gusts with the discussion confined to the 00 UTC cycle in order to streamline the presentation. In this effort, data from the New York State Mesonet (NYSM; Brotzge et al. 2020) were also analyzed and gust forecasts made using a simple parameterization suggested by Cao and Fovell (2018, hereafter CF18) were considered. As version 4 of the HRRR (HRRRV4 or V4) became operational in December 2020, an analysis of April 2021 is also provided to highlight improvements and identify remaining challenges.

This work diagnoses systemic errors and weaknesses of a very skillful operational model for the purposes of highlighting areas for potential future improvements. Another goal was to identify and understand issues with available observational data. This paper is organized as follows. Section 2 describes the data and methods used in this study and Sections 3 and 4 present our analyses of April 2019 (HRRRV3) and April 2021 (HRRRV4), respectively, the latter emphasizing comparisons with the Section 3 findings. Finally, Section 5 presents some conclusions and recommendations.

2. Data and methods

Anemometers of different types, including the sonic, cup and vane, and propeller varieties, are used to sample the wind at some period we will term the sampling interval. These samples are then averaged over a certain period, the averaging interval. The World Meteorological Organization (WMO) standard (WMO 2018) specifies averaging intervals of 3-s and 10-min for the gust and sustained (mean) wind, respectively. In a given report consisting of sustained wind (hereafter usually termed simply as "wind") and gust readings, the gust is conventionally the highest 3-s value within the averaging interval used for the wind[1]. The standard also specifies an anemometer mounting height at 10 m above ground level (AGL) with adequate clearance around the instrument. Ideally, the surrounding environment would consist of open flat terrain with obstacles no taller than 4 m and more than thirty times their height (2° above the horizon) away from the anemometer (WMO wind class 1). Adherence to these guidelines, however, is not all that common in practice.

NOAA makes HRRR model outputs available hourly and on the hour, providing forecasts of 10-m AGL wind speed representing an instant of time[2]. However, because the winds at any grid point only vary over time periods that are much longer than the model time step (20 s), these are interpreted as sustained winds. As in FG20, 1-min ASOS observations were obtained from the National Centers for Environmental Information (NCEI) archive, which are available for more than 850 sites in the CONUS. The 1-min observations provide measurements of sustained winds and gusts made from sonic anemometers nominally at 10 m AGL. Although the internal processing is complicated[3], the sustained wind readings we used effectively represent an average of samples

[1]In the United States, a significant exception to this is the Remote Automated Weather Station (RAWS) network, for which hourly reports consist of the past hour's highest speed sample (peak wind) along with the mean wind of the last 10 min prior to the report (National Wildfire Coordinating Group 2019). Thus, there is no guarantee the peak came from the samples used to compute the sustained wind.

[2]The lowest horizontal wind model level is close to 10 m AGL and the 10-m wind speed value is obtained via vertical interpolation. See Benjamin et al. (2021).

[3]See documentation at https://www.weather.gov/asos/

taken over the 2-min period prior to the report, with the highest 3-s average during the 1-min interval provided as the gust. The consequences of the relatively coarse (1 kt or 0.5144 m s^{-1}) precision of ASOS wind and gust reports will be noted in the analyses to come.

The FG20 analysis used top-of-the-hour ASOS reports and model fields were interpolated to station locations in the usual fashion. However, owing to the model's horizontal resolution, which does not resolve small turbulent eddies, there is very likely less temporal and spatial variability in the forecasts than in the observations. To assess whether this unduly influenced the results, we elected to pursue an alternative strategy in this new effort, using the observed *hourly mean wind speed* and *hourly maximum gust*. Sustained wind observations from each site were averaged through a 60-min window centered at the top of each hour and the largest gust report within that window was identified. For each station, only hours without missing or invalid data in a given hour were retained. Thus, we used hourly-averaged winds instead of 2-min averages in the sustained wind verifications. Owing to Harper et al. (2010), who argued that different averaging intervals represent "equivalent measures of the true mean wind but with differing variance", we expected that the results for the sustained wind would be nearly unchanged, and this proved to be true.

In contrast, the altered handling of the gusts did make a difference. In prior work using 1-min ASOS observations (including Cao and Fovell 2016, 2018; Fovell and Gallagher 2018), the gust in each station record represented the largest speed sample during the 1-min interval at the top of each hour. Because this covers only 1.7% of the hour, we believe the hourly maximum gust is a better measure of the wind threat. This caused a reasonable and anticipated change in the gust factor (GF), being the gust divided by the sustained wind. Averaged over the CONUS, the 1-min ASOS GF was about 1.29 and this increased to 1.86 with the new strategy. Further discussion may be found in the Appendix.

Although most ASOS stations are at airports there are some significant exceptions, such as the consistently windiest site (KDGP - Guadalupe Pass, TX), a non-airport installation sited near a steep cliff. There are some very low wind speed stations, including non-airport sites such as KMEH (Meacham, OR), KP69 (Lowell, ID), and KMHS (Mt. Shasta, CA), and small airports possessing significant along-runway obstructions, examples being KVPC (Carterville, GA) and K1JO (Bonifay, FL). A fraction of installations reportedly have anemometers mounted below 10

m AGL (e.g., KMTP - Montauk, NY). None of these problem stations were excluded from our analyses because they were not found to alter our results or conclusions.

The New York State Mesonet (Brotzge et al. 2020) contains 126 surface stations distributed across the state with an average spacing of 27 km. Each station possess sonic and propeller anemometers mounted (apart from five rooftop installations in New York City) at 10 m AGL. Retention of these rooftop sites did not change our results or conclusions. The precision of the sonic and propeller anemometer readings are 0.1 and 0.17 m s^{-1}, respectively (G. Lufft Mess und Regeltechnik GmbH 2021; R.M. Young Company 2000). Quality controlled, three-second observations from both sensors were obtained directly from the Mesonet. This would seem to represent an opportunity to evaluate the influence of hardware on the wind measurements but there are some unfortunate complications. The NYSM propeller instrument provided a 3-s average wind every 3 s, consistent with the WMO gust standard and being the same gust averaging interval employed by the ASOS sonic anemometers. In contrast, the NYSM's sonic instrument sampled once per second but only every third reading was recorded, meaning its gusts are actually 1-s and not 3-s averages.

As with the ASOS data, we used the NYSM readings to construct hourly average winds and hourly maximum gusts centered on the hour for both instruments, but retained only hours with valid data from both instruments. Over April 2019 and 2021, mean propeller winds were about 0.25 m s^{-1} (10.7%) lower than for the sonic, and gusts were 0.6 m s^{-1} (12%) slower, these differences being large enough to be relevant to our analyses. The propeller anemometer reported relatively more readings close to calm. The network-averaged GFs for April 2019 were 2.21 and 2.24 from the propeller and sonic instruments, respectively. The shorter interval used with the sonic gust data could be expected to increase the GF slightly (cf. Durst 1960).

FG20 did not consider gust forecasts. Herein we verified forecasts made using the simple CF18 parameterization for 10-m gusts, which consisted of multiplying the (sustained) wind forecast by the network-averaged GF after correcting for the mean network-averaged bias. We note the HRRR model also provides "gust potential" forecasts created using boundary layer depths and winds (Benjamin et al. 2021). However, in the hourly HRRR outputs, these forecasts are instantaneous values. It would be inappropriate to consider them as predictions of the hourly maximum gust and they do not verify well against them anyway (not shown). The HRRR makes subhourly (every 15-min) forecasts available, but these neither fully sample the hour nor are available beyond forecast

hour 18, even in HRRRV4. As a consequence, we did not consider the HRRR's gust forecasts in this study.

3. HRRRV3 wind and gust evaluation for April 2019

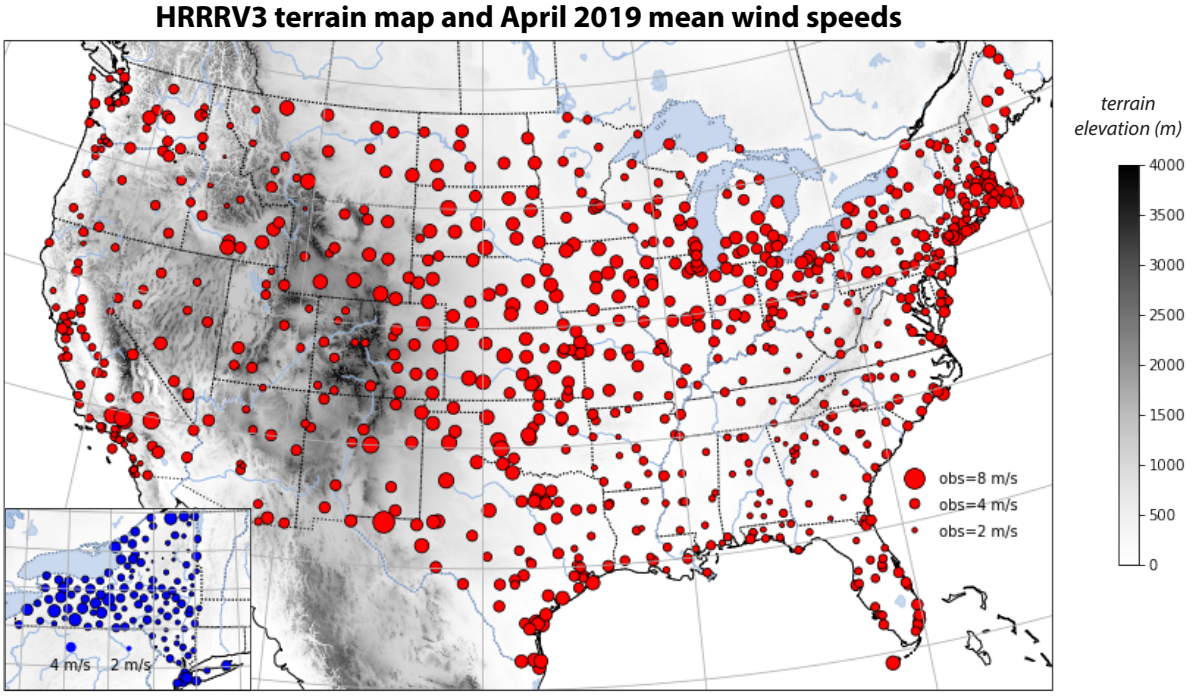
Figures 1 and 2a show the topography and primary landuse assignments used by the HRRRV3. Landuse and terrain information was obtained from the WRF Geogrid file made available on NOAA's HRRR website (http://rapidrefresh.noaa.gov/hrrr/). In WRF-ARW version 3, on which HRRRV3 was based, there were two separate landuse databases derived from MODIS (Moderate Resolution Imaging Spectroradiometer) satellite information available with HRRRV3 employing the lower-resolution version. Surface roughness lengths (z_0), constructed as described in He et al. (2021), were extracted from model outputs at forecast hour 12[4].

The locations of 807 ASOS and 126 NYSM sites are also shown on Fig. 1, with marker size reflecting mean wind speed for April 2019. Sites with fewer than 500 observations in the month were excluded from the analysis and are not shown. Owing to finite resolution, a few stations were misclassified as being over water (including having $z_0 < 0.01$ m), and these were also removed. WRF-ARW and the HRRR's Rapid Update Cycle (RUC) land surface model utilize fractional landuse assignments, and more than half (53%) of the ASOS stations were associated with more than one class (Fig. 2b). This can and does influence surface characteristics (including roughness) used in a given grid cell. That being said, the class representing the primary assignment had an average landuse fraction of 0.84 over the 807 ASOS sites, this ranging from 0.74 among the forested lands to 0.88 for the cropland and urban classes.

a. Analysis by forecast hour and local time

As in FG20, we first considered ASOS network-averaged winds expressed in terms of forecast hour, which extended out to 36 h for the 00 UTC cycle. The present result (Fig. 3a) is nearly identical to that shown in FG20 (their Fig. 7a), illustrating that the adoption of hourly mean observations made essentially no difference. Again, the model started with a small negative bias (defined as forecast minus observation) of about -0.5 m s^{-1} that became smaller in magnitude with time over the first 24 forecast hours. This bias is small compared to the spatial variation of the

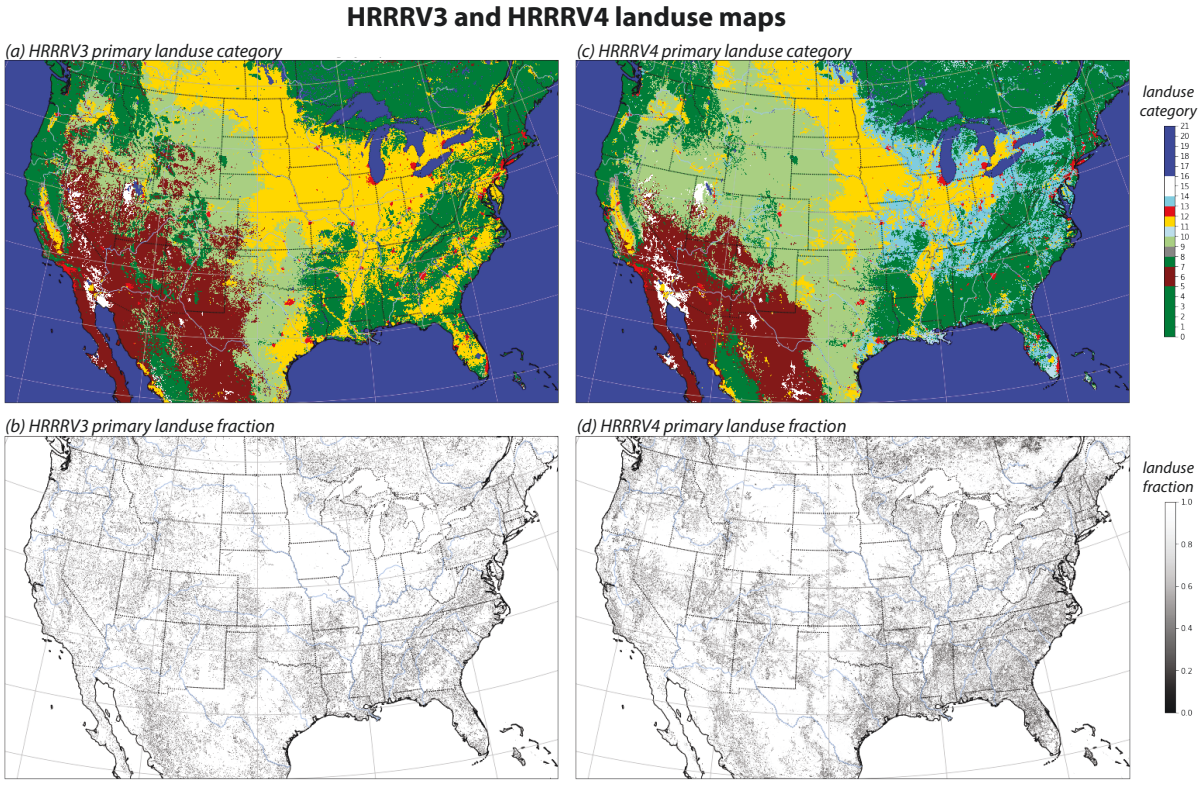
[4]In WRF-ARW, roughness lengths reported in the 0 h model output has not yet been updated, and thus may not be correct.



184 FIG. 1. Topography of the HRRRV3 domain, superposed with locations of 807 ASOS stations (red) and 126
185 NYSM sites (blue, in inset) retained in the April 2019 analysis. Marker sizes indicate monthly average wind
186 speed from April 2019, using observations from all times of day.

199 observations (illustrated by the grey vertical bars) owing to fact we are averaging across a very
200 wide area spanning four time zones.

201 New to this evaluation are examinations of forecast and observation spatial and temporal variabil-
202 ity and an analysis by local time (LT). Figure 4a reveals that the spatial variation of the forecasts
203 valid at ASOS sites (henceforth, "ASOS forecasts"), expressed as the standard deviation, was
204 smaller than that of the observations at all forecast hours. There is a diurnal cycle in both, again
205 smeared by averaging across time zones. This may be in part a consequence of local landscape
206 features (valleys, hills, obstacles and/or land surface variations) that cannot be resolved in the
207 model. Since the *mean* forecast and observed winds were quite similar, it can be anticipated that
208 the model would fail to represent the frequency of both lower and higher wind speeds. This will be
209 examined presently. Additionally, Fig. 4b presents time series of the difference between forecast



187 FIG. 2. Primary landuse assignments used in the (a) HRRRV3, and (c) HRRRV4, color coded by landuse
188 category, showing ocean and lakes (blue), croplands (gold), grasslands (light green), evergreen, deciduous, and
189 mixed forests and woody savannas (dark green), open shrublands (maroon), and urban (bright red) and barren
190 (white) lands. In (b) and (d), fraction (0-1) of the primary landuse classification in the HRRRV3 and HRRRV4,
191 respectively.

210 and observation spatial standard deviation and the forecast wind bias. They are similar in that they
211 both were negative but became less so with time.

220 Expressed in terms of LT, the network-averaged forecasts retained a negative bias through
221 the day (Fig. 3b), with the model apparently ramping up the late morning winds too slowly
222 and diminishing them too quickly into the evening[5]. The HRRR model employs the Mel-
223 lor–Yamada–Nakanishi–Niino Level 2.5 (MYNN2) planetary boundary and surface layer param-
224 eterizations (Nakanishi and Niino 2004) which have been refined in recent years (cf. Olson et al.
225 2019a). This finding may hold clues for further parameterization improvements. There was a

[5]The analysis time, forecast hour 0, was removed from this analysis owing to the shift in bias behavior seen between the analysis and forecast hour 1 in Figs. 4a and b.

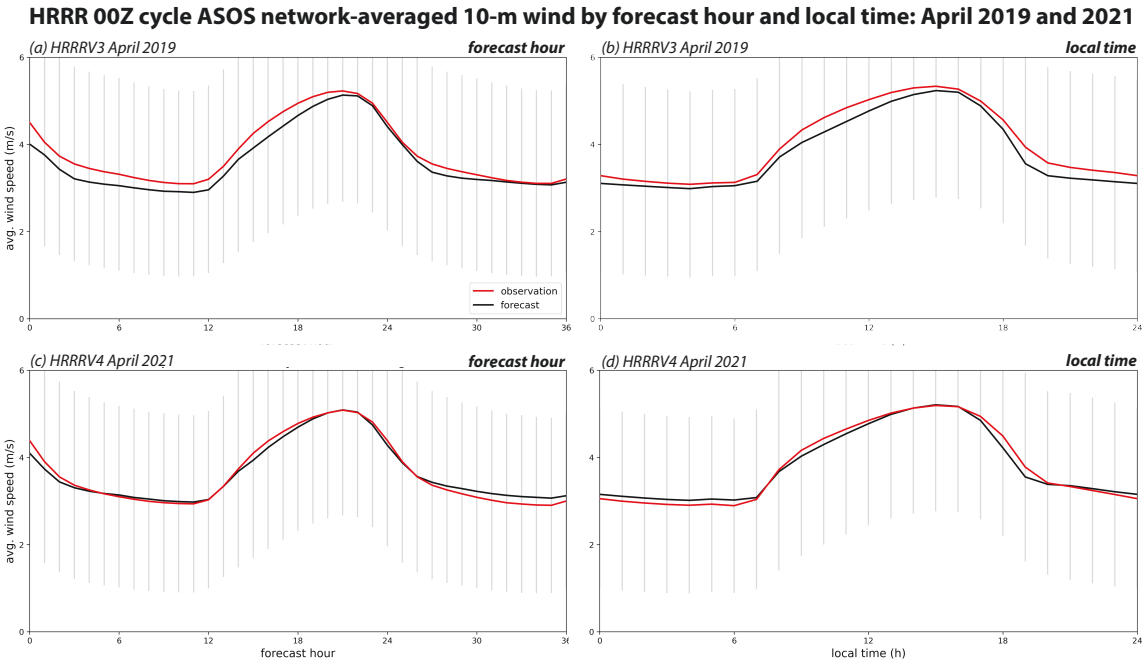


FIG. 3. Time series of ASOS observations (red) and HRRR forecasts (black) of 10-m sustained wind speeds, averaged spatially across the ASOS network and temporally over the month of (a, b) April 2019, and (c, d) April 2021, presented with respect to HRRR forecast hour (left) and local time (right). On all plots, the vertical grey bars denote ±1 standard deviation of the averaged observations.

diurnal cycle in both forecast and observation spatial variation (Fig. 5a) but again the forecast variability was slightly smaller and the diurnal variation in spatial standard deviation difference and forecast bias was very small (Fig. 5b). It is emphasized that this is an excellent, if not completely perfect, forecast, at least with respect to the network average.

In pointed contrast, the HRRRV3 overpredicted wind speeds averaged over the 126 NYSM sites by more than 1 m s^{-1} (Fig. 6a). Part of this gap is due to the propeller instrument that, as noted above, reports lower sustained wind speeds than its sonic counterpart. However, the forecast bias with respect to the sonic observations was 0.77 m s^{-1}, which is still sizable. Another difference is that the spatial variability of the forecasts (Fig. 6b) was larger than the observations at every forecast hour with the biases and spatial standard deviation differences being relatively constant with forecast hour (Fig. 6c). We need to emphasize at this point that the ASOS and NYSM networks

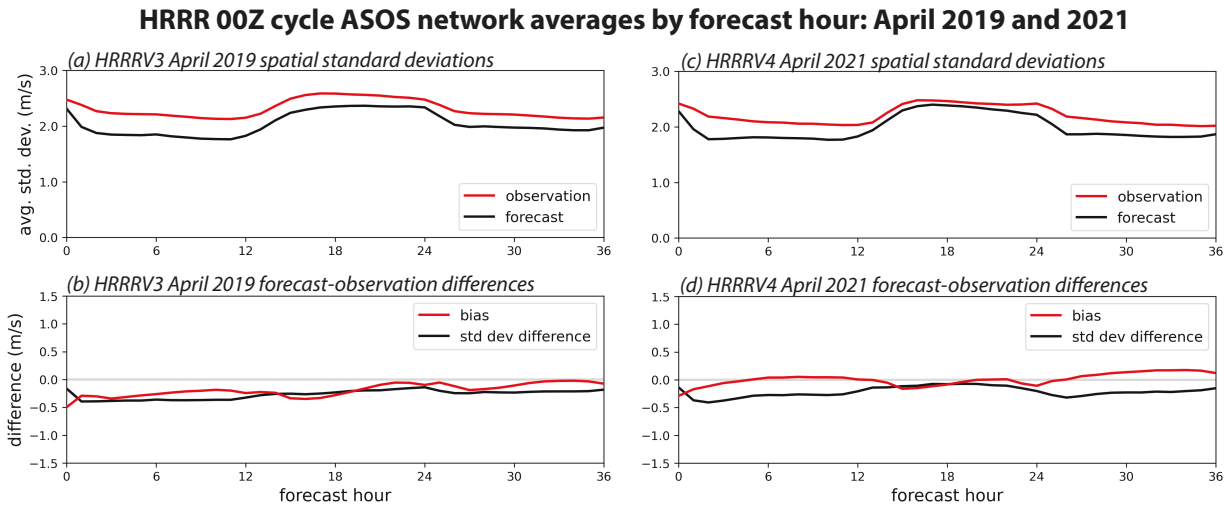


Fig. 4. Time series of the spatial standard deviation of ASOS observations (red) and HRRR forecasts (black) of 10-m wind speed for April 2019 and April 2021 (top row), and of forecast minus observation average wind speed (bias, red) and spatial standard deviation (black) for these same two months (bottom row). All are shown with respect to HRRR forecast hour.

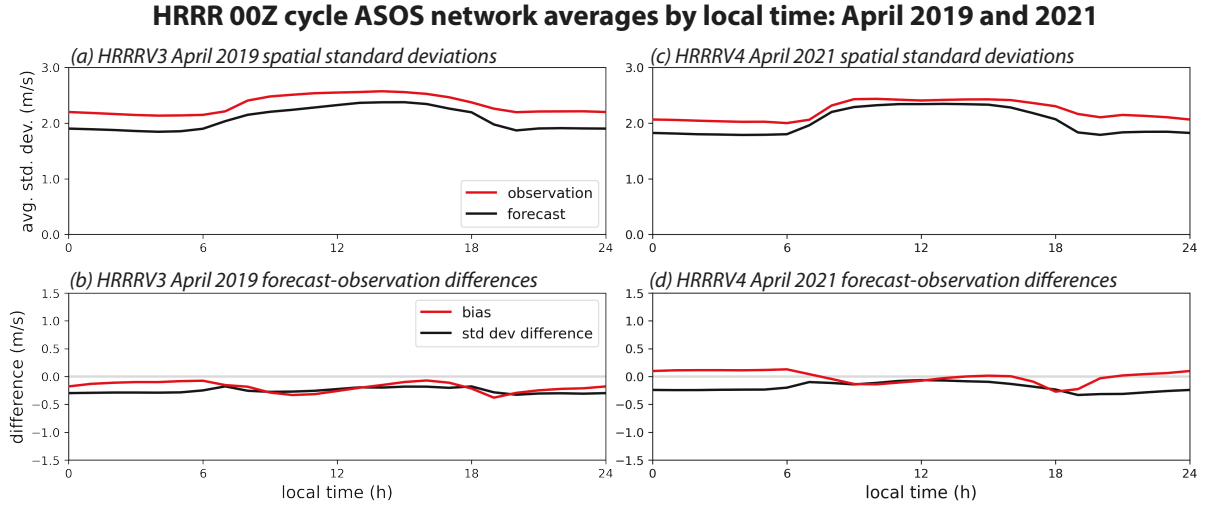


Fig. 5. Similar to Fig. 4 but expressed in terms of local time.

serve different needs and represent markedly different siting philosophies. Instead of being largely colocated with airports, NYSM stations sample the landscapes and geography of the state.

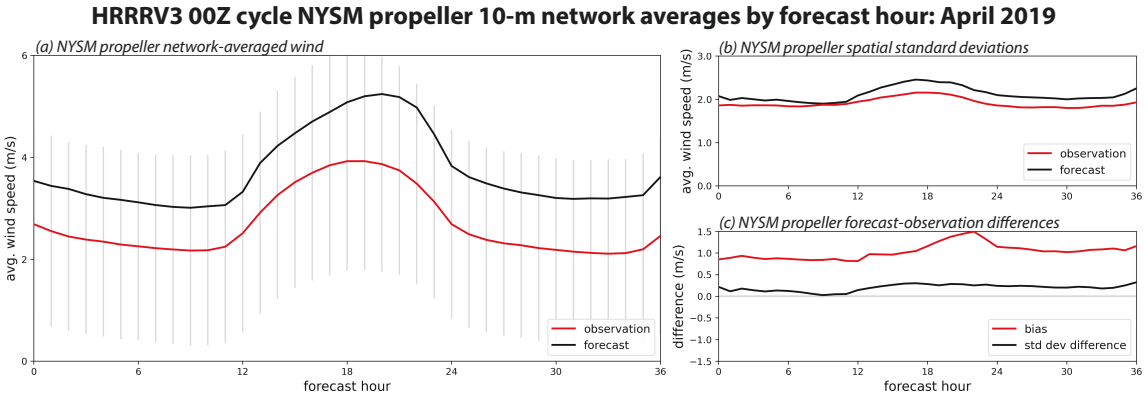


FIG. 6. Similar to Figs. 3a and 5a,b but for the NYSM propeller observations and forecasts.

b. Analysis by station

The present study also enhanced the station-based analysis of FG20 and the previously cited work on Santa Ana winds. We started by comparing forecast and observed sustained winds averaged over all available pairs for each station (Fig. 7a). Each dot is an ASOS (black) or NYSM (orange) station. Regarding the ASOS sites, while there are a few, non-impactful outliers, the squared linear correlation coefficient between the series is moderately high ($R^2 = 0.56$) and largely arrayed along the 1:1 correspondence line. NYSM stations are generally found beneath the 1:1 line, consistent with the overprediction already demonstrated. The least-squares fit shown was based solely on the 807 ASOS sites.

The relationship between forecast wind bias and various variables is examined in Fig. 8. Similar to previous studies already cited, the forecasts were not correlated with the bias (Fig. 8a), even for NYSM stations (orange circles). However, the observations were significantly and negatively correlated with bias (Fig. 8b), indicating overprediction of calmer sites and underprediction at windier locations. The NYSM stations do not appear to be exceptional, apart from the fact that as a relatively low wind speed network their sites are more likely to be associated with positive biases. A comparable analysis using the NYSM's sonic observations was only subtly different (not shown).

CF18 demonstrated (their Fig. 11d) that the forecast wind bias was also positively correlated with the station gust factor, which could be expected because GF incorporates the observed wind. They used station GF *relative to the network average value* to interpret the forecast bias and infer

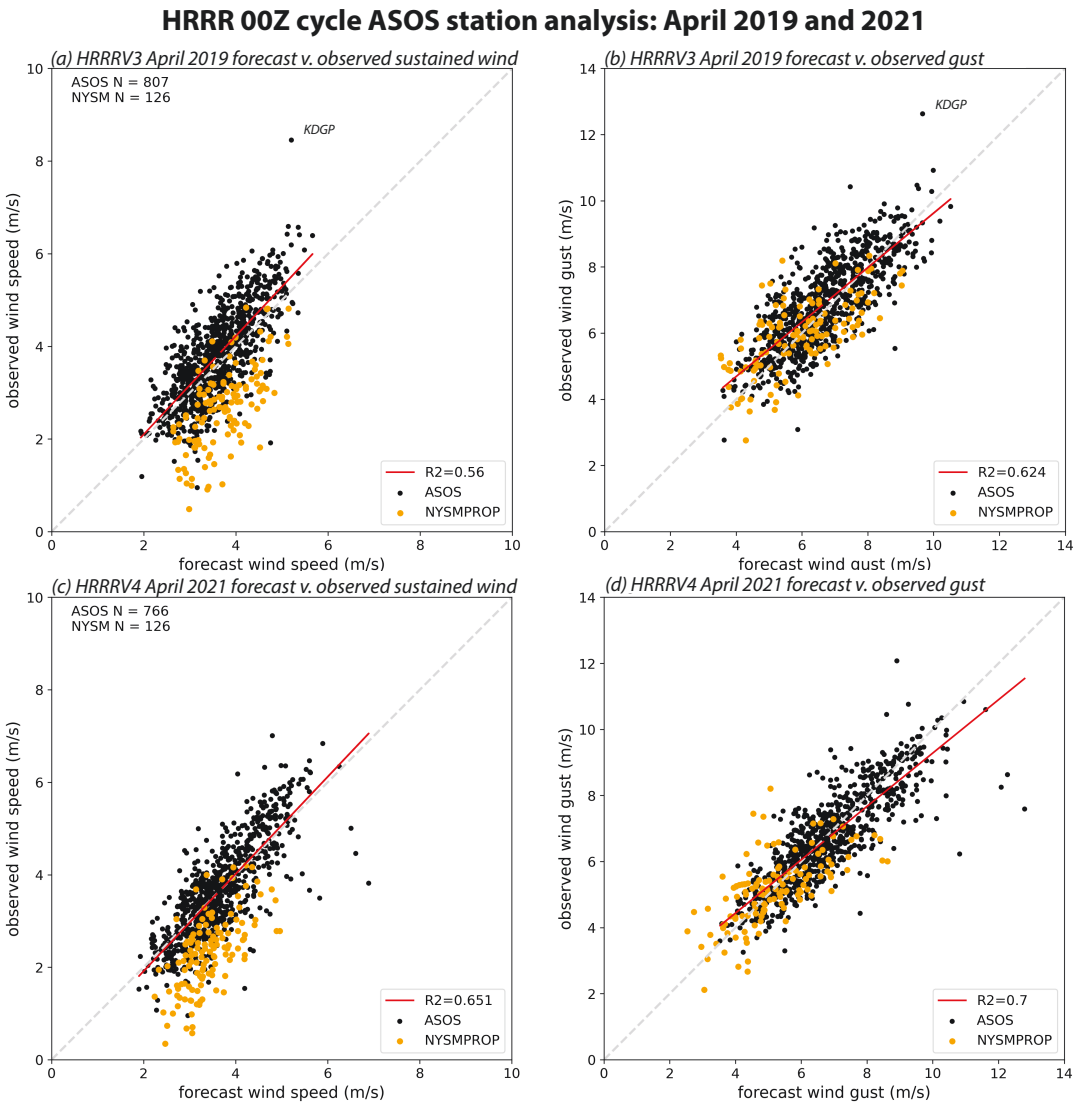


FIG. 7. Forecast vs. observed (a) wind and (b) and gust speeds averaged over all forecast/observation pairs from April 2019 or April 2021, aggregated by station. Here each dot is an individual station, either ASOS (black) or NYSM propeller (orange). Gust forecasts used the CF18 strategy (see text). Least squares fits (red lines) are based only on ASOS sites. Panels (c) and (d) are for April 2021. For April 2019, ASOS station KDGP is identified. This station had insufficient observations for April 2021.

site exposure. Locations with significant obstructions would be expected to have relatively lower wind speeds than similar although unobstructed sites, but short-period gusts might be anticipated to be less impacted, leading to higher GF values. Wind speeds at these stations would be expected to be overforecast because the model cannot "see" and account for these obstructions. In contrast,

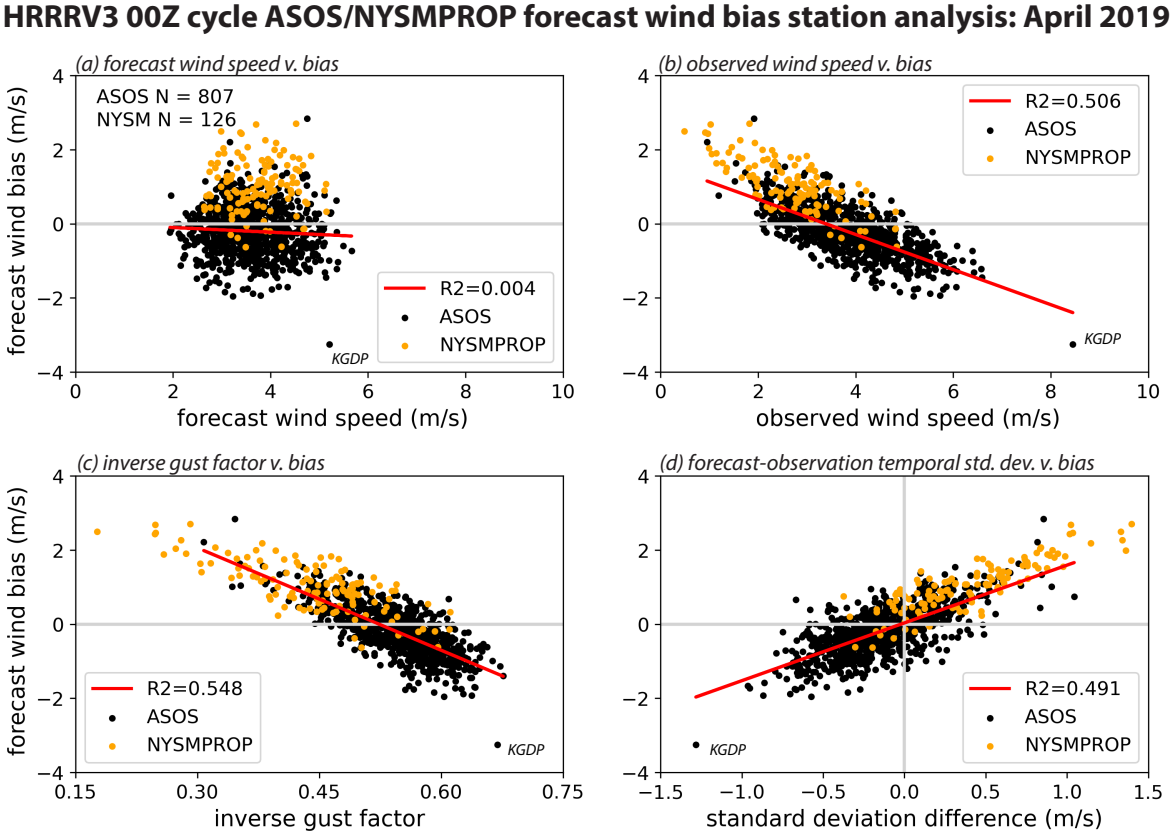


FIG. 8. Station averages from April 2019 of (a) forecast wind speed, (b) observed wind speed (c), inverse gust factor, and (d) temporal standard deviation difference (forecast-observation) presented vs. station average forecast wind bias for ASOS (black) and NYSM (orange) stations. Least squares fits (red lines) shown only incorporated ASOS sites. ASOS station KDGP is identified.

sites with lower GFs might have local features, such as hills, that might help speed up the wind relative to a more average setting. These stations would likely be underpredicted.

In Fig. 8c, we see a sizable negative correlation between bias and GF, although here we have instead elected to employ its reciprocal, the inverse gust factor (iGF), because it improves the linear relationship with bias and is bounded between 0 and 1. GF and iGF are functions of the observational data only and we see the model tended to overpredict when the sustained wind speeds were particularly small relative to the gust and underpredict when they were more comparable.

CF18 also considered a simple gust parameterization that was inspired by the association between bias and GF (and thus iGF). That strategy partially compensated for the biases in the sustained

wind forecasts by applying the network-average gust factor to *all* wind forecasts, yielding less biased gust predictions. Underpredicted stations also tended to have smaller GF (larger iGF) values than average, so multiplying the too-low speed forecasts by the network average at least partially mitigated the model's negative sustained wind bias. Similarly, overpredicted sites often had larger than average GFs (smaller iGFs) so multiplying the positively biased forecasts by the smaller network-average GF compensated for some of the overprediction.

This idea was applied to the April 2019 HRRR forecasts and is shown in Fig. 7b. In this case, ASOS wind forecasts were multiplied by 1.86, being roughly the network's average GF for the hourly maximum gust. This GF was applied to forecasts made for the top of the hour because we have insufficient information to determine the hourly mean forecast wind speed. With that caveat, we note this very simple gust parameterization performed quite well, with an even higher R^2 (0.62) than the forecast/observed wind relationship. Again there is a tendency for forecast/observation pairs to spread along the 1:1 line.

The CF18 parameterization implicitly presumed the network-averaged forecast wind bias was negligible so application of a single GF value could mitigate errors relative to the average. That is not the case for the NYSM. Figure 7b also shows (again in orange) NYSM gust forecasts made using that network's average GF (2.21), after adjustment for the mean forecast wind bias of about 1 m s^{-1}. Compared to the sustained winds, these gust forecast/observation pairs clustered much closer to the 1:1 line.

Finally, Fig. 8d demonstrates that the difference between forecast and observation temporal standard deviation was also well-correlated with forecast bias. Note now the standard deviations represent the *temporal* variability of the forecasts and observations at each station. Stations at which the forecasts have more variability than the observations tended to be overpredicted with respect to wind speed and underprediction often resulted at stations where the observations had more variation. However, as with GF and iGF, this variable is not independent of the observed wind. The standard deviation of a variable like wind speed, which has the hard constraint of being non-negative, can (and, although not shown, generally does) increase with the variable magnitude.

Spatial plots (Fig. 9) were examined to look for patterns. While the average forecast wind bias, computed over all stations and forecast hours, was only -0.2 m s^{-1} (cf. Fig. 3a), it remains that 507 of the 807 stations (63%) were underpredicted in the mean. Figure 9a shows that the

positively biased stations were concentrated in the Southeast, the Appalachians generally, and into the Northeast where forested land is more common (Fig. 2a). In Fig. 9b, marker size reflects the squared linear correlation between the forecast and observed winds, based on an average of 1000+ forecast/observation pairs from each site. R^2 values ranged between 0.03 (KP69) and 0.77 (KARR - Aurora, IL) with a mean of 0.57 and median of 0.59. Correlations were high throughout most of the country, even in the Southeast where mean winds were relatively light, and lowest in the mountainous West. Like the correlation coefficient, R^2 is not sensitive to means or mean differences between series and is most likely low where the predictions are somewhat out of phase with the measurements. The concentration of low correlations in the western CONUS may reflect the influence of local features on diurnal winds that the model fails to properly represent.

Figure 9c reveals how the temporal standard deviation difference between the forecasts and observations varied spatially. Figure 8d showed that the former tended to be the larger when observed wind speeds were low and forecasts were positively biased. The mean and median differences were -0.15 and -0.17 m s^{-1}, respectively, with 581 (72%) of the sites having less variability among the forecasts than the observations. Note that the large red dots (representing larger forecast than observation variability) are few in number and widely scattered. These are stations having significant local obstructions near the ASOS installations. For those sites, observation variability was likely suppressed by limited anemometer exposure. This measure could be used to identify problem sites for potential removal from analyses and data assimilations.

Taken together, this analysis suggests that the small negative forecast bias seen in the network averaged winds (Fig. 3) is more significant than it might appear at first glance. The majority of locations have insufficient forecast variability that is strongly correlated with negative biases. This suggests the model is not capturing something that is important to determining real winds measured in the field. However, this is partly compensated by the inclusion of stations that are not at airports and/or have obvious siting issues. Had those sites been removed from the analysis, the underprediction would have been more pronounced. The model is still very skillful but steps could be taken to address its tendency to understate the mean winds at better exposed locations.

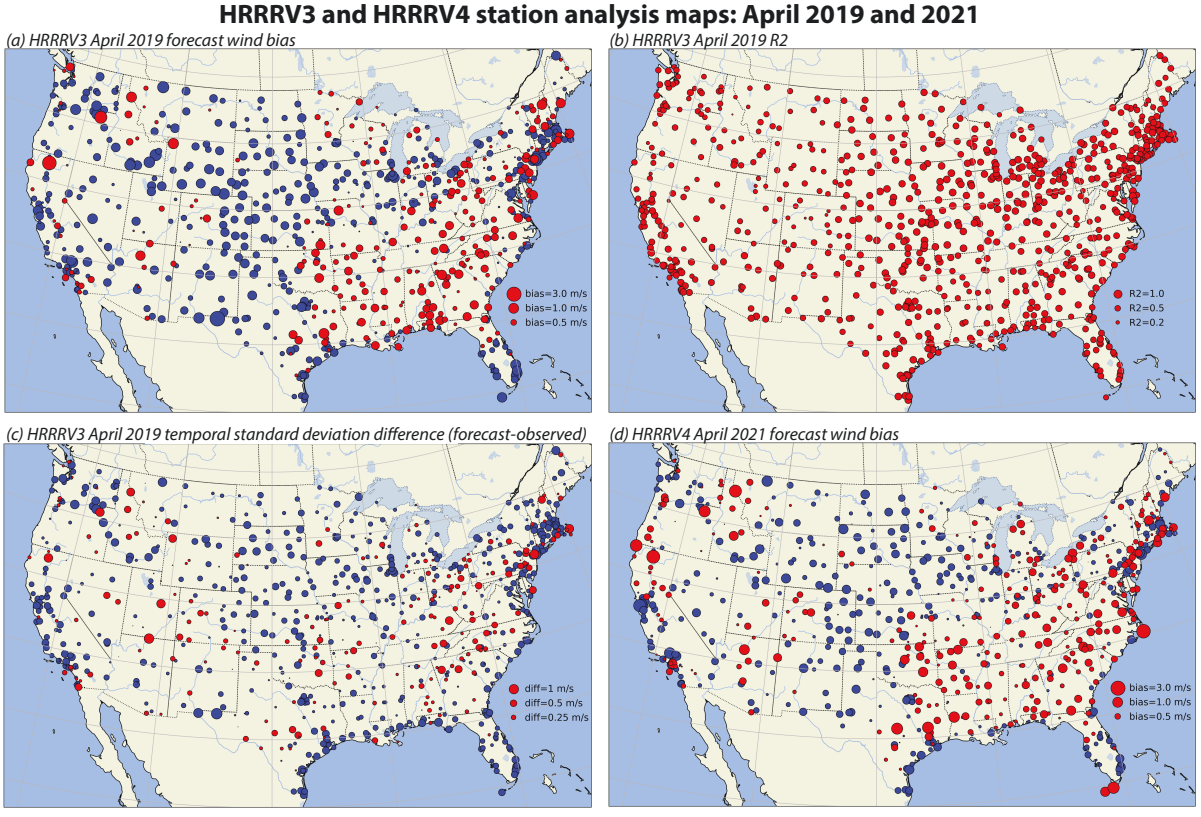


FIG. 9. Spatial plots of (a) average forecast wind bias, (b) forecast-observation squared correlation R^2, and (c) temporal standard deviation difference (forecast-observed) for April 2019. Panel (d) shows average forecast wind bias for April 2021. For (a), (c), (d) positive values are red, negative are blue.

c. Analysis of forecast/observation pairs

In their analysis, FG20 examined scatterplots involving all individual ASOS forecast and observation pairs over a full month and this provided insight into the source of forecast biases. Here, we improve and extend that analysis, examining all 827,230 April 2019 pairs[6]. This represents the concatenation of forecasts and observations from 807 ASOS stations and all forecast hours from the daily 36-h HRRRV3 00 UTC cycle forecasts. Note that many observations were paired with more than one forecast.

All ASOS forecast/observation pairs are presented as a heatmap, color coded by point density, in Fig. 10a. Although there is scatter about the 1:1 correspondence line, there is a reasonably good

[6]There are fewer pairs in the present analysis than in FG20 (851,550) owing to the more stringent restrictions employed in the construction of hourly-averaged observations.

relationship (R^2 = 0.56) between these variables, comparable to that seen in the station-averaged analysis (Fig. 7a). The majority of observations and forecasts represented speeds less than 5 m s^{-1}, and this fact drives the relationship. For higher observed winds, however, the forecasts still largely spread along the 1:1 line, indicating some usable skill. Similarly, all forecast gusts – created via the constant GF of 1.86 – are plotted against observed gusts in Fig. 10b. As was the case with the station-averaged analysis, the correlation is higher for the gust forecasts than their sustained wind counterparts.

However, these same data viewed as histograms (Fig. 11) demonstrate that the forecast and observed wind and gust distributions had distinctly different shapes. The forecasts possessed a narrower peak such that the occurrence of both lower and higher observed winds was relatively more frequent. This result was suspected in the discussion of Fig. 4a above. Motivated by Fig. 8c, we also examined histograms of winds and gusts partitioned into lower and higher GF segments (Fig. 12). Forecast and observation pairs were separated into two groups based on the GF associated with the observation relative to the median value (about 1.81). With respect to winds (panels a, b) there is a much larger shift between the segments in the shapes of the observed wind distributions than for the forecasts. When the GF is lower, the observed distribution is shifted rightward, resulting in more observations than forecasts of values exceeding 3.5 m s^{-1}. In contrast, observations in the high GF half are skewed towards lower speeds, resulting in a mean positive bias.

To reiterate, the network mean bias of ASOS forecasts was nearly zero (Fig. 3a,b), but the bias was biased such that stations having lower average wind speeds were overpredicted while windier ones were underforecast (Figs. 8c, 11a). The constant GF algorithm exploits this systemic tendency to underpredict at sites where GFs lower than the network average and overpredict at the others by multiplying these biased wind forecasts by a single number (the network average GF), the result being *less biased gust forecasts* (Fig. 12c,d). For locations in space and/or instances in time where the observed GF was lower than the network average, multiplying by the larger average value helped shift the forecast gusts more into alignment with the observations (Fig. 12c). Similarly, multiplying forecasts of high GF instances or locations by the smaller network average helped correct for the deficiencies seen among the sustained winds. The result is not perfect and we have already seen that when the two segments are recombined (i.e., Fig. 11a), the forecast range is too narrow relative

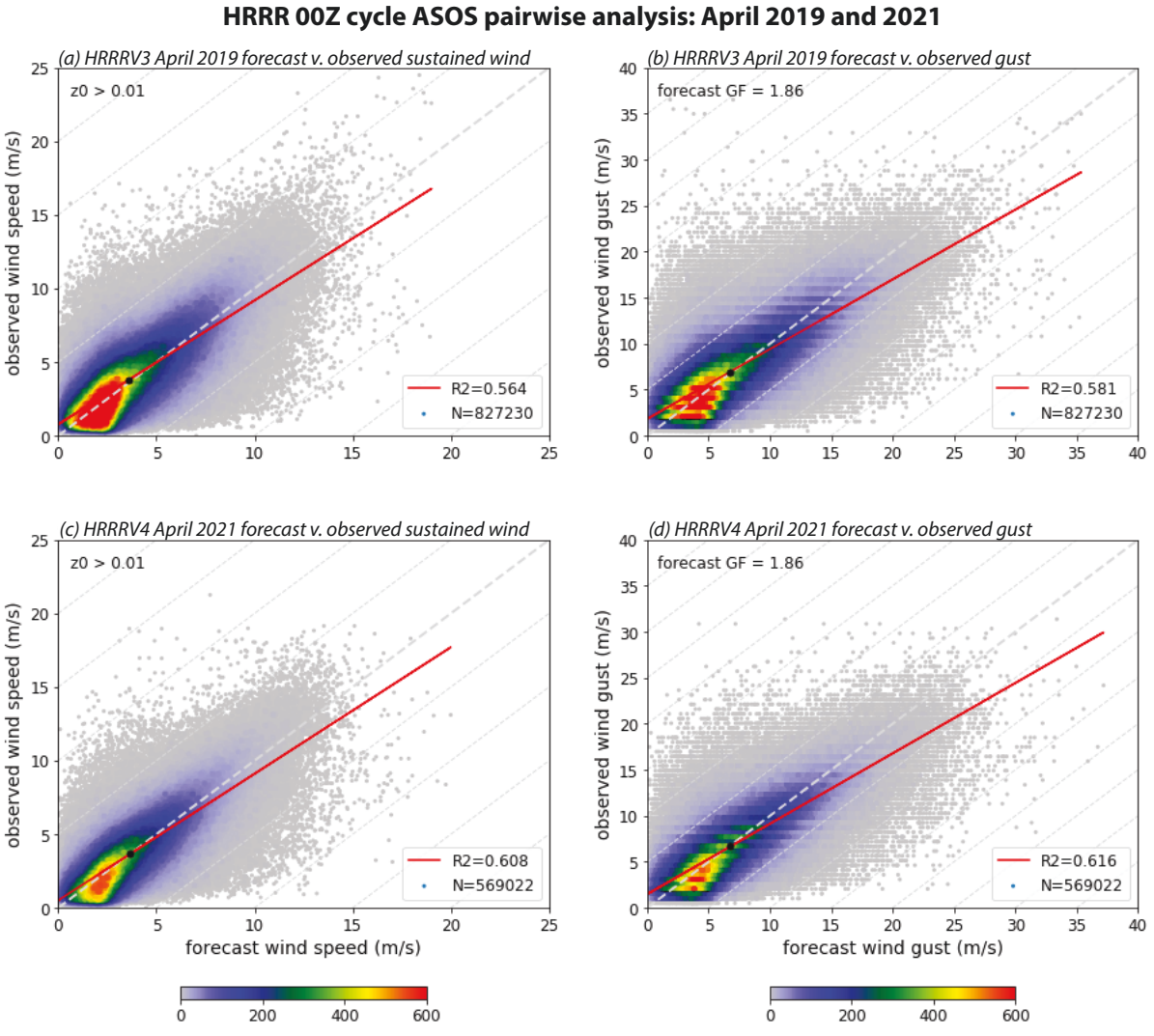


FIG. 10. All forecast/observation pairs of wind speed (left) and gust (right) for ASOS stations during April 2019 (top) and April 2021 (bottom). Color shading indicates point density and the linear regression line for each is shown in red. Gust forecasts were produced using the ASOS network average GF. The black dot is the joint mean. $z_0 > 0.01$ indicates that sites misclassified as being over water have been removed. This is true for all analyses in this study.

to the observations. In the next section, we will discover reasons for the excessive sharpness in the forecast distributions.

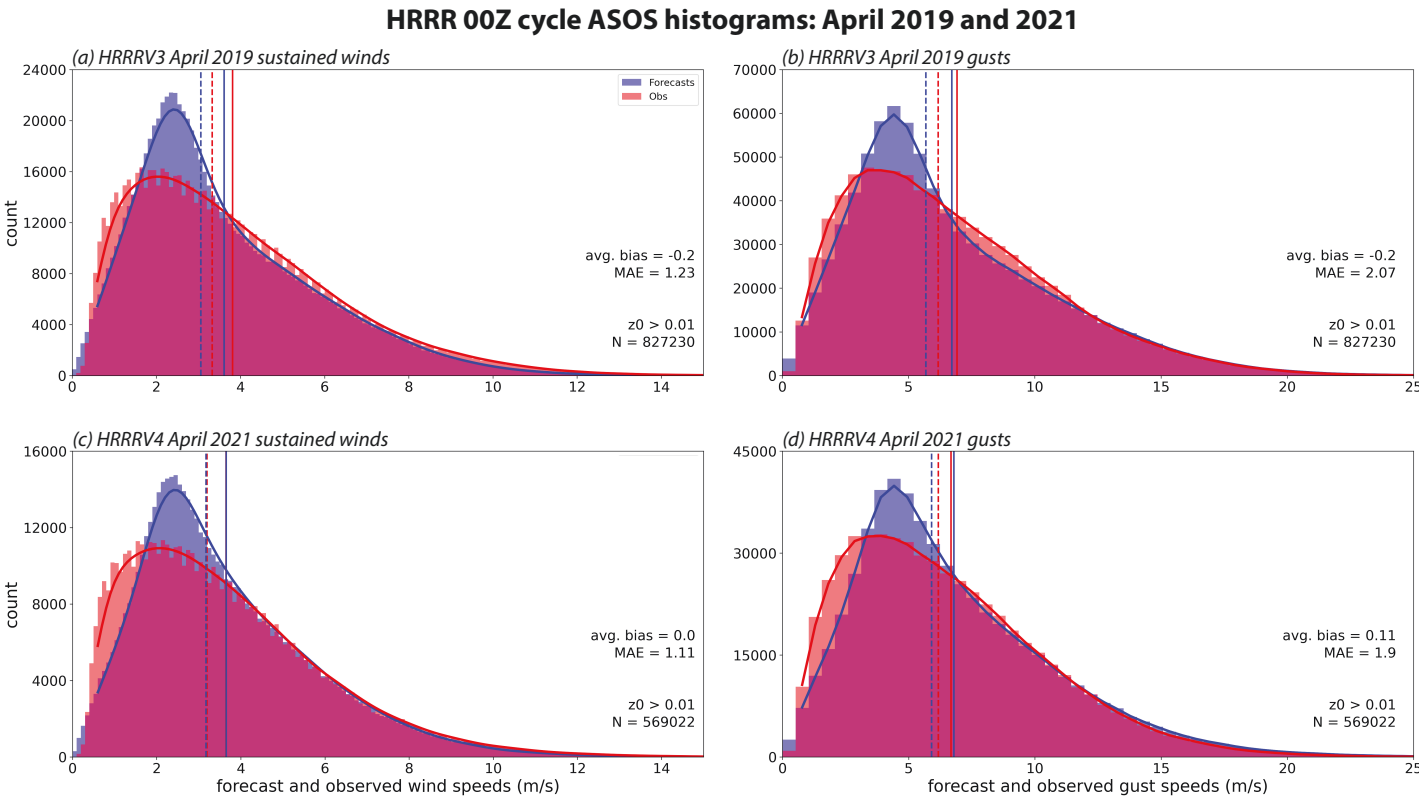


FIG. 11. Histograms of all forecast (blue) and observed (red) wind speeds (left) and gusts (right) for April 2019 (top) and April 2021 (bottom). MAE stands for mean absolute error. Vertical solid lines indicate mean values and dashed lines represent median values. A wider bin size was used for gusts owing to the coarse (1 kt) precision of hourly maximum gust observations.

d. The roles of landuse and local time

The potential roles of landuse type and local time were investigated to understand the differences between the observations and forecasts, especially with respect to their distributional shapes as seen in Fig. 11. As noted earlier, WRF-ARW uses fractional landuse allocations (cf. Fig. 2b) and the focus here is on the largest, or primary, assignment. For HRRRV3 and April 2019, 41% of the ASOS stations had a primary classification of cropland, 24% were urban, 14% had grassland, and 6% were given open shrubland assignments. The various forested land classes, including deciduous, evergreen, and mixed forests, accounted for about 11% of the ASOS sites. While unsurprising, it is clear that the urban landuse type is substantially overrepresented in the ASOS network relative to the CONUS landscape (see, for example, the bright red areas in Figs. 2a,c).

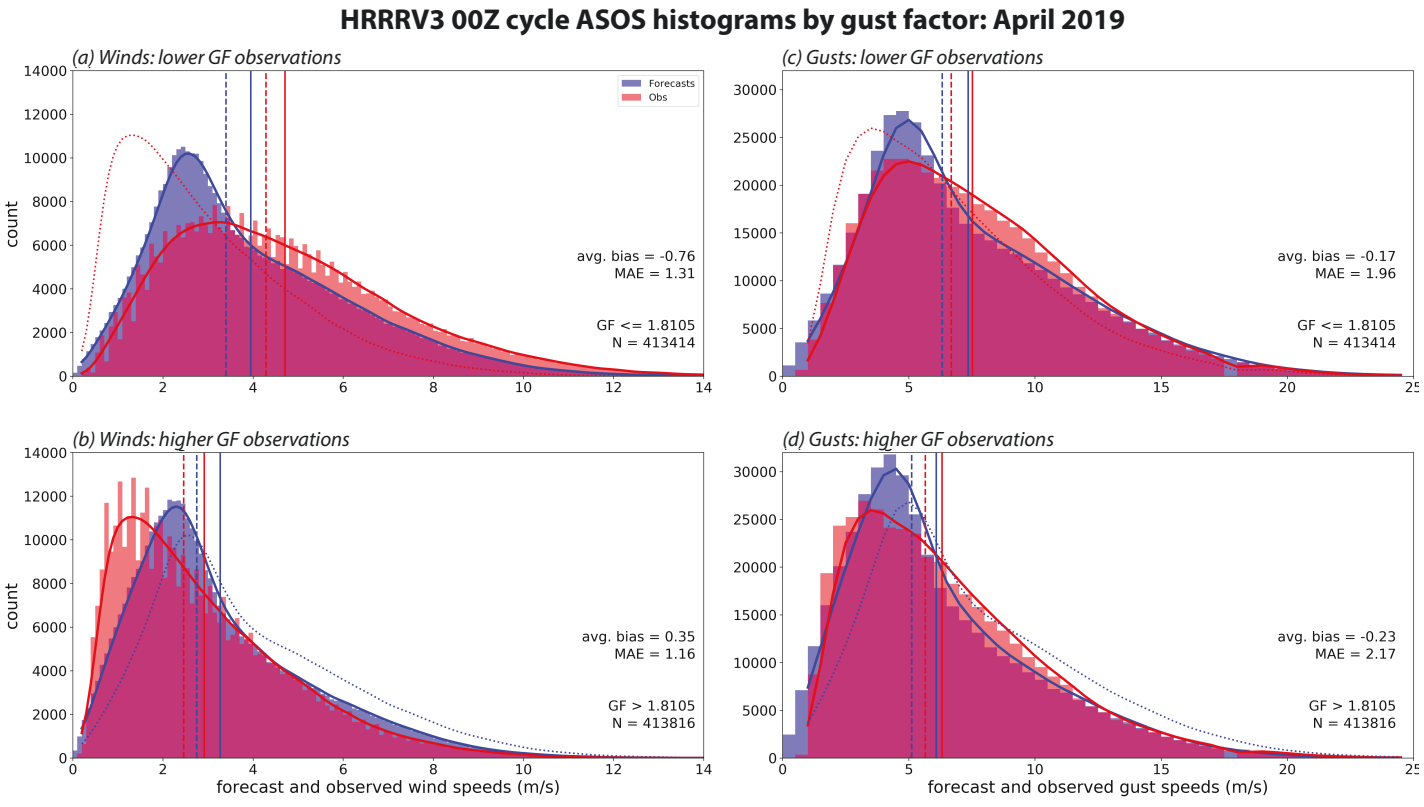


FIG. 12. Similar to Fig. 11a,b but showing the April 2019 wind and gust distributions subdivided at the median GF (about 1.81). The dotted red curves in (a) and (c) represent the higher GF observation distributions, to facilitate comparison. Similarly, the dotted blue curves in (b) and (d) represent the lower GF forecast distributions. The sawtooth variation in the observations in (a) and (b) is another consequence of the gust data precision.

Figure 13 reveals the existence of a robust association between primary assignment and forecast wind bias. Each class possesses two horizontal bars, representing the average bias (blue, units m s^{-1}) among stations with that classification and their weighted contribution (red, units dm s^{-1} for convenience) reflecting station count towards the network-average bias of -0.2 m s^{-1}. The most negative bias (-0.6 m s^{-1}) was associated with the open shrublands stations but the urban and grassland sites had larger weighted shares owing to their larger station counts. Similarly, although cropland stations had a small class-average bias (-0.08 m s^{-1}), their aggregate effect was not minor owing to their ubiquity (41% of stations). In contrast, the roughly 11% of installations residing in forested grid cells were positively biased, by as much as +0.52 m s^{-1} in the evergreen needleleaf

cells[7]. If these overpredictions were resolved in isolation, the network-averaged skill would actually *decrease*.

Figure 14 presents histograms of forecast and observed sustained wind similar to Fig. 11a but have been segregated by selected primary landuse classes. All of the forecast distributions are too sharp and narrow relative to the observations. In urban areas (panel a), the observed wind distribution has spread farther to the right, revealing underforecasts of speeds exceeding about 3.5 m s^{-1}. That tendency was even more pronounced in the grassland and open shrubland group (panel b), which have been combined owing to their similarity. The small negative bias in the cropland class (panel c) occurred despite general overprediction of winds weaker than 1.5 m s^{-1}.

Importantly, the model has obviously failed to properly represent the general slowness of the winds in the forested areas (Fig. 14d). This elucidates why the network-averaged sustained winds from the NYSM were so overpredicted. Note that the Mesonet's sustained wind histograms (Fig. 15) bear a strong resemblance to that of the ASOS forested class, independent of anenometer type. While only 11% of the ASOS sites were classified as forested in the HRRRV3, that category represented 43% of the Mesonet stations, and thus it exerts a powerful influence on this network's average. Landuse type can affect wind forecasts through the roughness length, z_0. Although this would require testing, it is not clear that simply raising z_0 would improve these predictions because the more serious issue is site exposure.

When the day is subdivided into four 6-hour segments as in Fig. 16, we clearly see the underprediction of observed ASOS winds exceeding 4 m s^{-1} seen in Fig. 11 is largely confined to the nocturnal period between 6 PM and 6 AM local time (LT), when the boundary layer is likely to be stable[8]. This period is also largely responsible for the distributional differences between the forecasts and observations noted above. The frequency of relatively larger observed wind speeds at night was sufficient to make the mean bias of forecast/observation pairs to be negative, even though the model generated too few low speed predictions. This may represent a problem with how the model handles the stable boundary layer and its intermittent, localized turbulence (cf. Medeiros and Fitzjarrald 2014, 2015). In contrast, the daytime period of 6 AM to 6 PM LT (panels b and c) seems to be rather well represented in the HRRRV3 forecasts, albeit with a small underrepresentation at higher wind speeds ($\geq$ 8 m s^{-1}) that also led to small negative net biases.

[7]Precise percentages vary slightly between the station and forecast/observation pair analyses owing to minor data dropouts.

[8]The number of forecast/observation pairs varies among the segments because we are only using the 00 UTC cycle and its 36 h simulations, which means some times have more forecasts than others.

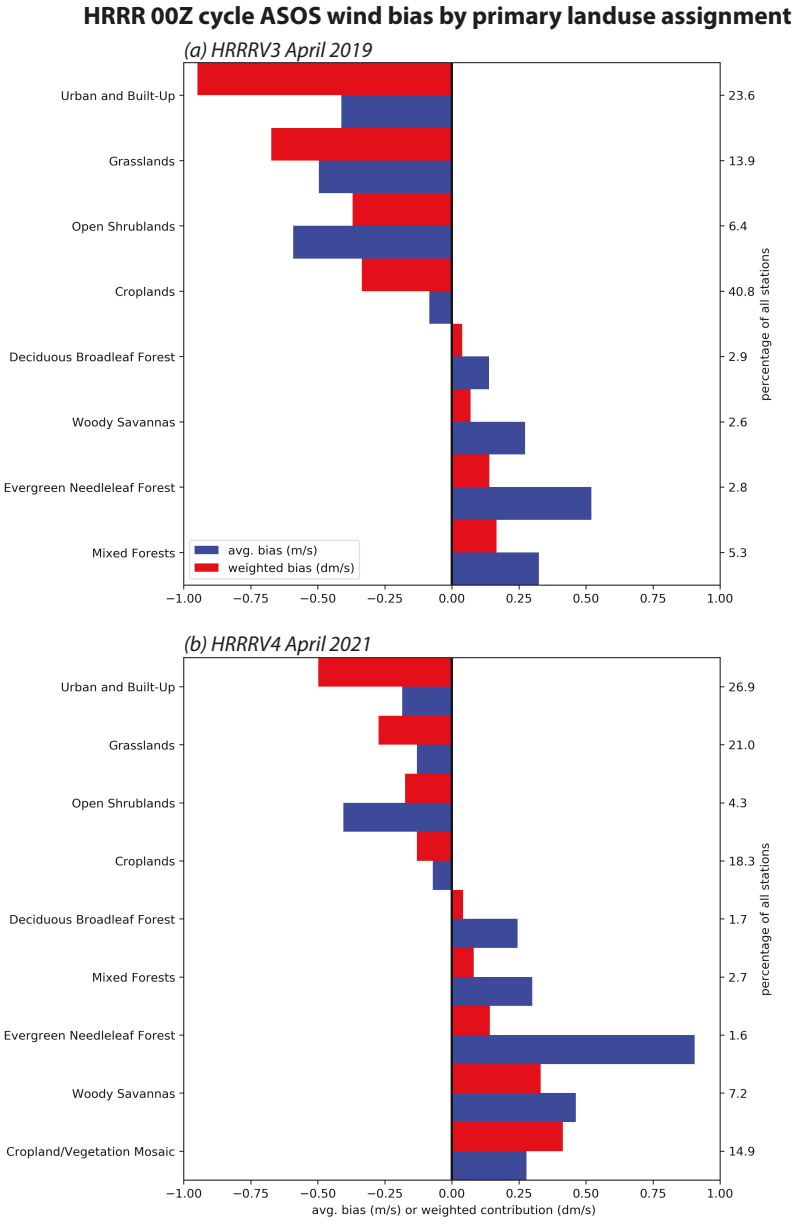
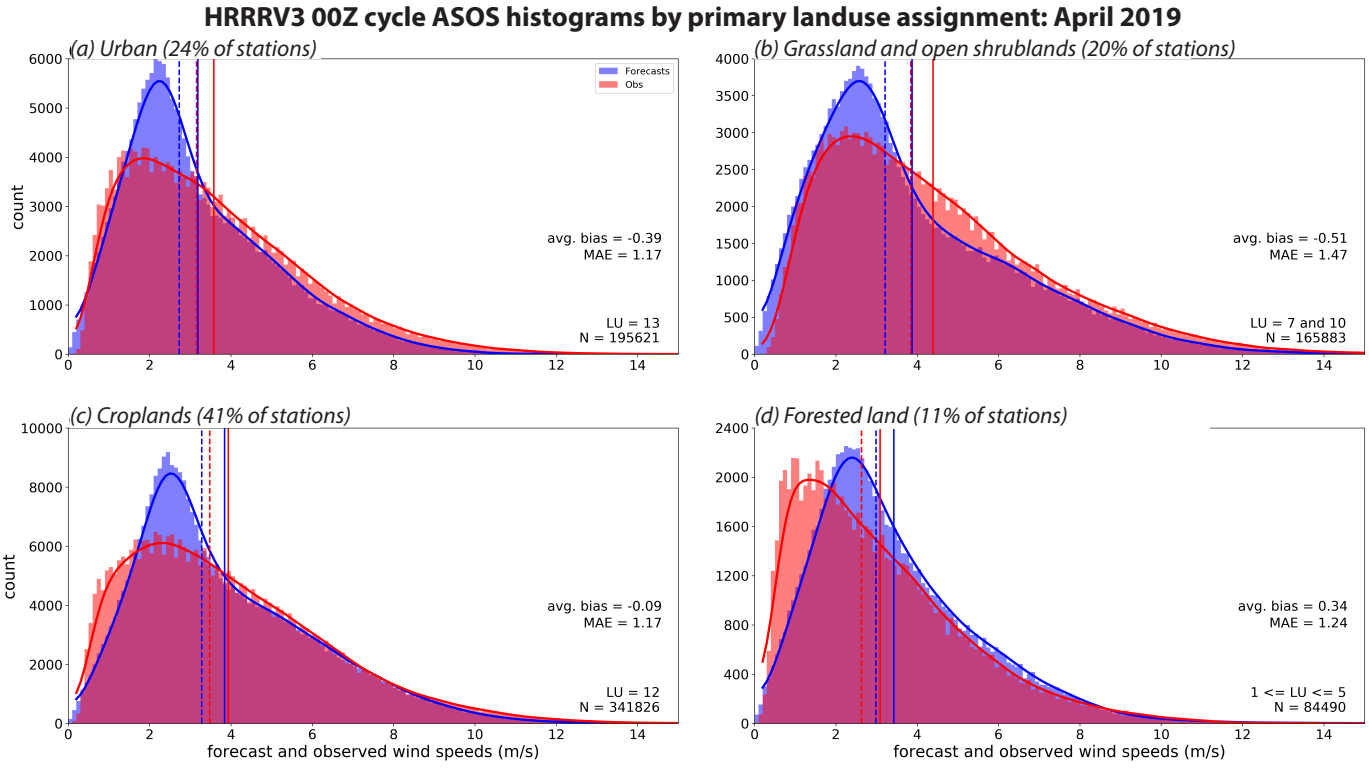
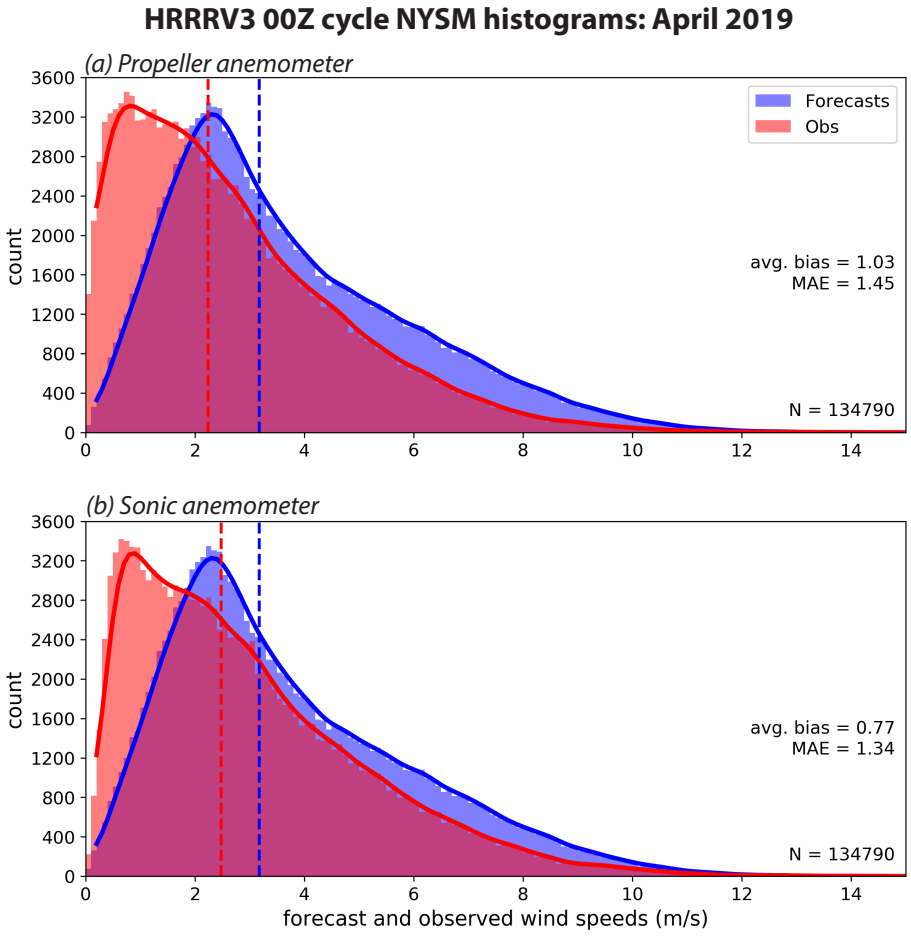


FIG. 13. Average forecast wind bias (blue bars) aggregated over ASOS stations having same HRRR primary landuse assignments for (a) April 2019, and (b) April 2021. Red bars represent the weighted contribution of that class towards the network-average bias. Landuse classes are ordered by weighted bias. Right axis: percentage of stations having this primary classification. Precise percentages vary slightly between the station and forecast/observation pair analyses owing to minor data dropouts.



427 FIG. 14. Similar to Fig. 11a but segregated by primary landuse (LU) category from the MODIS 21-class
428 database used by HRRRV3: (a) urban, (b) grasslands and open shrublands, (c) croplands, and (d) forested land
429 (including deciduous, evergreen, and mixed forest).

452 Those histograms aggregated all landuse classes. Figure 17 focuses on the 6 PM to midnight
453 LT period differentiated by the landuse groupings examined in Fig. 14. Only the forested lands
454 (panel d) did not have characteristic underprediction of relatively faster winds, again reflecting
455 the less than optimal handling of those areas in the model. For the afternoon (noon to 6 PM LT)
456 period (Fig. 18), however, only the urban classification (panel a) failed to capture the frequency of
457 stronger winds. Thus, except in the vicinity of cities, the model's inability to capture the frequency
458 of stronger winds appears to be a nocturnal issue and one that might be addressed by reconsidering
459 assumptions employed in the stable boundary layer regime. It is surmised that the urban issue
460 may also stem from overly high specifications of surface roughness in those areas. While many
461 airports are located in grids designated as urban, that does not mean that the local environment
462 of the airport is truly city-like. Finally, we reiterate that resolving the issue with forested land or



439 FIG. 15. Similar to Fig. 11a but for April 2019 wind forecasts for NYSM sites compared to (a) propeller, and
440 (b) sonic observations.

463 removing those stations from the analysis would tend to make the nocturnal underprediction issue
464 appear worse.

467 4. HRRRV4 wind and gust evaluation for April 2021

468 Version 4 of the HRRR became operational on 2 December 2020. The revised model incorporated
469 a number of improvements to the planetary boundary layer and radiation schemes, the land surface
470 model, and numerical methods and diffusion, and adopted a new gravity wave drag treatment
471 (cf. Dowell and co authors 2022). It also shifted to the higher-resolution version of the MODIS

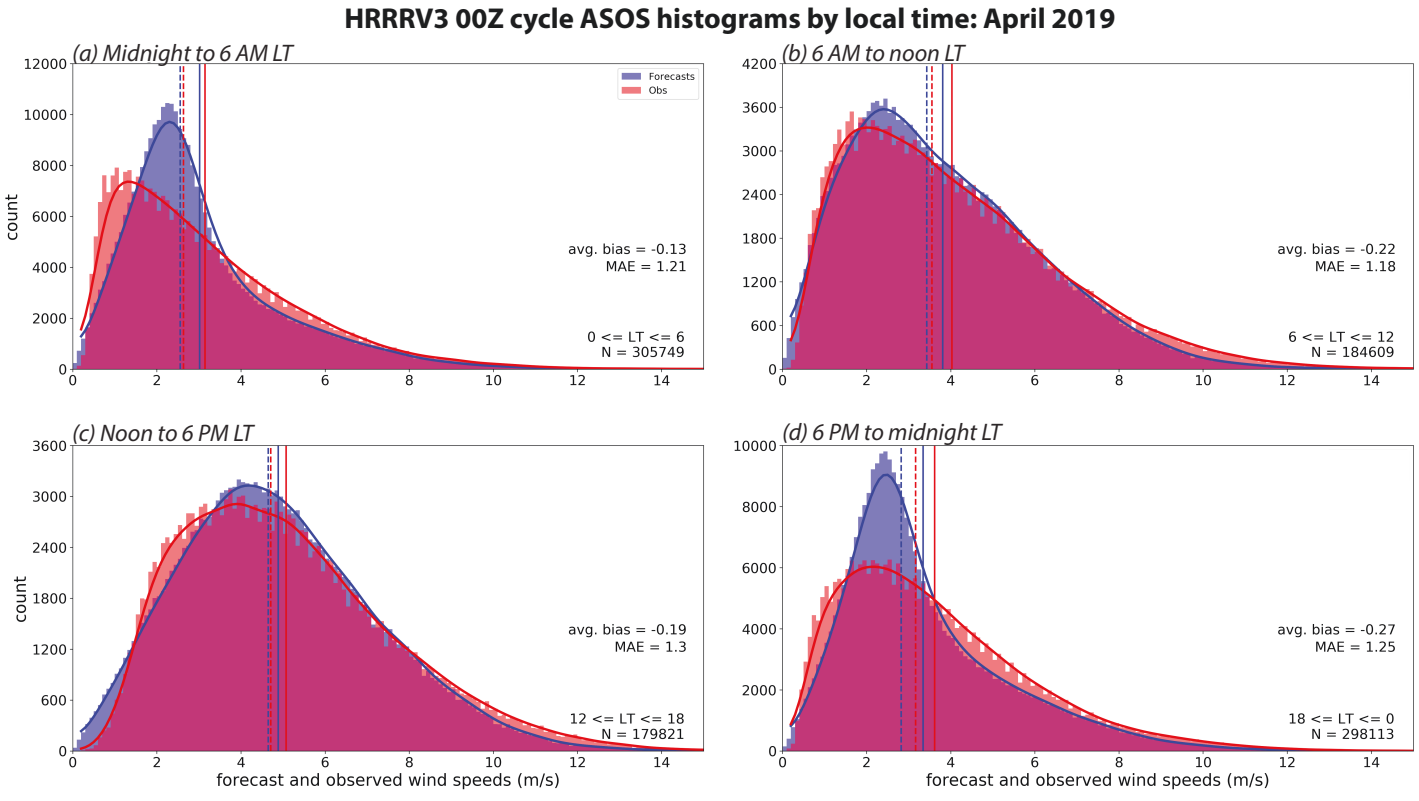


FIG. 16. Similar to Figs. 11a and 14 but segmented with respect to local time (LT).

landuse database that was released with WRF Preprocessing System (WPS) version 3.9 in 2017, with consequences as discussed presently.

Our examination of 10-m wind forecasts at ASOS stations from April 2021 from HRRRV4 emphasizes points of similarity and difference with the April 2019 HRRRV3 results. The verification was again restricted to the 00 UTC cycle and through forecast hour 36, even though V4 now integrates out two full days for that start time. Although not shown, monthly mean wind speeds were very comparable to April 2019 (Fig. 1). Unfortunately, relative to April 2019, there were more missing observations in the ASOS 1-min database in April 2021. As a consequence, the database of hourly mean sustained wind and maximum gust had 32% fewer observations than for April 2019, averaging about 15400 observation/forecast pairs per forecast hour instead of 22650. Only 766 sites remained after removal of misclassified stations and those with 500 or fewer observations. In our judgment, this does not negatively affect the evaluation.

There are more differences between these two MODIS-derived databases than just the resolution enhancement. In HRRRV4 (Fig. 2c), a large fraction of the original croplands class (#12, gold),

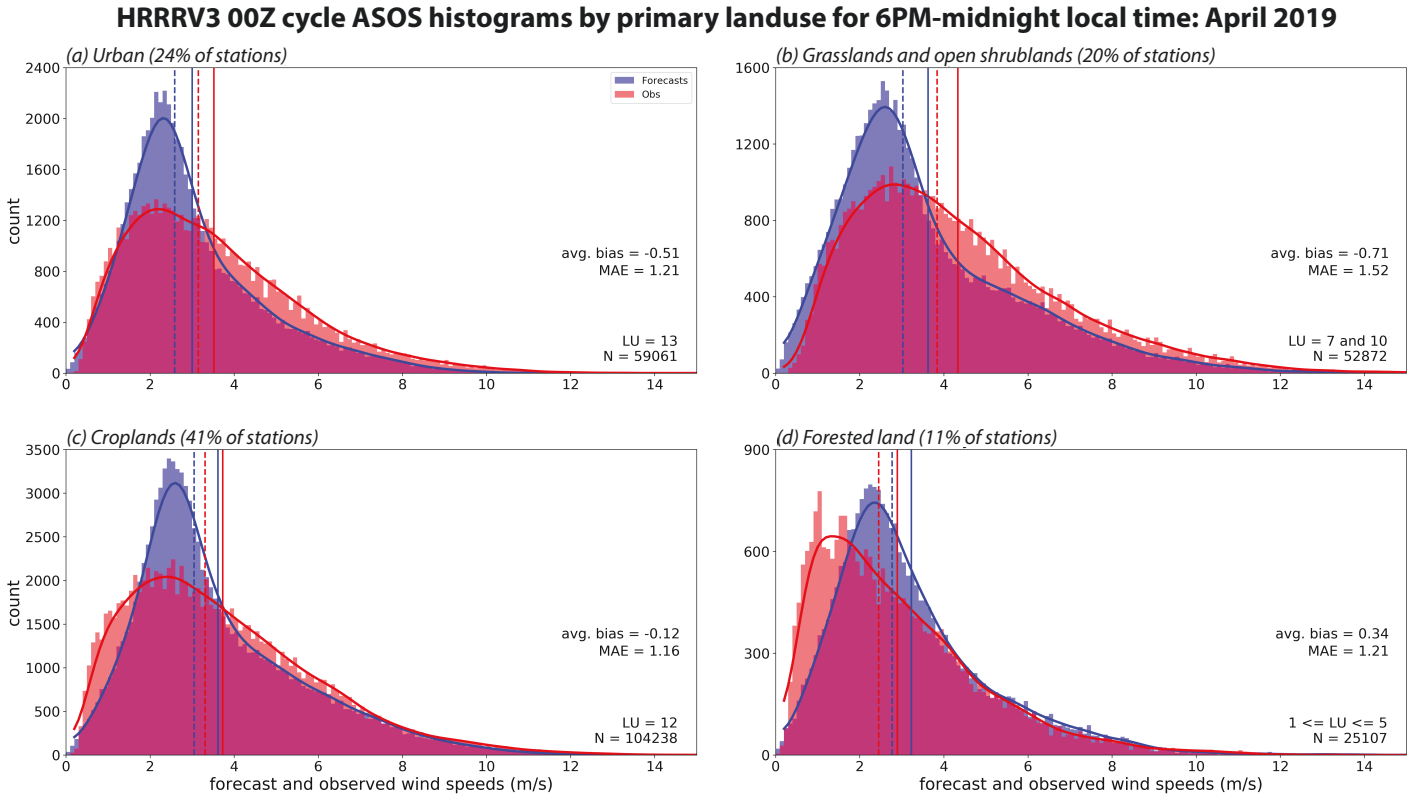


FIG. 17. Similar to Fig. 16a but focusing on the 6 PM to midnight LT period and separated into different landuse groupings.

especially in the eastern CONUS, has been transferred into the previously existing but unused "cropland/vegetation mosaic" group (#14, cyan). The croplands category presently accounts for only 18.3% of ASOS station primary assignments while the mosaic claims 14.9%. In the west, a portion of the open shrublands (#7, maroon) primary assignments have been reassigned as grasslands (#10, light green), constituting 4.3% and 21.0% of ASOS sites in the newer MODIS database, respectively. We have continued combining those landuse types owing to their similarity with respect to model performance. The HRRRV4 grassland area has also spread eastward into the former croplands, so the grassland and open shrubland combination now represented 25% of the April 2021 ASOS primary assignments, an increase of 5 percentage points. Some areas that had been assigned to one of the forest classes (categories 1-5) have been reclassified as woody savannas (#8), increasing its share of the network from 2.6% to 7.2%. Owing to their similarity,

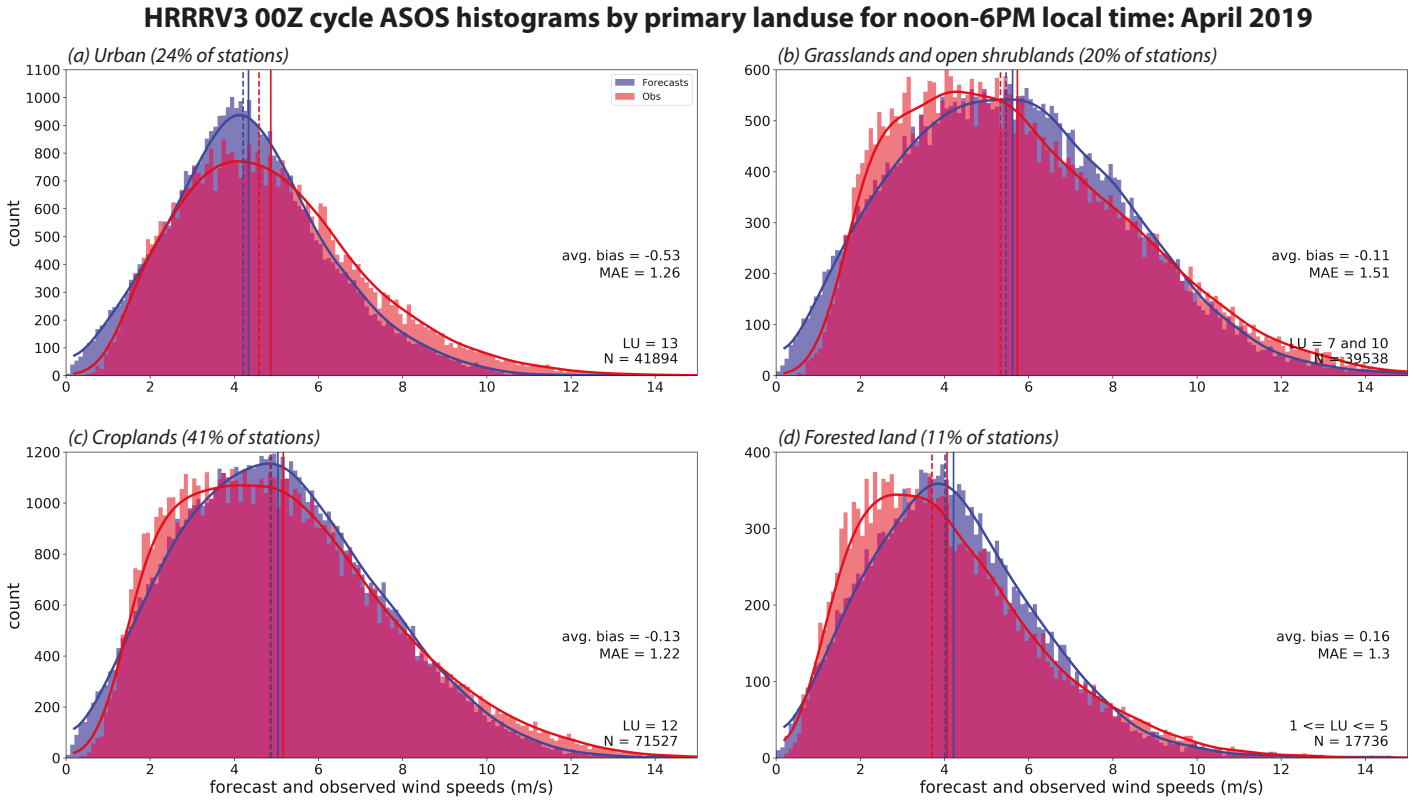


FIG. 18. As in Fig. 17 but focusing on the noon to 6 PM LT period.

class 8 was analyzed with the forested land, and this combination represented 13% of the ASOS stations retained in the April 2021 analysis.

As in Section 3, above, these are *primary* landuse assignments. The fractional landuse apportionments represent another difference with HRRRV3. In HRRRV4, 87% of ASOS stations reside in grid cells assigned more than one landuse class, up from 53% in V3 (compare Figs. 2 panels d and b), a consequence of V4's higher landuse resolution. The average fraction claimed by the primary class was 0.7, a decrease from 0.83 for V3. Again, this was relatively smaller for the forested group and also the new cropland/vegetation mosaic classes (both about 0.6) than for the urban and croplands (both ≈ 0.7) and grasslands (0.8). The HRRRV4 landscape is more finely divided and this makes analyzing by primary landuse assignment less precise, but again we find some value in this effort.

Figures 3c and 4c,d present the April 2021 forecast hour analysis. The small negative forecast bias that was previously seen in V3 has vanished (indeed, the mean bias is now essentially zero) although the spatial standard deviation of the forecasts was still smaller than that of the observations

at all forecast hours. The local time versions of these figures also revealed some improvements (Figs. 3d and 5c,d). Despite involving fewer sites, the station analysis results and conclusions were little changed. R^2 values for the sustained wind and gust fits were higher for both station-average (Figs. 7c,d) and pairwise (Fig. 10c,d) comparisons and (although not shown, see Gallagher 2021) the average forecast wind was again uncorrelated with bias but the higher wind stations were still underpredicted and lower sites overforecast in a manner that is predictable from iGF or GF[9]. In addition, the association between bias and the difference between forecast and observed temporal standard deviation remained (also not shown, cf. Gallagher 2021). Viewed spatially (Fig. 9d), forecast bias was still concentrated in the east CONUS in general and southeast in particular, although errors were somewhat smaller in magnitude.

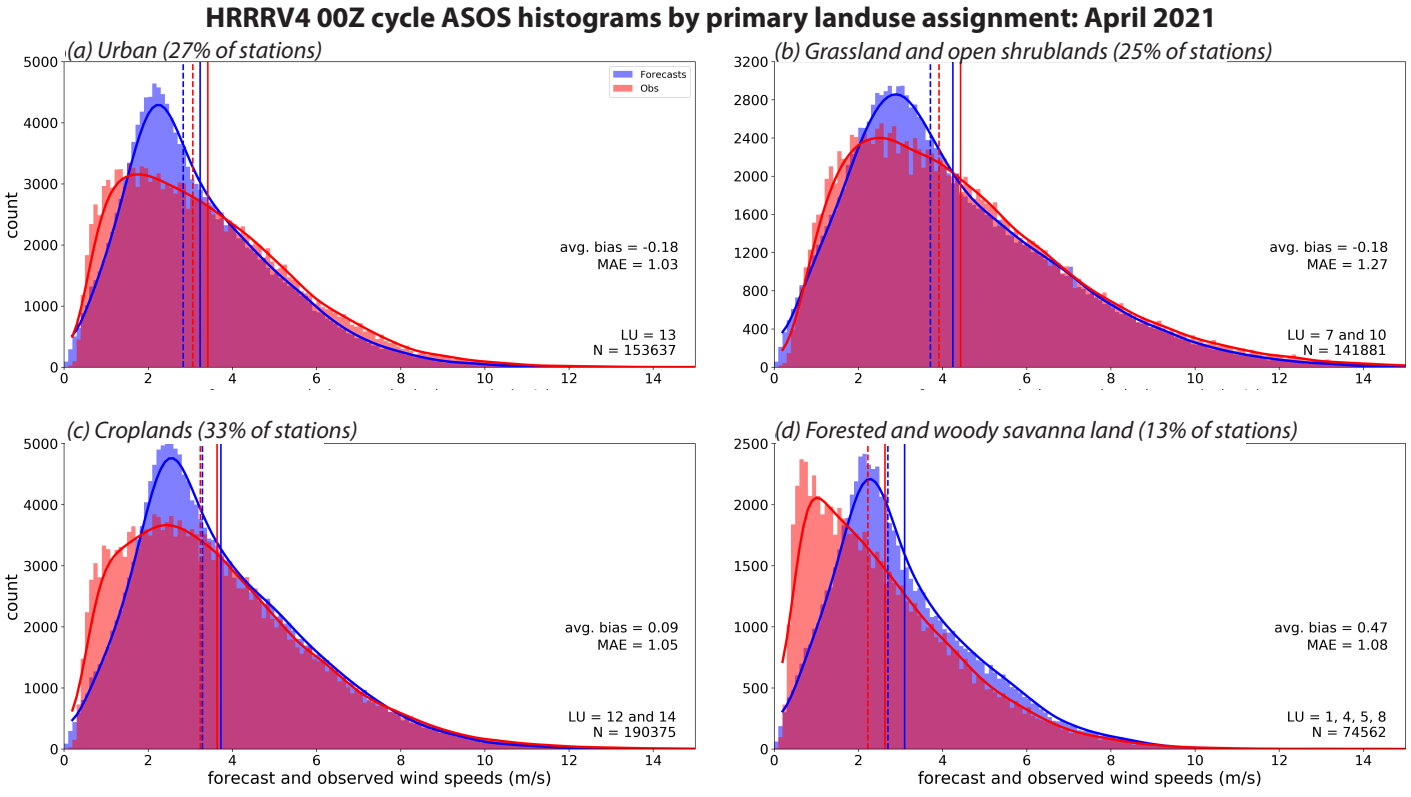
The wind and gust histograms (Fig. 11c,d) also suggest improvements relative to April 2019. However, the compensating errors between more densely treed areas (the forest and woody savannas categories) and the urban and grassland areas persisted (Fig. 13b). The now more spatially confined croplands class was still the best modeled and the newly separate mosaic group had a positive bias, which is unsurprising because much of the this group's stations are in the southeast, the site of lower wind observations (not shown, but similar to Fig. 1) and positive biases (Fig. 9d). Still, the histograms representing the urban and combined grassland and open shrubland categories (Fig. 19, top row) also reveal better model behavior at relatively higher wind speeds compared to HRRRV3 (Fig. 14). For convenience, we have combined the cropland and mosaic classes in Fig. 19c, despite their differences, and note that the forested and woody savanna grouping remained the most poorly handled (Fig. 19d).

In the end, and despite the improvements in model performance, we see that the glaringly different distributional shapes noted previously are still present and that this is still driven by the 6 PM to 6 AM period (Fig. 20). Clearly, more work on the stable boundary layer remains to be done. Although 10-m wind speeds during this period are typically not strong, sizable wind errors may have implications for boundary layer pollution transport, wind energy, etc..

5. Summary and recommendations

Our previous study, Fovell and Gallagher (2020, FG20), presented a detailed verification of Version 3 of the HRRR model focusing on surface and boundary layer winds and temperatures. It

[9]Station KDGP, which was an outlier in the April 2019 analysis, did not have sufficient April 2021 observations for inclusion.



532 FIG. 19. Similar to Fig. 14, but for April 2021 and referencing primary assignments from the higher resolution
533 MODIS landuse database used by HRRRV4.

542 was motivated by prior findings of systemic biases in forecast wind speeds at individual locations
543 even when network-average bias was insignificant (Cao and Fovell 2016; Fovell and Cao 2017;
544 Cao and Fovell 2018; Fovell and Gallagher 2018). FG20 leveraged underutilized observations (1-
545 min ASOS and high-frequency radiosonde) to investigate pervasive background biases across the
546 entirety of the CONUS in the operational HRRR model. The conclusions of FG20 were consistent
547 with previous work, detailing a pervasive bias in forecasts of surface sustained wind speed that
548 was highly (negatively) correlated with the observed value itself. Stations having lower average
549 wind speeds were being overpredicted while the wind threat was being underestimated at windy
550 locations.

551 The present examination represented a deeper analysis into the nature and cause of these biases
552 and also covered the now current version, HRRRV4. Analysis enhancements included verification
553 against hourly mean winds, consideration of local time and landuse classification, inspection of the

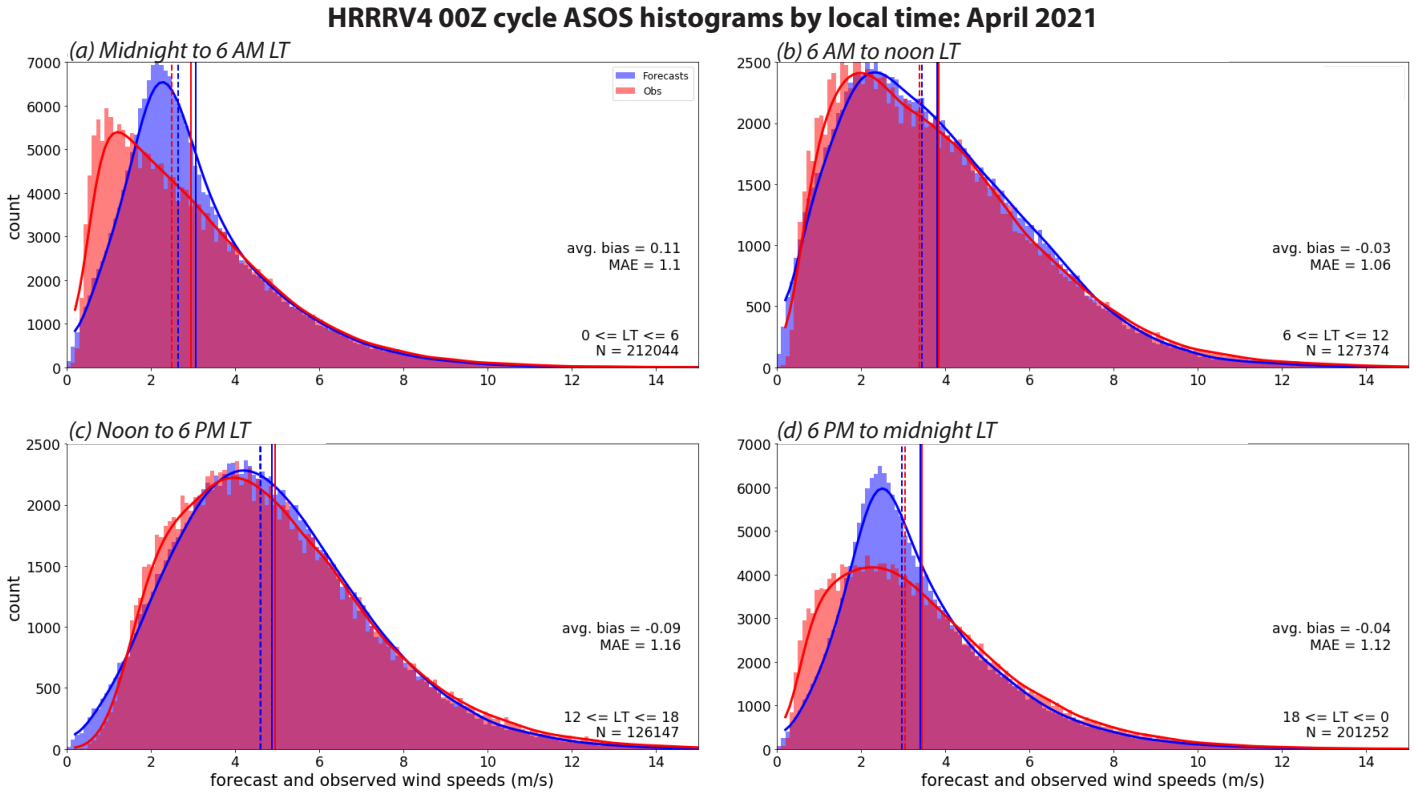


FIG. 20. Similar to Fig. 16, but for April 2021.

temporal and spatial variability of forecast and observed winds and biases, and the incorporation of additional surface observations from the New York State Mesonet (NYSM). Additionally, hourly maximum gusts were assessed and verified, using the network-average gust factor (GF) approach as proposed in Cao and Fovell (2018, CF18). Since GF was also correlated with bias, with smaller and larger factors associated with under- and overprediction, respectively, multiplying the biased wind forecasts by a fixed value (the network average) was found to reduce the bias in the gust predictions compared to those of the sustained winds.

For two spring months in 2019 and 2021, we showed the network average sustained wind forecasts for ASOS stations were excellent in Version 3 and even better in the current configuration. That said, the negative correlation between bias and mean observed wind speed persisted in Version 4, and we also demonstrated that the forecast and wind distributions were distinctly different overall, with ASOS forecasts in both versions having less spread about their modal value of about 2.5 m s^{-1} than in reality. Furthermore, observations associated with below-median GFs skewed towards higher speeds and those with above-median values skewed sharply leftward, characteristics not

captured in the forecasts. The inclusion of stations classified as forested land in the model actually worked to obscure the model's tendency to *underpredict* winds across the bulk of the ASOS network. A large fraction of the NYSM sites are situated in forested areas and that explained why the wind speeds at those stations were substantially overpredicted in the model.

Regarding local time, forecast wind distributions during the daytime looked quite good but less so at night, when the boundary and surface layer are usually stable. This demonstrates that further work needs to be done in the nocturnal regime. Even that systemic bias was landscape-dependent, however. Especially in urban and grassland areas, stronger winds at night were more common in the observations than in the model forecasts.

Taken together, we see evidence of further improvement in the HRRRV4 relative to its already skillful predecessor, at least in the spring month selected for close analysis. The gust parameterization inspired by Cao and Fovell (2018) continued to work well, despite its simplicity. Because it helped mitigate systemic biases, the CF18 gust can supply a starting point for a more sophisticated approach that might also factor in boundary layer depth, winds, and stability for even better-verifying predictions, especially in particularly challenging or dangerous situations (e.g., downslope windstorms, tropical cyclones, convective storms, etc.). Challenges with respect to the stable boundary layer and the treatment of some landuse classes (especially forested areas) remain. Other important variables, such as temperature, moisture, and the HRRR's own gust potential, have not yet been assessed. These should be foci of future work.

Acknowledgments. This research depended on support from National Science Foundation grants 1450195 and 1921546, the USD/R&E (The Under Secretary of Defense-Research and Engineering), National Defense Education Program (NDEP) / BA-1, Basic Research, SMART Scholarship Program, and Atmospheric Data Solutions LLC.

Data availability statement. HRRR model outputs can be obtained from Amazon Web Services and Google Cloud as described on the NOAA Big Data Program web page, https://www.noaa.gov/organization/information-technology/big-data-program. Versions 3 and 4 of the HRRR are based on the WRF-ARW model available via https://www2.mmm.ucar.edu/wrf/users/. HRRR model landuse information was extracted from the WRF model system Geogrid files made available at https://rapidrefresh.noaa.gov/hrrr/. One-minute ASOS data are available at https://www.ncdc.noaa.gov/data-access/land-based-station-data/. New York State Mesonet data are not publicly available and were used with permission. Mesonet data may be requested at http://www.nysmesonet.org/weather/requestdata.

APPENDIX

In Figs. 16 and 20, we demonstrated that the 10-m wind speed distributions for forecasts and observations were less comparable at night for both versions of the HRRR examined. A reviewer pointed out that our study design incorporated more nighttime than daytime hours and wondered how that influenced the results. We thank the reviewer for alerting us to this. We performed our analyses again, limiting them to forecast hours 0-24, inclusive, which makes the number of night and day hours much more comparable. However, we found few discernible changes to the figures, with the caveat noted below, and there was no impact on our conclusions. As a consequence, we have retained all forecast hours (0-36) common to both HRRR versions for the 00 UTC cycle.

The reason for the insensitivity is that many of our analyses involved medians, means, and differences between means (i.e., biases). Even at night, the means and medians of the forecasts and observations were very similar, as were biases and mean absolute errors (Figs. 16 and 20), and that is why removing some of the nocturnal hours did not materially alter the results. The *distributional* differences at night, however, imply larger scatter among forecast/observation pairs. Figure A1 is a version of Fig. 10 in which only pairs for forecast hours 0-24 were retained. Reflecting the reduced

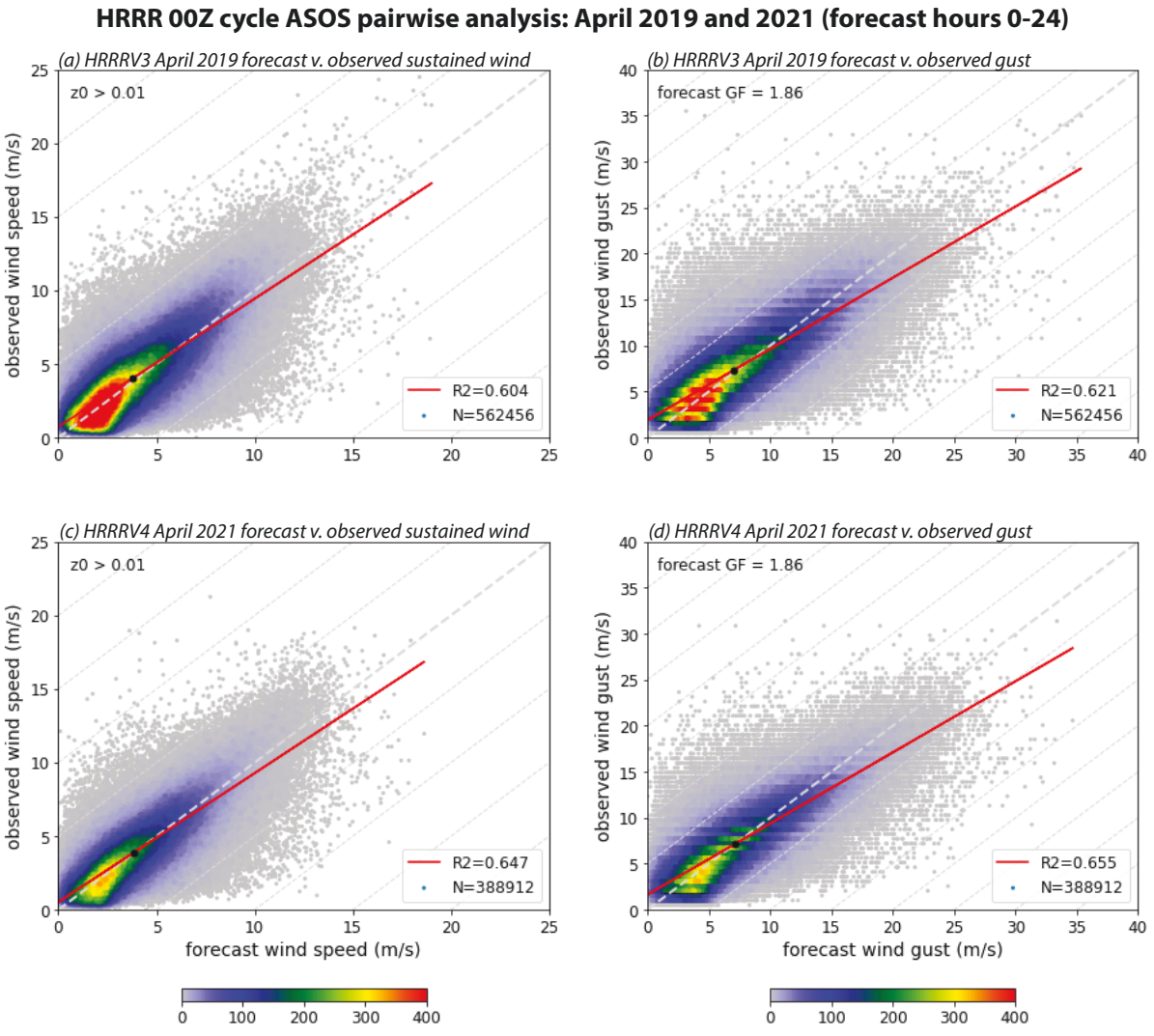


FIG. A1. As in Fig. 10 but for analyses restricted to forecast hours 0-24, inclusive.

scatter, the R^2 values are somewhat higher in this version, but the conclusions from our analyses remain unchanged.

Another reviewer wondered about the large change in gust factor (GF), from 1.29 to 1.86, that occurred when we shifted from top-of-hour to hourly maximum gusts. First, we note that our GFs are usually computed as ratios-of-means, such that a station's mean gust is being divided by its average sustained wind. The network-averaged GF then represents the ratio of the average of the gusts and sustained winds over all included stations. The mean-of-ratios approach is also

valid but typically results in higher gust factors because wind and gust distributions have long tails (cf. Fig. 11 and Gallagher 2021).

In previous work (e.g., Fovell and Gallagher 2018), we used ASOS reports from the 1-min database, each of which consisted of a 2-min running average wind (i.e., sustained wind) and the peak 3-sec average (gust) during that one minute interval. Over the ASOS network, the gust factor for the 1-min reports averaged to about 1.29. For this study, we adopted the hourly maximum gust as a better measure of the wind threat. This GF is an hour's fastest 3-s gust report divided by that hour's mean sustained wind, so both the numerator and denominator of the gust factor have been redefined. However, consistent with Harper et al. (2010), the mean wind is nearly the same when averaged over 2- and 60-min periods. Yet, the largest gust discoverable within a given interval logically increases with interval length.

Figure A2 presents the ratio-of-means GFs obtained from about 840 ASOS sites vs. the time interval for which the maximum 3-s gust was identified. For each station, for each of four months considered, the station's entire record length T was subdivided into nonoverlapping segments of length τ in minutes, where $1 \leq \tau \leq 60$. Then, for each segment without missing data, the maximum gust report was identified and the mean sustained wind was computed. These were first averaged over all available segments of length τ and then over all stations and the four months, yielding the ratio-of-means network-averaged GF representing time interval τ. Because the average sustained wind for each interval represented the same information, only the numerator of the GF varied among the time intervals. Figure A2 demonstrates that the 1-min GF is about 1.29 (red star) while the 60-min value is about 1.84 (green star), about 1.4 times larger. This curve varies somewhat among seasons and more prominently among networks owing to differences in mean wind speeds, mounting heights, anemometer hardware, characteristic exposures, and possibly other factors, but the shape of the curve is typically logarithmic in time.

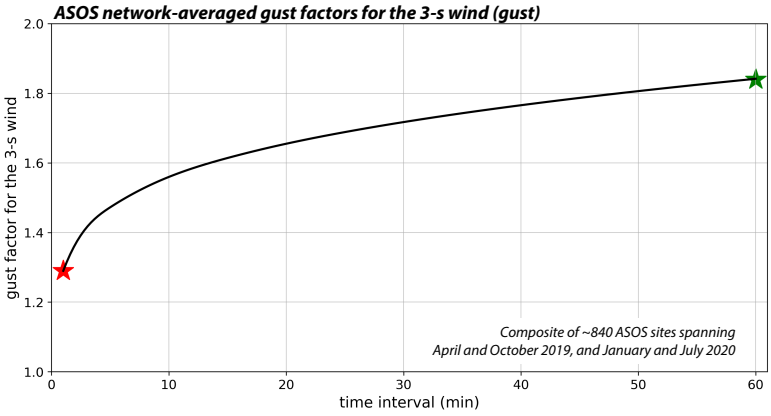


FIG. A2. Gust factor curve for the ASOS network representing a composite of about 840 stations sampling four seasons. For each time interval considered, the network-average maximum 3-s wind (gust) was divided by the network-average sustained wind representing that interval. Red and blue stars represent the one-minute GF used in Fovell and Gallagher (2018) and the 60-min GF used in this study, respectively.

References

Arya, S., 1999: *Air Pollution Meteorology and Dispersion*. Oxford University Press, 305 pp.

Benjamin, S. G., E. P. James, J. M. Brown, E. J. Szoke, J. S. Kenyon, R. Ahmadov, and D. D. Turner, 2021: Diagnostic fields developed for hourly updated NOAA weather models. Tech. rep., Earth System Research Laboratory, Global Systems Laboratory, NOAA Tech. Doc. OAR GSL-66, 55 pp. https://doi.org/10.25923/f7b4-rx42, URL https://repository.library.noaa.gov/view/noaa/32904.

Benjamin, S. G., 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.

Brasseur, O., 2001: Development and application of a physical approach to estimating wind gusts. *Mon. Wea. Rev.*, **129**, 5–25, https://doi.org/10.1175/1520-0493(2001)129$<$0005:DAAOAP$>$2.0.CO;2.

Brotzge, J. A., and Coauthors, 2020: A technical overview of the New York State Mesonet standard network. *J. Atmos. Oceanic Technol.*, **37 (10)**, 1827 – 1845, https://doi.org/10.1175/JTECH-D-19-0220.1, URL https://journals.ametsoc.org/view/journals/atot/37/10/jtech-d-19-0220.1.xml.

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.

Cao, Y., and R. G. Fovell, 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.

Dowell, D. C., and co authors, 2022: The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part 1: Motivation and System Description. *Wea. Forecasting*, **XXX**, YYY–ZZZ.

Durst, C. S., 1960: Wind speeds over short periods of time. *Meteor. Mag.*, **89**, 181 – 186.

Fovell, R. G., and Y. Cao, 2017: The Santa Ana winds of Southern California: Winds, gusts, and the 2007 Witch fire. *Wind Struct.*, **24 (6)**, 529–564, https://doi.org/10.12989/was.2017.24.6.529.

Fovell, R. G., and A. Gallagher, 2018: Winds and gusts during the Thomas fire. *Fire*, **1 (3)**, https://doi.org/10.3390/fire1030047.

Fovell, R. G., and A. Gallagher, 2020: Boundary layer and surface verification of the High-Resolution Rapid Refresh, version 3. *Wea. Forecasting*, **35 (6)**, 2255–2278, https://doi.org/10.1175/WAF-D-20-0101.1, URL https://journals.ametsoc.org/view/journals/wefo/35/6/waf-d-20-0101.1.xml.

G. Lufft Mess und Regeltechnik GmbH, 2021: *Technical Data V200A Ultrasonic Wind Sensor*. Fellbach, Germany, URL https://www.lufft.com/products/wind-sensors-anemometers-289/v200a-ultrasonic-wind-sensor-2295/productAction/outputAsPdf/.

Gallagher, A. R., 2021: Exploring Environmental and Methodological Sensitivities of Forecasted and Observed Surface Winds and Gusts Using Underutilized Datasets. Ph.D. thesis, University at Albany, State University of New York, 260 pp.

Gray, M. E. B., 2003: The use of a cloud resolving model in the development and evaluation of a probabilistic forecasting algorithm for convective gusts. *Meteor. Appl.*, **10**, 239–252, https://doi.org/10.1017/S1350482703003049.

Gutiérrez, A., and R. G. Fovell, 2018: A new gust parameterization for weather prediction models. *J. Wind Eng. Ind. Aerodyn.*, **177**, 45–59, https://doi.org/10.1016/j.jweia.2018.04.005.

Harper, B., J. D. Kepert, and J. D. Ginger, 2010: Guidelines for converting between various wind averaging periods in tropical cyclone conditions. Tech. rep., World Meteorological Organization Tech. Doc. WMO/TD-1555, 64 pp. URL https://www.wmo.int/pages/prog/www/tcp/documents/WMO_TD_1555_en.pdf.

He, S., T. G. Smirnova, and S. G. Benjamin, 2021: Single-Column Validation of a Snow Subgrid Parameterization in the Rapid Update Cycle Land-Surface Model (RUC LSM). *Water Resources Res.*, **57**, e2021WR029 955, https://doi.org/10.1029/2021WR029955.

Holmes, J. D., A. C. Allsop, and J. D. Ginger, 2014: Gust durations, gust factors and gust response factors in wind codes and standards. *Wind Struct.*, **19 (3)**, 339–352, https://doi.org/10.12989/WAS.2014.19.3.339.

Medeiros, L. E., and D. R. Fitzjarrald, 2014: Stable boundary layer in complex terrain. Part I: Linking fluxes and intermittency to an average stability index. *J. Appl. Meteor. Climatol.*, **53 (9)**, 2196 – 2215, https://doi.org/10.1175/JAMC-D-13-0345.1.

Medeiros, L. E., and D. R. Fitzjarrald, 2015: Stable boundary layer in complex terrain. Part II: Geometrical and sheltering effects on mixing. *J. Appl. Meteor. Climatol.*, **54 (1)**, 170 – 188, https://doi.org/10.1175/JAMC-D-13-0346.1.

Nakamura, K., R. Kershaw, and N. Gait, 1996: Prediction of near-surface gusts generated by deep convection. *Meteor. Appl.*, **3**, 157–167, https://doi.org/10.1002/met.5060030206.

National Wildfire Coordinating Group, 2019: NCWG Standards for Fire Weather Stations (PMS 426-3). National Wildfire Coordinating Group, URL https://www.nwcg.gov/sites/default/files/publications/pms426-3.pdf, 46 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 Technical Memorandum, National Oceanic and Atmospheric Administration, Office of Oceanic and Atmospheric Research, 42 pp., https://doi.org/10.25923/n9wm-be49.

Olson, J. B., 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.

Panofsky, H. A., H. Tennekes, D. H. Lenschow, and J. C. Wyngaard, 1977: The characteristics of turbulent velocity components in the surface layer under convective conditions. *Bound.-Lay. Meteorol.*, **11**, 355–361, https://doi.org/10.1007/BF02186086.

Petersen, E., N. Mortensen, L. Landberg, J. Højstrup, and H. Frank, 1998: Wind power meteorology. Part 1: Climate and turbulence. *Wind Energy*, **1**, 2–22.

729 Piccardo, G., and G. Solari, 1998: Closed form prediction of 3-D wind-excited response of slender
730 structures. *J. Wind Eng. Ind. Aerodyn.*, **74-76**, 697–708, https://doi.org/10.1016/S0167-6105(98)
731 00063-4.

732 Pichugina, Y. L., and Coauthors, 2019: Spatial variability of winds and HRRR–NCEP model error
733 statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia River
734 Basin. *J. Appl. Meteor. Climatol.*, **58**, 1633–1656, https://doi.org/10.1175/JAMC-D-18-0244.1.

735 R.M. Young Company, 2000: *Meteorological Instruments Instructions Wind Monitor-HD*
736 *Model 05108*. Traverse City, MI, USA, URL https://s.campbellsci.com/documents/ca/manuals/
737 05108-10_man.pdf.

738 Rolinski, T., S. B. Capps, R. G. Fovell, Y. Cao, B. J. D'Agostino, and S. Vanderburg, 2016: The
739 Santa Ana wildfire threat index: Methodology and operational implementation. *Wea. Forecast-*
740 *ing*, **31**, 1881–1897, https://doi.org/10.1175/WAF-D-15-0141.1.

741 Rolinski, T., S. B. Capps, and W. Zhuang, 2019: Santa Ana winds: A descriptive climatology.
742 *Wea. Forecasting*, **34 (2)**, 257 – 275, https://doi.org/10.1175/WAF-D-18-0160.1, URL https:
743 //journals.ametsoc.org/view/journals/wefo/34/2/waf-d-18-0160_1.xml.

744 Sheridan, P., 2011: Review of techniques and research for gust forecast-
745 ing and parameterisation forecasting. Tech. rep., UK Meteorological Of-
746 fice Research Technical Report 570., 64 pp. URL https://www.researchgate.
747 net/profile/Peter-Sheridan-2/publication/268744498_Review_of_techniques_and_
748 research_for_gust_forecasting_and_parameterisation/links/5474c0b00cf245eb436e0791/
749 Review-of-techniques-and-research-for-gust-forecasting-and-parameterisation.pdf.

750 Skamarock, W. C., and Coauthors, 2019: A description of the Advanced Research WRF version
751 4. Tech. rep., NCAR Technical Note NCAR/TN–556+STR, Boulder, CO, 145 pp. URL https:
752 //www2.mmm.ucar.edu/wrf/users/docs/technote/v4_technote.pdf.

753 Stucki, P., S. Dierer, C. Welker, J. J. Gómez-Navarro, C. C. Raible, O. Martius, and S. Brönnimann,
754 2016: Evaluation of downscaled wind speeds and parameterised gusts for recent and historical
755 windstorms in Switzerland. *Tellus A: Dynamic Meteorology and Oceanography*, **68 (1)**, 31 820,
756 https://doi.org/10.3402/tellusa.v68.31820.

Westerling, A. L., D. R. Cayan, T. J. Brown, B. L. Hall, and L. G. Riddle, 2004: Climate, Santa Ana winds and autumn wildfires in Southern California. *Eos, Trans. Amer. Geophys. Union*, **85**, 289–296, https://doi.org/10.1029/2004EO310001.

Wilczak, J. M., and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): Observational field campaign. *Bull. Amer. Meteor. Soc.*, **100 (9)**, 1701–1723, https://doi.org/10.1175/BAMS-D-18-0035.1.

WMO, 2018: Guide to Meteorological Instruments and Methods of Observation. WMO no. 8, Geneva, Switzerland. World Meteorological Organization, URL https://library.wmo.int/index.php?lvl=notice_display&id=12407#.YWR0VC-B17h, 1177 pp.