Forecasting High Wind Events in the HRRR Model over Wyoming and Colorado. Part I: Evaluation of Wind Speeds and Gusts

ETHAN COLLINS,[a] ZACHARY J. LEBO,[b] ROBERT COX,[c] CHRISTOPHER HAMMER,[c] MATTHEW BROTHERS,[c] BART GEERTS,[a] ROBERT CAPELLA,[d] AND SARAH MCCORKLE[e]

[a] *Department of Atmospheric Science, University of Wyoming, Laramie, Wyoming*
[b] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[c] *National Weather Service, Cheyenne, Wyoming*
[d] *SciTec, Inc., Boulder, Colorado*
[e] *National Weather Service, Monterey, California*

(Manuscript received 27 February 2023, in final form 9 February 2024, accepted 12 February 2024)

ABSTRACT: Strong wind events cause significant societal damage ranging from loss of property and disruption of commerce to loss of life. Over portions of the United States, the strongest winds occur in the cold season and may be driven by interactions with the terrain (downslope winds, gap flow, and mountain wave activity). In the first part of this two-part series, we evaluate the High-Resolution Rapid Refresh (HRRR) model wind speed and gust forecasts for the 2016–22 winter months over Wyoming and Colorado, an area prone to downslope windstorms and gap flows due to its complex topography. The HRRR model exhibits a positive bias for low wind speeds/gusts and a large negative bias for strong wind speeds/gusts. In general, the model misses many strong wind events, but when it does predict strong winds, there is a high false alarm probability. An analysis of proxies for surface winds is conducted. Specifically, 700- and 850-mb (1 mb = 1 hPa) geopotential height gradients are found to be good proxies for strong wind speeds and gusts at two wind-prone locations in Wyoming. Given the good agreement between low-level height gradients and surface wind speeds yet a strong negative bias for strong wind speeds and gusts, there is a potential shortcoming in the boundary layer physics in the HRRR model with regard to predicting strong winds over complex terrain, which is the focus of the second part of this two-part study. Last, the sites with the largest strong wind speed bias are found to mostly sit on the leeward side of high mountains, suggesting that the HRRR model performs poorly in the prediction of downslope windstorms.

SIGNIFICANCE STATEMENT: We investigate the performance of the High-Resolution Rapid Refresh (HRRR) model with respect to strong wintertime wind speeds and gusts over the complex terrain of Wyoming and Colorado. We show that the overall performance of the HRRR model is low with regard to strong wind speed and wind gust forecasts across the investigated winter seasons, with a large negative bias in predicted strong wind speeds and gusts and a small positive bias for weak wind speeds and gusts. The largest biases are found to be on the leeward side of high mountains, indicating poor prediction of downslope winds. This study also utilizes National Weather Service forecasting metrics to understand their performance with respect to strong wind forecasts, and we find that they provide skill in forecasting these events.

KEYWORDS: Wind; Wind gusts; Forecast verification/skill; Numerical weather prediction/forecasting

1. Introduction

While the fundamental forces that drive winds in Earth's atmosphere have been understood for decades, the accurate prediction of winds, in particular strong winds, has been a challenge for the numerical weather prediction (NWP) community. Such strong winds occur within a variety of different weather systems, including but not limited to tropical cyclones (TCs; e.g., Emanuel 2000), thunderstorms (e.g., Ashley 2007; Ashley and Mote 2005), midlatitude cyclones (e.g., Collier et al. 1994), and even topographically forced flows, e.g., downslope winds and gap flows, such as those commonly referred to as Santa Ana winds, chinook winds, and sundowner winds (e.g., Klemp and Lilly 1975; Cao and Fovell 2018; Carvalho et al. 2020). In most parts of the continental United States (CONUS), winds associated with the first three categories are often the most severe. The predictability of strong winds associated with these events has been greatly improved in recent years through a combination of model improvements and advanced observational capabilities (e.g., DeMaria et al. 2014; McGovern et al. 2014). In general, winds associated with convective storms have received considerable attention in the literature (e.g., Taszarek et al. 2020). On the contrary, the prediction of strong to severe winds associated with non-thunderstorm events (Pokharel et al. 2017) have received less attention.

In some parts of the CONUS, clear-sky non-"weather"-related winds can and are often stronger than those associated with thunderstorms, midlatitude cyclones, and TCs. In particular, Martner and Marwitz (1982) provided an in depth look into the climatology of winds over southeastern (SE) Wyoming, focusing on wind energy and wind speeds and directions both seasonally and diurnally. They found that wintertime wind speeds across SE Wyoming are ≈50% stronger than in

Corresponding author: Zachary J. Lebo, zachary.lebo@ou.edu

DOI: 10.1175/WAF-D-23-0036.1

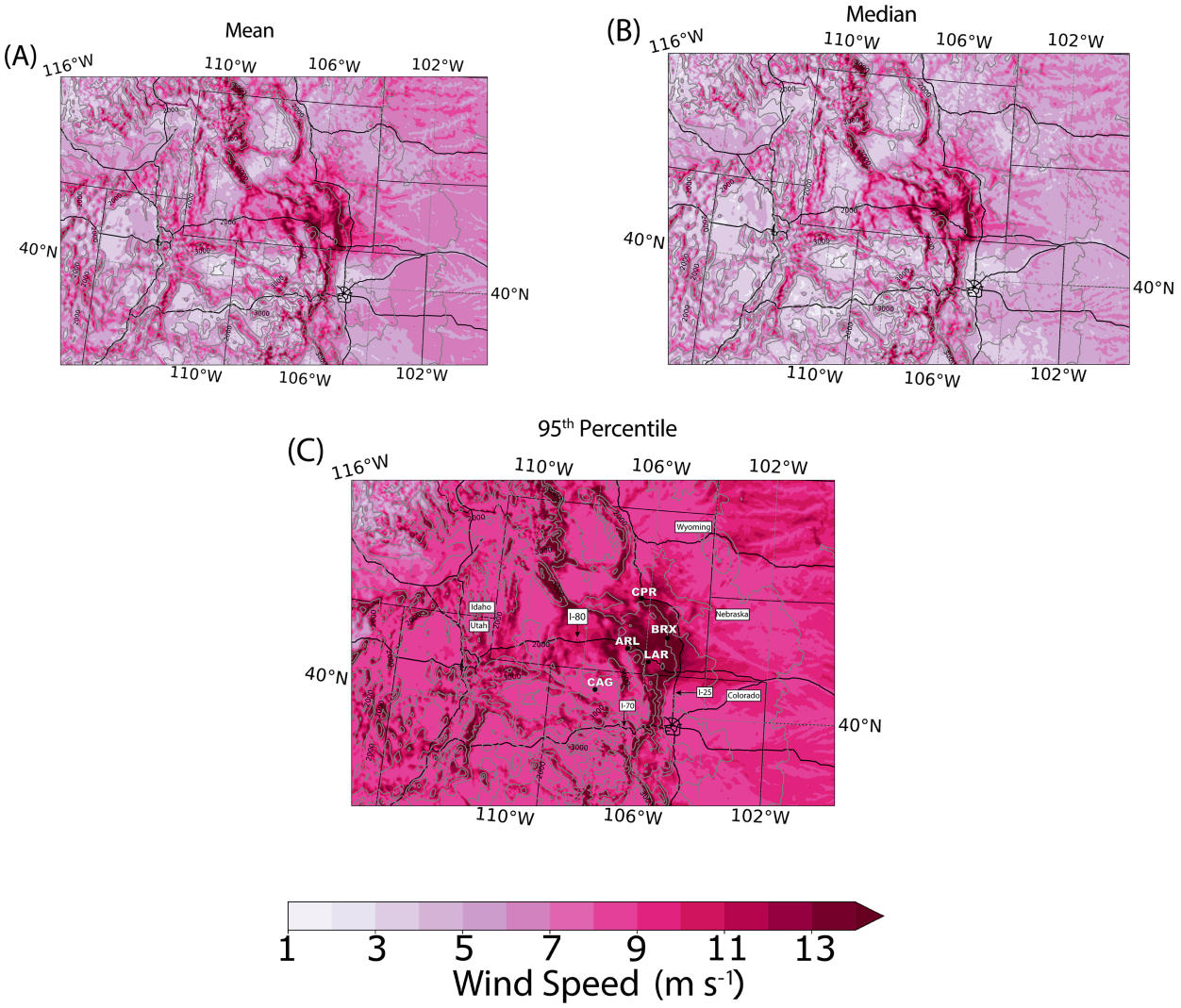

FIG. 1. (a) Mean December–January–February (DJF) wind speeds over Wyoming based on the 2016–22 HRRR forecast hour 0 output; gray lines indicate the elevation. (b) As in (a), but for the median wind speed. (c) As in (a), but for the 95th percentile wind speed. Major interstates are overlaid and labeled accordingly. Laramie, Arlington, Bordeaux, and Casper in Wyoming and Craig in Colorado are labeled as LAR, ARL, BRX, CPR, and CAG, respectively. Note that the color bar in (c) differs owing to the different magnitudes.

the summer months. To demonstrate just how strong the wind can be over the Mountain West of the United States, Fig. 1 shows the mean, median, and 95th percentile wind speeds spanning December–February based on the High-Resolution Rapid Refresh (HRRR; Benjamin et al. 2016) model forecast hour 0 output from 2016 to 2022.[1] Many regions, especially over the higher terrain, experience mean wind speeds in excess of 9 m s^{-1} (approximately 20 mph) according to the HRRR. Moreover, especially across the higher terrain, median wind speeds are similar in magnitude to the mean, while the 95th percentile wind speeds in the most wind-prone locations exceed 14 m s^{-1} (more than 30 mph). Motivation for this study lies in the fact that these areas are collocated with major interstates, including Interstate 70 (I-70), I-80, I-25, and I-15, as well as the Denver International Airport, thus presenting a serious hazard to both road and air transportation.

The Front Range of Colorado (spanning from Fort Collins to Colorado Springs and farther south) also frequently experiences strong wind events. These areas experience strong downslope windstorms (Klemp and Lilly 1975; Abatzoglou et al. 2021). Downslope windstorms are typically formed due to synoptic conditions that interact with high terrain and are formed via two different mechanisms on the synoptic scale (Mercer et al. 2008). The first mechanism leads to strong downslope winds occurring ahead of a fast-moving shortwave trough, often accompanied by a surface lee trough forming

[1] We use HRRR-simulated winds here as opposed to observations to better convey the spatial heterogeneity in wind speeds, which cannot be fully appreciated with the observation network over the complex terrain. As shown below, these winds are likely an underestimate of the actual strong wind speeds over the region.

across the High Plains. The second mechanism of downslope windstorm formation is associated with cold air advection behind a surface cold front that has moved across the region. The accepted theory (Mercer et al. 2008) is that the overturning of a large-amplitude vertically turning gravity wave leads to the formation of downslope wind events. This can be triggered by flow over a mountain range (i.e., the Rockies). Gap winds, on the other hand, are formed by the acceleration of air through lower terrain or a canyon in a mountain range.

These terrain-induced strong wind events have been known to cause property damage and fuel the rapid spread of grass fires, as recently illustrated in the rapid expansion of the Marshall Fire in December 2021. Fovell et al. (2022) recently showed that forecast models, even with a day lead time, failed to predict a strong downslope windstorm in this case, which they attributed to errors on the synoptic scale (e.g., errors in the jet stream and trough locations). Furthermore, Bergen and Murphy (1978) found that improved short-term forecasting of windstorms in Boulder, Colorado, could substantially reduce property and business losses. Understanding model performance with regard to strong winds over these regions could allow for better warning and decreased societal/economic impacts.

To improve predictions of strong wind events, one must first understand why models struggle to accurately predict them. Kunz et al. (2010) used a regional numerical weather prediction (NWP) model and found that the grid spacing is too coarse to represent strong wind events, and higher-resolution models are needed to more accurately predict strong wind events. For example, the Global Forecasting System (GFS) has a nominal grid spacing of 13 km, with most data products available publicly at even coarser resolutions, resulting in a smoothed and coarsened representation of terrain. This could result in errors simulating strong wind events, similar to the errors in NWP forecasts of precipitation, especially over high and complex terrain (e.g., He et al. 2019; Rahimi et al. 2022; Wyngaard 2004; Gutiérrez and Fovell 2018). More specifically, the inaccurate and simplified/smoothed terrain over Wyoming and Colorado could also impact the forecasting of gap flows (e.g., Brennan et al. 2010) and downslope windstorms (e.g., Oltmanns et al. 2015), both of which are influenced by the terrain.

Unlike the GFS and regional models commonly used for NWP in the United States, e.g., the North American Mesoscale Model (which has a grid spacing of 12 km, with a nest of 3 km), the HRRR model has a high horizontal resolution (3 km) and a higher temporal frequency of new model runs (hourly as opposed to every 6 h). Although the grid spacing used in the HRRR is more capable of more accurately representing the complex terrain of Wyoming and other Rocky Mountain states, the accuracy of HRRR wind speed forecasts have not been specifically analyzed over the Wyoming and Rocky Mountain region. While wind speeds have not been evaluated over this region, there have been studies conducted to evaluate the accuracy and performance of HRRR wind forecasts over other regions. For example, a study conducted by Cao and Fovell (2016) looked into the ability to simulate Santa Ana winds in San Diego County, California. This study found that the Weather Research and Forecasting (WRF) Model (which is the basis for the HRRR) was able to capture details of the selected Santa Ana events, although the wind speed was generally overestimated. Cao and Fovell (2018) showed that this overestimate could be attributed to model physics, including the planetary boundary layer (PBL) scheme, surface layer (SL) scheme, and land surface model (LSM). Moreover, Zhong and Whiteman (2008) also found that there was sensitivity to the model parameterizations over the Salt Lake Valley using the Regional Atmospheric Modeling System (RAMS; Pielke et al. 1992) at grid spacings of 250 and 1000 m. They found that the LSM had a significant impact due to the differing representation of surface roughness. More recently, Fovell and Gallagher (2022) studied the performance of HRRR wind speeds and wind gusts (wind gusts were derived from a gust factor rather than the HRRR-forecasted wind gust which our study looks at). They found that strong wind speeds and wind gusts are underforecasted by the HRRR compared to Automated Surface Observing Systems (ASOS) and the New York State Mesonet. Moreover, Fovell and Gallagher (2020) showed that biases in the HRRR-predicted winds extend through the boundary layer. While these studies are important, they did not focus on the Rocky Mountain region and/or did not evaluate the wintertime climatological performance of an operational model with the highest horizontal resolution (in the United States), i.e., the 3-km HRRR or NAM nest models.

The above review identified a gap in our understanding and prediction of strong to severe wind events across the complex terrain of the Rocky Mountains of the United States in NWP models (which are primarily related to gap flows and downslope windstorms). These events have important societal and economic risks, and thus it is important to both understand how well state-of-the-art NWP models are at predicting high wind events and analyze the physical reasons why or why not. This paper (Part I) is the first in a two-part series targeted at addressing these objectives and is mainly focused on an evaluation of HRRR-forecasted winds over multiple winter seasons in the Rocky Mountains of the United States, a region conducive to gap flows and downslope windstorms, especially in Wyoming and Colorado. Collins et al. (2024, hereafter Part II) describes a suite of sensitivity simulations targeted at understanding the shortcomings of HRRR-forecasted winds and to identify a more applicable model physics configuration for this region.

The remainder of this paper is organized as follows. Section 2 describes the observations and HRRR model data used in the analysis and the methodology. Section 3 describes the performance of the HRRR wind speed and wind gust forecasts, and a discussion of other model fields related to strong low-level winds is also presented. HRRR model performance as a function of forecast hour is presented in section 4. Furthermore, in section 5, an analysis of the strong wind speed biases in the HRRR model in relation to the complex terrain in the study domain is presented. This paper concludes with section 6, which provides the conclusions and discusses the motivation for the subsequent work presented in Part II.

2. Data and methodology

a. *HRRR overview*

The HRRR model is an operational NWP model with an hourly update cycle, initialized utilizing the Rapid Refresh

(RAP) model, and a horizontal grid spacing of 3 km, making it currently one the highest-resolution NWP model used operationally in the United States (the other being the NAM Nest, also at 3 km). Key advantages of this model are that its high horizontal resolution allows for convection in the model without a cumulus parameterization as well as a better representation of terrain (both the heterogeneity of the terrain height and slope), which has large impacts over Wyoming and Colorado owing to the complex terrain and the role the terrain plays in gap flow and downslope wind events, as described in detail in section 1. In theory, this high resolution compared to other operational NWP models is expected to improve the prediction of terrain-driven strong wind events.

The HRRR model is initialized at hourly intervals; however, when initialized at 0000, 0600, 1200, and 1800 UTC, it is run for a longer duration (48 h beginning with version 4) compared with the other times (18 h). For this study, we analyzed the 0000, 0600, 1200, and 1800 UTC HRRR forecasts. While we analyze many of the forecast hours, the majority of the evaluation presented here is conducted utilizing forecast hour 6 as the basis for our analysis. This forecast hour is utilized as it would provide adequate lead time for a forecaster to warn strong wind events.

The time frame for this study spans December, January, and February from 2016 to 2022 (HRRR versions 2–4; version 1 is not utilized as versions 2–4 extend to 18 h, while version 1 extends to 15 h (see Table 4 in Dowell et al. 2022). These months are chosen to coincide with the peak in occurrence of strong winds over the region (see Fig. 6 in Abatzoglou et al. 2021) as well as the likelihood of such strong winds coinciding with snow on the ground, thus creating potentially hazardous blowing snow conditions.

Two different wind fields are evaluated in this study from the HRRR model: the 10-m wind speed and wind gust (which is intended to be a gust potential and often should exceed observed wind gusts). The HRRR wind gust potential is postprocessed by the Earth System Research Laboratory (ESRL) utilizing the following equation:

$$v_{\text{gust}} = v_{\text{sfc}} + \max\{f(z)[v(k) - v_{\text{sfc}}]\}. \quad (1)$$

To compute the wind gust potential (v_{gust}; Benjamin et al. 2020), the depth of the well-mixed PBL is calculated. This is done by identifying the altitude where the virtual potential temperature equals or exceeds the 2-m surface virtual potential temperature. Once this depth is computed, at each model level k in the PBL, the wind speed $v(k)$ is used to compute the excess wind speed relative to the (10-m) surface speed v_{sfc}. This excess is multiplied by a coefficient that is a function of height [$f(z)$, where z refers to AGL height], which decreases linearly from 1 at the surface to 0.5 at 1 km above ground level (AGL). For all heights higher than 1 km AGL, 0.5 is used. Then the maximum of this weighted wind excess is added to the 10-m wind to obtain the wind gust potential. A direct comparison of the wind gust potential to observed wind speeds at any given time is not consistent. Therefore, to reflect the fact that the HRRR wind gusts are intended to represent the "potential" gusts, we compare the potential wind gusts at the HRRR model output times to the maximum observed wind gusts over a 1-h period surrounding the model forecast time.

Prior to 2018, the 10-m wind speed was assumed to be equivalent to the wind speed at the lowest model level [approximately 7 m above ground level (AGL) at sea level]. Beginning in 2018, the HRRR now uses the 10-m wind output from the MYNN surface layer scheme, which utilizes a neutral-log relationship to derive the zonal and meridional components of the wind at 10 m from the lowest-model-level wind field (Nakanishi and Niino 2006). Our focus on multiple years will allow us to determine if this change in the diagnoses of the 10-m wind speed has changed model performance.

b. MADIS overview

The wind speed and wind gust observations utilized in this study are from the National Oceanic and Atmospheric Administration (NOAA) Meteorological Assimilation Data Ingest System (MADIS) dataset (Miller et al. 2007). This dataset is composed of numerous different data networks, including but not limited to, the National Weather Service (NWS), ASOS, as well as transportation and other state and regional networks. This high-density network allows for an ample regional analysis and evaluation effort of the HRRR dataset over the target area, namely, Wyoming, Colorado, and portions of the surrounding states. Shown in Fig. 2 is a density plot of the 2446 sites utilized for the HRRR evaluation effort of this study. There is clear spatial heterogeneity in the distribution of the surface sites, with far more locations confined to metropolitan areas (e.g., Denver, Colorado, and Cheyenne, Wyoming) as well as interstates (e.g., the I-70 corridor in Colorado). In addition to an analysis of all sites in the database, which could provide biased results owing to the spatial distribution of sites, we also analyze individual sites known to have a high frequency of strong winds, namely, Laramie (LAR), Arlington (ARL), and Bordeaux (BRX), Wyoming, which are denoted in Fig. 1.

For the model–observation comparison, the HRRR model output is bilinearly interpolated to the locations of the observations. Using hourly HRRR output, we only consider observations that occur within ±5 min of the top of the hour. Owing to the differing heights of wind speeds in the HRRR compared to actual observations, HRRR wind speeds are adjusted from 10 to 6.1 m for all stations in which the anemometer height is 6.1 m (e.g., RAWS)[2] following the methodology of Cao and Fovell (2016), Eq. (2) therein, which uses an assumed logarithmic wind profile in the near-surface region to adjust heights. Furthermore, all wind speeds and wind gusts in the observations greater than 50 m s^{-1} are removed; while such wind speeds are possible in the study domain, given the locations of the observation sites (e.g., not on mountain tops),

[2] The MADIS dataset does not contain specific information on the height of the anemometers. While standards are in place for these sites, we cannot be assured that they are followed. For example, the MADIS dataset includes some WYDOT sites, which follow no standard to the best of the authors' knowledge. In such instances, without visual inspection of each site, we cannot be absolutely certain of the heights. In general, this will only quantitatively impact the results of this work, as anemometers are not expected to be *higher* than assumed in this study. If in fact the anemometers are lower than assumed, this would only further increase the negative bias in strong wind speeds shown below.

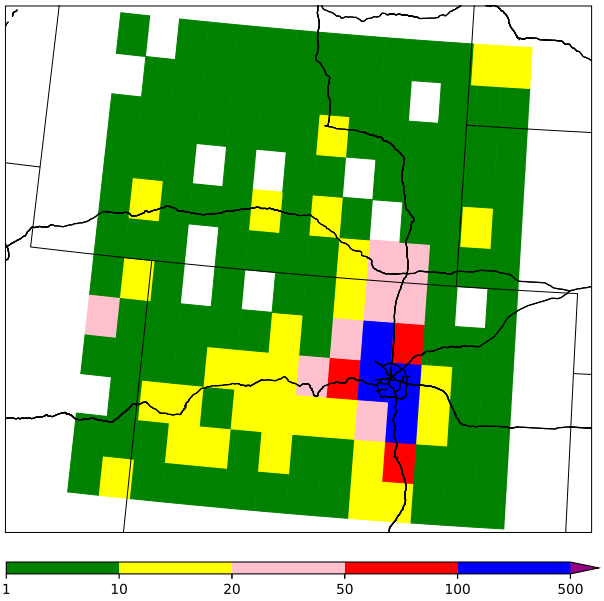

FIG. 2. Density plot of MADIS sites utilized for the HRRR analysis. The sites are binned into 0.5° × 0.5° grid boxes.

such data points are likely erroneous. It should be noted that ASOS METARs set winds less than 3 kt (1 kt ≈ 0.51 m s^{-1}) to 0. However, as the focus of this study is on strong wind events, this adjustment should be less of a concern in our study but certainly could enhance the weak wind biases.

c. *Verification statistics*

To measure the HRRR performance with regard to wind speeds and gusts, we compute the coefficient of determination (r^2) and bias between the HRRR-forecasted winds and observed winds. Moreover, several other forecast verification statistics are used, as adapted and modified from Wilks (2020). To compute these statistics, the observed and HRRR wind speeds are categorized into a contingency table in the format of Fig. 3. We focus on the strong winds for their hazardous nature and because, as will be shown below, if we focus on the weak wind speeds, our conclusions will differ as the model inherently performs reasonably well given how often they occur (the number of weak wind speed hits would overwhelm the results), although they are a nuisance in general. Four categories are created based on a threshold of 20 m s^{-1}. The 20 m s^{-1} (approximately 45 mph) filter is commonly used by the Wyoming Department of Transportation (WYDOT) and the NWS Cheyenne office (CYS) to begin warning severe wind events over this region. This threshold results in the following categories: weak wind speed (<20 m s^{-1}) hit, missed strong wind speed, false alarm strong wind speed, and strong wind speed (>20 m s^{-1}) hit. These four categories are then utilized to compute the hit rate, miss rate, false alarm rate, and false alarm ratio for strong winds, which are defined as follows:

$$\text{hit rate} = \frac{\text{strong wind speed hit}}{\text{strong wind speed hit} + \text{missed strong wind speed}}, \quad (2)$$

		Observations	
		Weak Wind Speed	Strong Wind Speed
Model	Forecasted Weak Wind Speed	Weak Wind Speed Hit	Missed Strong Wind Speed
	Forecasted Strong Wind Speed	False Alarm Strong Wind Speed	Strong Wind Speed Hit

FIG. 3. Example contingency table for the HRRR model evaluation of strong winds and wind gusts.

$$\text{miss rate} = \frac{\text{missed strong wind speed}}{\text{strong wind speed hit} + \text{missed strong wind speed}}, \quad (3)$$

$$\text{false alarm ratio} = \frac{\text{false alarm strong wind speed}}{\text{false alarm strong wind speed} + \text{strong wind speed hit}}, \quad (4)$$

$$\text{false alarm rate} = \frac{\text{false alarm strong wind speed}}{\text{false alarm strong wind} + \text{weak wind speed hit}}. \quad (5)$$

The hit rate indicates the fraction of observed strong winds that were correctly predicted by the model, while the miss rate represents the fraction of strong winds that were not predicted. The false alarm ratio indicates the fraction of predicted strong winds that did not occur. Finally, the false alarm rate represents the fraction of weak winds that were incorrectly forecasted as strong winds. A well performing model would result in a high hit rate, as well as a low false alarm rate, ratio, and miss rate.

We further compute the bias between the observations and the HRRR model output utilizing the following equation (where N refers to the number of observations):

$$\text{bias} = \frac{1}{N}\sum_{i=1}^{N}(x_{\text{observations}} - x_{\text{HRRR}}). \quad (6)$$

3. HRRR evaluation

a. *Wind speed*

1) CORRELATION AND BIAS

We begin this study by examining the HRRR wind speed for the 6th forecast hour (note that an analysis of the forecast hour is provided in section 4). Displayed in Fig. 4 are joint probability distribution functions (PDFs) of the observed wind speeds compared to HRRR-forecasted wind speeds for each of the analyzed years [2016/17–2021/22 winter seasons; (Figs. 4a–f), respectively], as well as a combination of all years (Fig. 4g), including all available observation sites and HRRR initialization times of 0000, 0600, 1200, and 1800 UTC. Here, the model performance is examined in terms of r^2 and biases, as shown in the upper right corner. The r^2 values are similar across all years (ranging from 0.25 to 0.31) with the exception

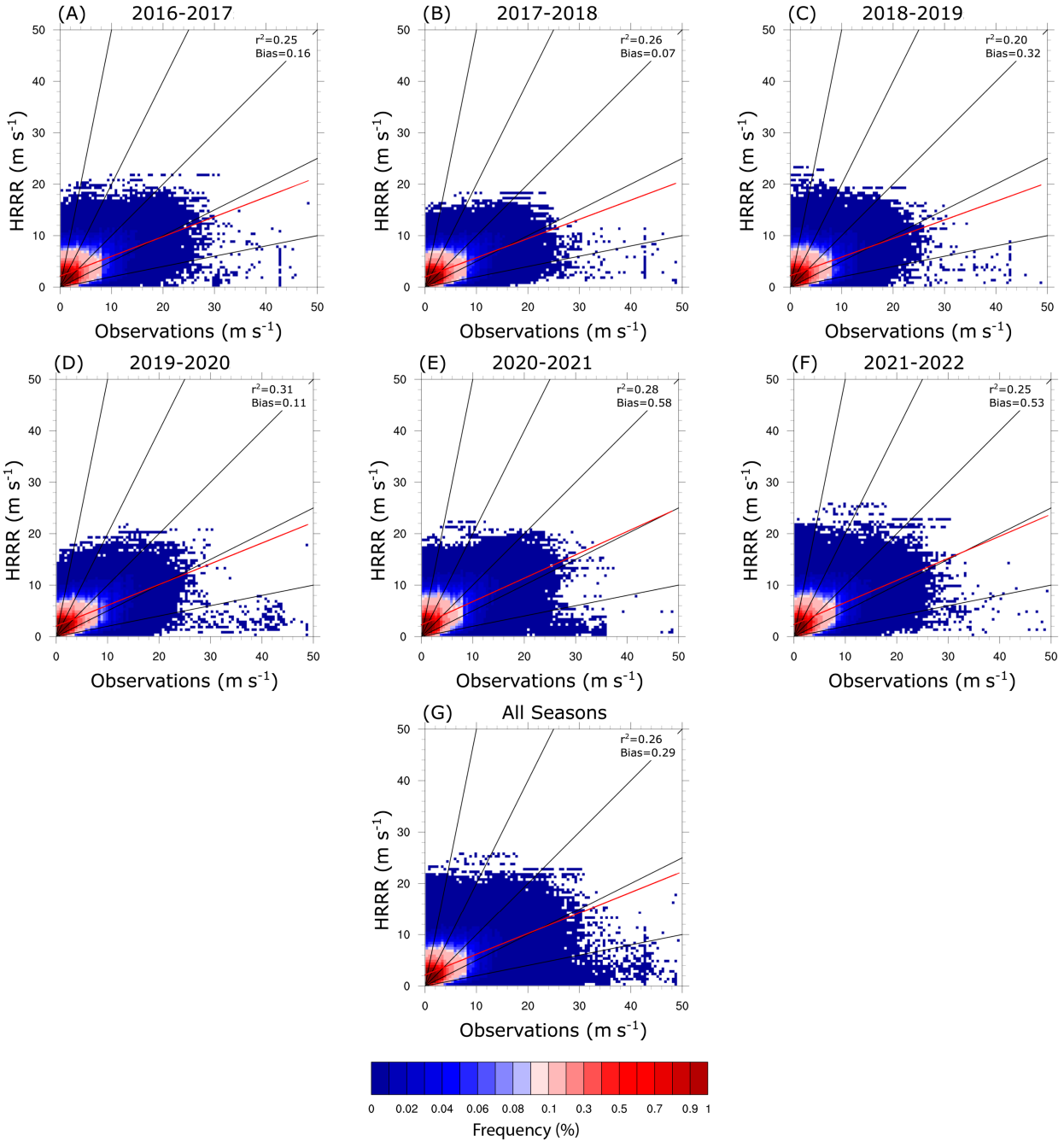

FIG. 4. Joint PDFs (%) of wind speeds (m s^{-1}) for DJF of the (a) 2016/17, (b) 2017/18, (c) 2018/19, (d) 2019/20, (e) 2020/21, and (f) 2021/22 seasons and (g) all years combined. The r^2 values and biases (m s^{-1}) are shown in the top right of each panel. The solid black lines represent the 1:1 line, as well as the 2:1, 1:2, 5:1, and 1:5 lines. The red lines show the best fit line for each panel.

of 2018/19 (0.20). Further, biases across all years are also similar, with a range of 0.07–0.58 m s^{-1}, with the largest bias occurring during 2020/21. Therefore, the change in the diagnosis of the 10-m wind in the HRRR that occurred in 2018/19 appears to have had a negligible impact on the model's performance thereafter. Interestingly though, 2018 is the year with the lowest r^2 and third highest bias, coinciding with the fewest high wind warnings issued by the NWS offices in the study domain (Iowa Environmental Mesonet 2022). Given the general similarity across all years, the majority of the subsequent analysis will be focused on the combination of all years (Fig. 4g), which results in an r^2 value of 0.26 and a bias of 0.29 m s^{-1}. It is important to note that due to the large fluctuations in winds in nature and the high spatiotemporal variability, a low r^2 is

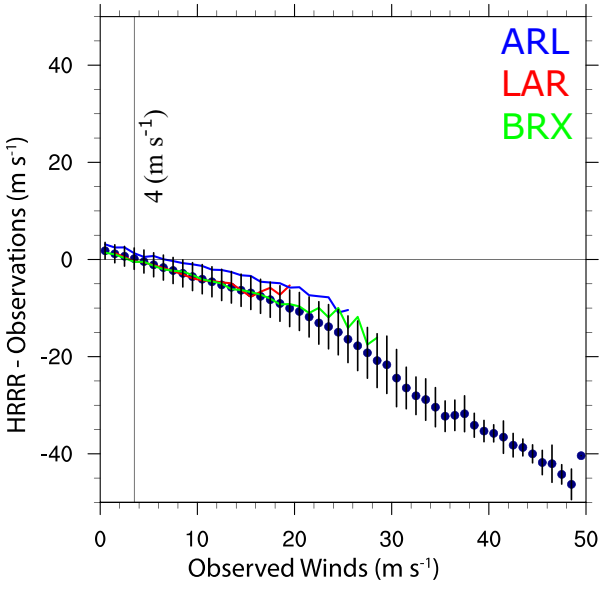

FIG. 5. HRRR wind speed bias at 1 m s^{-1} wind speed intervals for all years combined. The vertical line denotes the demarcation (i.e., zero bias) between positive and negative biases for weak and strong winds, respectively. For reference, overlaid are the HRRR wind speed biases for selected wind-prone sites: ARL, LAR, and BRX (blue, red, and green, respectively). Vertical bars are ±1 standard deviation.

not necessarily unexpected when comparing simulated versus observed wind speeds, and this is motivation for the use of contingency tables and additional metrics, as described in section 2.

While the bias across all years is rather small (≤0.58 m s^{-1}), the following analysis will focus on how this bias is affected by the disproportionately high occurrence of weak winds compared to strong winds, the latter of which is the focus of this study. For wind speeds in excess of 20 m s^{-1}, the HRRR model has a bias of −13.70 m s^{-1}, while the bias is much smaller for wind speeds below this threshold, i.e., 0.31 m s^{-1}. In other words, the wind speed bias is wind speed dependent (e.g., Fovell and Gallagher 2020, 2022). Moreover, for the strongest observed winds, i.e., wind speeds in excess of 40 m s^{-1}, which can cause significant damage, the HRRR model rarely predicts winds in excess of 10 m s^{-1}. In fact, according to Fig. 4g, there are a very small number of instances in which the HRRR correctly forecasts wind speeds greater than 20 m s^{-1} (only 128 times).

Given the identified wind speed–dependent bias in HRRR-simulated wind speeds, we further extend the HRRR model and observation comparison by examining the forecast bias in 1 m s^{-1} bins, which is shown in Fig. 5 for all years combined. For observed wind speeds less than 4 m s^{-1}, the HRRR model has a small positive bias. Moving to the stronger winds, the bias becomes increasingly more negative. That is, as the observed wind speeds get stronger, the HRRR generally performs worse. For the most extreme cases with wind speeds between 40 and 50 m s^{-1}, the HRRR generally underpredicts the wind speeds by 30–40 m s^{-1}. Such drastic underprediction

		Observations	
		0-20 (m s^{-1})	20-50 (m s^{-1})
HRRR	0-20 (m s^{-1})	13,713,891	18,936
	20-50 (m s^{-1})	538	128

FIG. 6. Contingency table for observed wind speeds and HRRR-simulated wind speeds. Red shaded boxes indicate correct forecasts, while blue shaded boxes indicate incorrect forecasts. Darker shades correspond to more instances occurring in the category. Refer to Fig. 3 for guidance on interpreting the contingency table.

would be considered a forecast miss. The inability of the HRRR model to predict the most extreme winds is in line with the results of Fovell et al. (2022) for the Marshall Fire in Colorado, except that here we demonstrate that this issue is present across a much larger spatial domain and over a longer period of time, indicative of a systematic bias in the HRRR model's prediction of strong winds.

2) CONFUSION MATRIX ANALYSIS

As noted above, winds naturally exhibit high spatiotemporal variability. As such, a model with an Eulerian grid will likely not be able to accurately predict the wind speed at each point in space and time. However, with regard to NWP model forecasts of high winds, we are less concerned about the exact wind speed and more concerned about the predicted wind speed exceeding some threshold, e.g., the threshold for advisory or warning issuance. As such, we shift our analysis to confusion matrices of wind speeds binned between 0 and 20 m s^{-1} and between 20 and 50 m s^{-1} (Fig. 6). As a reminder, the 20 m s^{-1} (approximately 45 mph) filter is commonly used by WYDOT and the NWS CYS office to begin warning severe wind events over this region. Moreover, the 50 m s^{-1} (approximately 112 mph) upper limit is imposed because, as discussed in section 2, all data with values exceeding this threshold are removed from the dataset. These thresholds are specific to the study domain and may not be applicable to other regions due to differing mechanisms driving strong winds. This approach enables us to focus on the ability of the HRRR model to predict the relatively rare strong winds compared to its ability to predict the far more common calmer conditions.

Overall, Fig. 6 shows that across all years and sites in the study area, as expected, the vast majority of the observations fall in the 0–20 m s^{-1} category (13 713 891 total instances), with just 0.13% of the data points falling in the 20–50 m s^{-1} category (19 064 total instances). While that may not seem like a large number of strong winds, this 0.13% of the data represent times of unsafe driving conditions, potential wind damage, rapid wildfire expansion. When the observed wind speeds exceed 20 m s^{-1}, the HRRR model only predicts 128 of these instances accurately, missing approximately 19 000 of them. In other words, if interested in below-advisory-level winds speeds, the HRRR predicts the occurrence of such instances very well (missing just 0.01% of them), but for the strong winds, it performs quite poorly (missing 99% of them).

Taking a closer look at the HRRR performance through the forecast evaluation metrics described in section 2c, we find that the hit rate is <0.01 (miss rate of 0.99). This provides even further quantitative evidence that the strong winds are underforecasted by the HRRR. Moreover, the false alarm ratio and false alarm rate, which are 0.81 and <0.01, respectively. These results provide another perspective on the aforementioned biases in HRRR-predicted wind speeds, namely, approximately 81% of the time that the HRRR model predicts strong winds, it is a false alarm, whereas <1% of the time that weak winds are observed the model predicts strong winds. The latter statistic indicates good model performance, whereas the former statistic indicates poor model performance. It is important to highlight the fact that part of the reason for the disparity in the evaluation metrics is the disproportion of weak to strong winds. However, the take home message is that if the HRRR predicts strong winds, it is highly likely that they will not occur based on the analysis in this study for Wyoming and Colorado.

3) Site-specific analysis

A potential shortcoming of the analysis thus far is the spatial distribution of observing sites (Fig. 2) compared to the spatial distribution of strong winds (Fig. 1). The bias in site location to population centers could affect the results owing to the generally weak wind speeds in these areas, especially compared to other regions of the study domain (cf. Figs. 1 and 2). As such, we target the subsequent analysis on three sites with climatologically strong winds and a high occurrence of high wind warnings (see Fig. 1): LAR, ARL, and BRX. We note that ARL and BRX are WYDOT sites, and there is no standard height for those anemometers. For simplicity in our analysis, we assume a height of 10 m and note that if they are lower, the negative strong wind bias would only be increased in magnitude. Figure 5 shows that the wind speed forecasts from the HRRR model may be less biased at these sites than others. To analyze this conclusion in a more statistical manner, contingency tables for these three sites are shown in Fig. 7.

We begin by looking at the hit rate for strong wind speeds at all three sites. LAR (Fig. 7a) has a hit rate of 0.24 (miss rate of 0.76), while ARL (Fig. 7b) has a hit rate of 0.72 (miss rate of 0.28), and BRX (Fig. 7c) has a hit rate of 0.30 (miss rate of 0.70). These results are further motivation for a site-specific analysis because the HRRR actually performs quite well at the ARL location, especially compared to the hit rates of the other sites and all sites together. This good performance may be related to the high elevation of the ARL site and lack of influence of downsloping and gap flow. Moreover, this is promising for forecasters as the ARL site could be used as an indicator of strong winds, even if the HRRR does not suggest such winds will occur elsewhere; however, care must be taken in using the ARL site as a proxy for high winds over a larger domain so as not to induce increased false alarms. We also find a generally smaller miss rate when analyzing these selected wind-prone locations compared to all sites in the domain of interest.

Moving onto a discussion of false alarms, LAR has a false alarm ratio and rate of 0.36 and 0.02, respectively, i.e., approximately 36% of the time the HRRR predicts a strong wind event

(A) LAR

HRRR \	0-10 (m s^{-1})	10-50 (m s^{-1})
0-10 (m s^{-1})	1,526	185
10-50 (m s^{-1})	33	59

(B) ARL

HRRR \	0-10 (m s^{-1})	10-50 (m s^{-1})
0-10 (m s^{-1})	763	354
10-50 (m s^{-1})	97	907

(C) BRX

HRRR \	0-10 (m s^{-1})	10-50 (m s^{-1})
0-10 (m s^{-1})	1,181	632
10-50 (m s^{-1})	29	268

FIG. 7. Contingency tables for observed wind speeds and HRRR-simulated wind speeds at specific locations in Wyoming prone to strong winds: (a) LAR, (b) ARL, and (c) BRX.

for LAR, it is a false alarm, and when a weak wind event occurs, only 2% of the time is it incorrectly identified as a strong wind event. The large difference in these indices is again due to the substantially higher number of weak wind instances compared to the strong wind events. At ARL, which was shown above to have a higher hit rate compared to the other sites, the false alarm ratio is substantially lower at 0.10, whereas the false alarm rate is somewhat higher, at 0.11. These indices are essentially identical owing to the similar number of correctly forecasted weak and strong wind instances, and their closeness to 0 indicates generally good model performance for this site. Furthermore, at BRX, the false alarm ratio is 0.10, and the false alarm rate is 0.02. In general, these results show that for the 3 selected wind-prone locations, the model rarely (<10% of the time at most) predicts strong winds when weak winds occur; however, at two locations, LAR and BRX, when the HRRR model predicts strong winds, approximately a third of the time it does not occur, which provides further evidence of the forecast challenge with regard to strong winds in the study domain and suggests that attention needs to be given to understanding why the HRRR model exhibits this low performance at these two sites but not at ARL. The differences in altitude and surrounding topography of these sites could provide evidence to understand this shortcoming in the HRRR model, which is discussed in section 5 (topography) and Part II (model sensitivity to grid spacing and physics parameterizations).

4) Wind speed proxies

The above analysis demonstrates clear biases in the HRRR model forecasts of winds over the domain of interest. We also

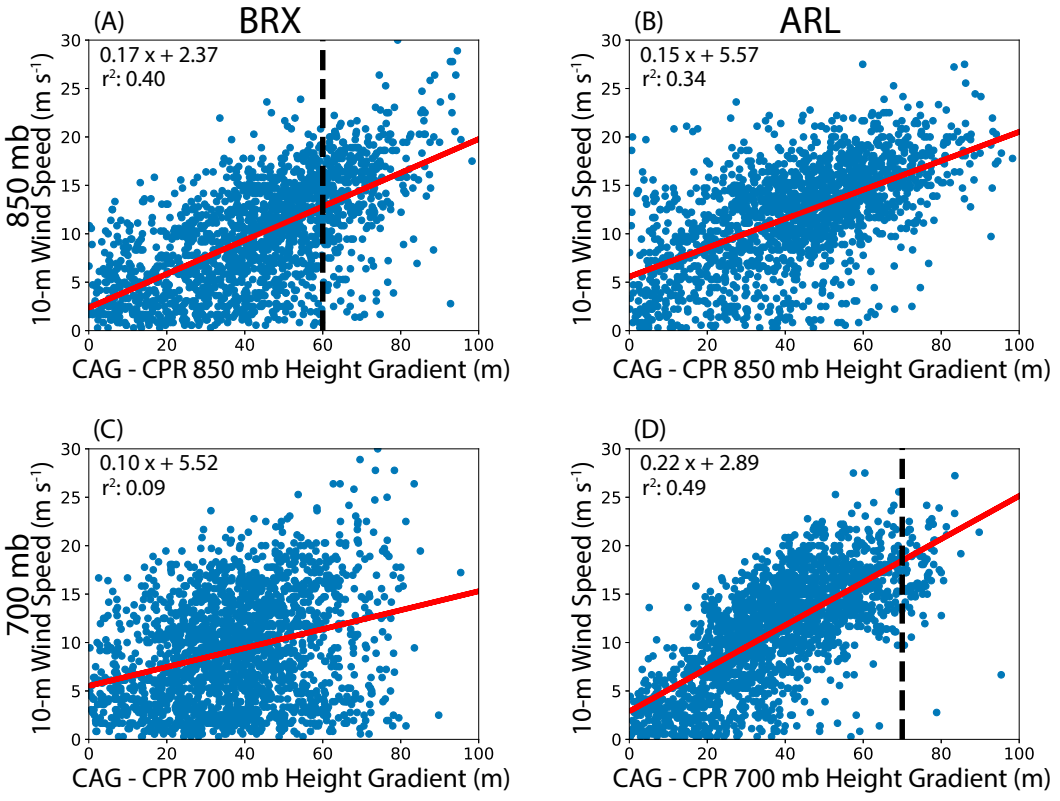

FIG. 8. (a) Scatterplot of the observed 10-m wind speeds (m s^{-1}) compared to the CAG-CPR HRRR 850-mb geopotential height gradient (m) at BRX. (b) As in (a), but for ARL. (c),(d) As in (a) and (b), but for the 700-mb geopotential height gradient (m). Overlaid on (a) and (d) are dashed lines denoting the height gradient threshold used for each location by the CYS NWS office.

elucidate potential forecasting issues related to the low hit rate of strong winds but also the high false alarm ratio. The natural objective then becomes to explain why the model underperforms in this context and provide guidance for improving the forecasts of high winds. Inherently, forecasting surface variables is challenged by topography and its representation in the model but also the myriad of land surface models, surface layer parameterizations, and PBL parameterizations used in NWP forecast models. A thorough analysis of the sensitivity of simulated winds over the domain of interest to these parameterizations is the focus of Part II.

In this study, we focus on an analysis of other fields that could be used as a proxy for fast low-level winds but are not as affected by the aforementioned low-level/surface parameterizations and model grid spacing. Such proxies are already used in some forecast offices in the Mountain West (e.g., Hammer 2015; Finch 2015), and in general, proxies are commonly used in forecasting applications (Hannigan and Godek 2020; Reymann et al. 1998). While several metrics were explored (e.g., 700-mb winds, 80-m winds, PBL depth; not shown; 1 mb = 1 hPa), only one presented significant skill with regard to forecasting strong winds in the study domain, which is already used by the NWS CYS office (Hammer 2015; Finch 2015), that is, 850- and 700-mb height gradients between Craig, Colorado (CAG), and Casper, Wyoming (CPR) (Fig. 8). These two height gradients are utilized as a proxy to forecast when strong wind speeds occur at BRX and ARL, respectively. For consistency with the above analysis, HRRR height gradients for forecast hour 6 are used in comparison with observed winds.

At ARL, there is a stronger correlation between large height gradients and strong wind speeds, especially for the 700-mb height gradient (i.e., r^2 of 0.49 at 700 mb; Fig. 8d versus 0.34 at 850 mb; Fig. 8b). For BRX, this relationship is less robust for the 700-mb height gradient, but shows a stronger relationship when the 850-mb height gradient is applied (i.e., r^2 of 0.40 at 850 mb versus 0.09 at 700 mb). This finding justifies the use of 700-mb height gradients for ARL and 850-mb height gradients for BRX, owing to their differences in elevation. Namely, ARL is located at an elevation of 2368 m, whereas BRX is 1562 m. For a standard atmosphere, this results in surface pressures of 760 and 839 mb. As such, 700 mb is not far above the surface at ARL, and 850 mb is just below the surface at BRX.

However, to determine whether these height gradient proxies provide skill over the HRRR-forecasted winds at the specific locations, an analysis of the confusion matrix for each site is needed. At ARL (BRX), the height gradient threshold performed better at 700 mb (850 mb), thus this level was

TABLE 1. Comparison of verification statistics for HRRR-simulated 10-m wind speeds compared to observations, as well as the HRRR-simulated isobaric height gradient proxy. Dashes indicate that the calculation cannot be performed. Height gradients at BRX (60 m) are derived at 850 mb, while at ARL (70 m), they are for 700 mb.

	HRRR-simulated 10-m wind speeds						HRRR-simulated height gradient					
	r^2	Total bias	Bias > 20 m s^{-1}	Hit (miss) rate	False alarm rate	False alarm ratio	r^2	Total bias	Bias > 20 m s^{-1}	Hit (miss) rate	False alarm rate	False alarm ratio
LAR	0.31	−1.57	0	0.24 (0.76)	0.02	0.36	—	—	—	—	—	—
ARL	0.57	−1.69	−7.03	0.72 (0.28)	0.11	0.10	0.49	—	—	0.17 (0.83)	0.01	0.03
BRX	0.49	−3.21	−10.83	0.30 (0.70)	0.02	0.10	0.40	—	—	0.16 (0.84)	0.02	0.08

utilized for the statistics. Height gradient thresholds of 70 and 60 m were used for the 700- and 850-mb height gradients, respectively (based on the findings of Hammer 2015; Finch 2015). A summary of the forecast statistics for the HRRR-forecasted winds and the height gradient proxy are shown in Table 1. At ARL, the HRRR results in an r^2 of 0.57, while at BRX, r^2 is 0.49, both of which are slightly higher than that for the height gradient proxy. Further, when comparing forecast statistics, the HRRR results in a higher hit rate at ARL (0.72 versus 0.17). This result is consistent at BRX with a hit rate of 0.30 for the HRRR compared to 0.14 for the height gradient proxy. At both sites, the false alarm rate and ratio for the height gradient proxy are less than or equal to those for the HRRR-simulated wind speeds. Therefore, while the false alarms decrease, care must be taken when applying these height gradient proxies to predict wind speeds, as wind speeds below the 20 m s^{-1} threshold commonly occur in conjunction with large gradients, leading to misses.

b. *Wind gust evaluation*

The above discussion highlighted biases in HRRR-forecasted wind speeds over the domain of interest. However, wind gusts can also be particularly hazardous owing to their temporal variability and magnitude (e.g., Fovell and Gallagher 2022). Thus, we turn our attention to an evaluation of HRRR-forecasted wind gusts.

STATISTICAL ANALYSIS

Displayed in Fig. 9 are joint PDFs of HRRR-predicted potential wind gusts (see section 2 for a discussion of how the potential wind gusts are computed for the HRRR model; hereafter, when referring to wind gusts, we these are potential gusts from the HRRR model and actual observed gusts from the observations) compared to observations for all years examined (Figs. 9a–f), as well as a combination of all years (Fig. 9g), which is similar to Fig. 4 for wind speeds. The results of the joint PDF analysis are consistent across all years, and thus, the majority of this analysis will once again be focused on the combination of all years (Fig. 9g). Overall, similar to the wind speed analysis, the HRRR does not perform well with respect to wind gusts, i.e., the HRRR overestimates low wind gusts but significantly underestimates strong wind gusts. These underestimations from the HRRR are more significant as the HRRR wind gust is a potential wind gust and is intended to be larger than the actual observed wind gusts. However, when compared to the wind speed analysis, the HRRR-forecasted wind gusts are slightly improved over the wind speeds, with r^2 values ranging from 0.16 to 0.35, compared to the wind speed r^2 values of 0.20–0.31. When looking at all years combined, the r^2 value is 0.26, which is the same as that for the wind speeds, and the biases are similar between the wind speed and gust forecasts (0.07–0.58 m s^{-1} and 0.06–0.53 m s^{-1}, respectively).

We can further see the improvement in the wind gust prediction compared with the wind speed forecasts in Fig. 10, which is similar to Fig. 5 for wind speeds. Similar to wind speeds, as the observed wind gusts become stronger and more severe, the HRRR bias becomes increasingly more negative. Comparing this result to the wind speeds (Fig. 5), there is a slight decrease in the bias at low speeds (below 20 m s^{-2}), decreasing from 0.31 m s^{-2} for wind speeds to −0.14 m s^{-2} for wind gusts. However, for wind speeds and gusts in excess of 20 m s^{-1}, the HRRR model exhibits biases of −13.70 and −11.32 m s^{-1}, respectively. This indicates that for the strongest winds (i.e., wind speeds/gusts in excess of 20 m s^{-1}), the HRRR performs better with regard to wind gusts than wind speeds, albeit the bias is still quite large in magnitude, and this reduction in magnitude may be related to the use of a potential wind gust in the HRRR model, which again is intended to exceed the actual observed gusts. Our results are in agreement with those of Fovell and Gallagher (2022), i.e., strong wind gusts are underestimated, and slow wind gusts are overestimated by the HRRR. However, Fovell and Gallagher (2022) only focused on April 2019 and 2020 and the CONUS, whereas we focused on the high terrain of Wyoming and Colorado and over longer time periods. While these study periods and regions differ, the results are consistent, suggesting that these biases may be systematic to the HRRR model and should be a focal point of future studies to determine the extent of this underestimation.

Analogous to the wind speed analysis and owing to the high spatiotemporal variability of wind gusts, the wind gust results can be interpreted through confusion matrices binned between 0 and 20 m s^{-1} and between 20 and 50 m s^{-1} (Fig. 11). Throughout all years examined in this study, the majority of wind gusts fall in the 0–20 m s^{-1} category, with ≈1% of the instances falling in the 20–50 m s^{-1} category (which is 10 times more instances than for wind speeds). Similar to the wind speed analysis, the majority of the data points fall within the 0–20 m s^{-1} category, and thus the percent correct is ≈99%.

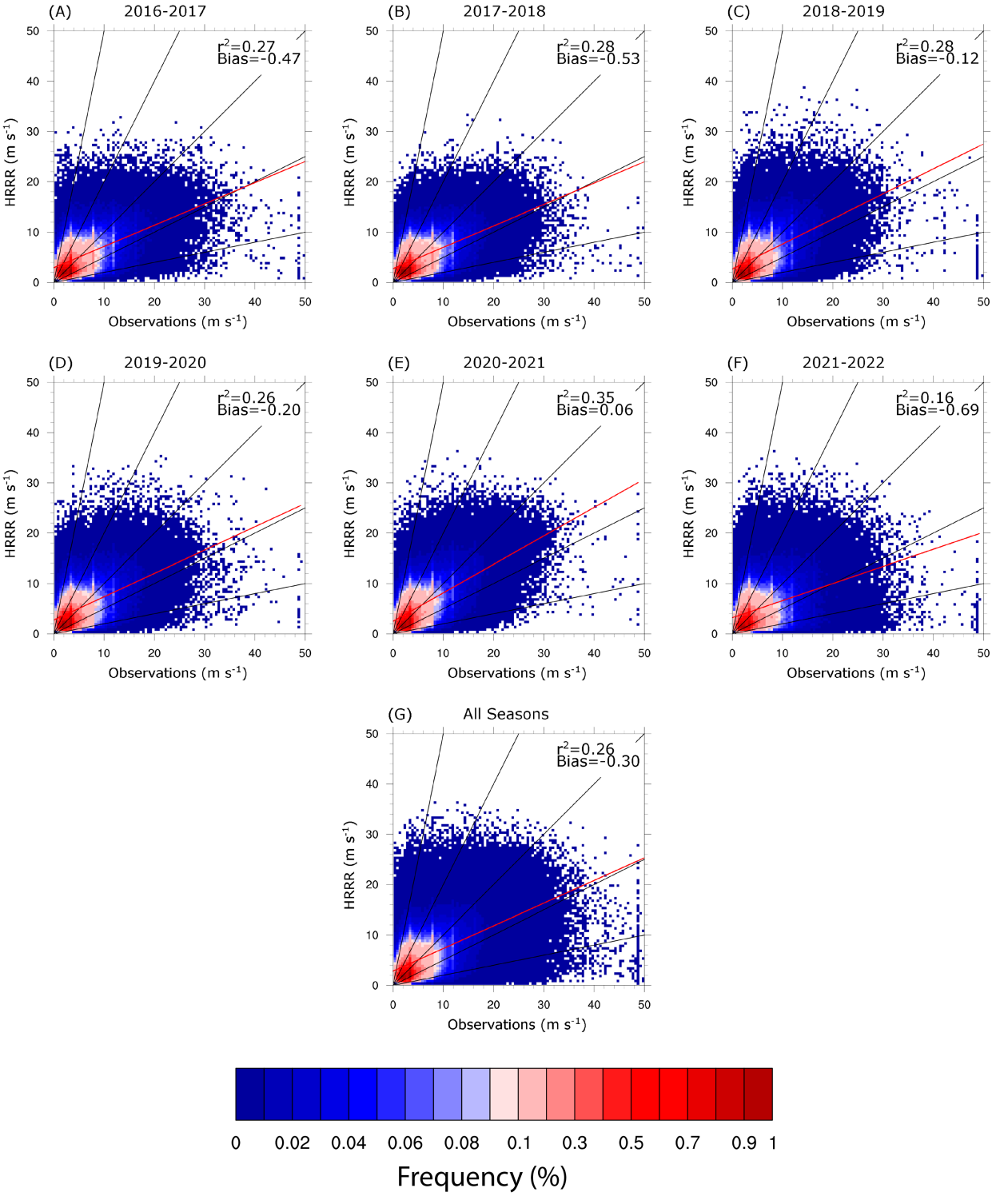

FIG. 9. As in Fig. 4, but for wind gusts.

While this is fairly high, when looking at the 20–50 m s^{-1} category, the hit rate is 0.22 (miss rate of 0.78), which is actually a drastic increase over the hit rate for strong wind speeds (0.01). However, even with this improvement, the HRRR still misses ≈ 32 000 of the strong wind gust data points, placing these instead in the 0–20 m s^{-1} category.

We further analyze the performance of the HRRR wind gust forecasts by an evaluation of the false alarm ratio and

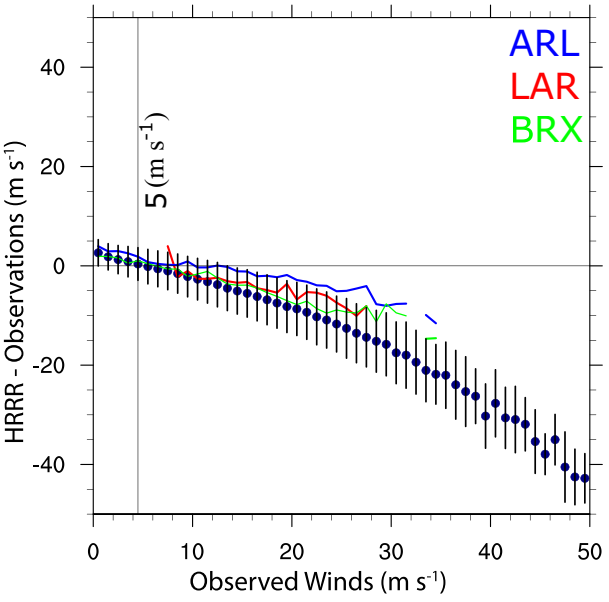

FIG. 10. As in Fig. 5, but for wind gusts.

rate, which are 0.26 and 0.001, respectively. This means that approximately 26% of the time the HRRR model predicts strong wind gusts, it is a false alarm, whereas only 0.1% of the time does the model predict strong wind gusts when weak gusts are observed. Compared to the wind speed analysis, these results indicate a small improvement with regard to the false alarm ratio (decrease from 0.81 for strong wind speeds to 0.26 for strong wind gusts) and a slightly worse false alarm rate (increasing to 0.001). However, the main conclusion for strong wind speeds remains, namely that it is unlikely that the model will indicate strong wind gusts when weak wind gusts are observed, but if the model does predict strong wind gusts, it is likely that these are false alarms, at least for the domain of interest in this study.

We once again investigate performance at strong wind prone locations, namely, LAR, ARL, and BRX for the HRRR-simulated wind gusts. We use Fig. 12 to compute the hit rate, false alarm rate, and false alarm ratio. Overall, the HRRR performs quite well across all three locations with hit rates of 0.63 (LAR), 0.91 (ARL), and 0.64 (BRX). There is also a low false alarm ratio across all sites, but a high false alarm rate. This indicates that when the model predicts strong wind gusts, it is not wrong often, but the model results in numerous instances in which there are predicted strong wind gusts when the wind gusts are in fact in the "slow" category. The wind gust

		Observations	
		0-20 (m s^{-1})	20-50 (m s^{-1})
HRRR	0-20 (m s^{-1})	2,409,773	32,386
	20-50 (m s^{-1})	3,188	9,191

FIG. 11. As in Fig. 6, but for wind gusts.

(A) LAR

HRRR	0-10 (m s^{-1})	10-50 (m s^{-1})
0-10 (m s^{-1})	13	172
10-50 (m s^{-1})	23	293

(B) ARL

HRRR	0-10 (m s^{-1})	10-50 (m s^{-1})
0-10 (m s^{-1})	487	141
10-50 (m s^{-1})	118	1391

(C) BRX

HRRR	0-10 (m s^{-1})	10-50 (m s^{-1})
0-10 (m s^{-1})	856	420
10-50 (m s^{-1})	103	745

FIG. 12. As in Fig. 7, but for wind gusts.

forecasts perform better that those of the wind speeds at these three locations, seen by the improvement in these statistics at all locations. There is an increase in the hit rate at LAR of 0.24 to 0.63, at ARL of 0.72 to 0.91, and at BRX of 0.30 to 0.64. Further, at all sites, there is an increase in the false alarm rate but a decrease in the false alarm ratio compared to the wind speed forecasts.

The wind speed analysis indicated the potential for isobaric height gradients to be used as a proxy for strong wind speeds; therefore, we repeat the analysis using the wind gusts, as shown in Fig. 13. The results are quite similar to those presented for the wind speeds. Specifically, the 850-mb height gradient is fairly well correlated with strong surface wind gusts at BRX owing to its lower elevation (r^2 of 0.42), whereas the 700-mb height gradient performs better at ARL (r^2 of 0.53). A summary of the forecast metrics for the HRRR-simulated wind gusts and the height gradient proxy is shown in Table 2. Here, similar to the results above for the wind speeds, the HRRR-simulated wind gust forecasts result in large r^2 values than the NWS metrics (0.67 versus 0.53 at ARL and 0.50 versus 0.42 at BRX). Moreover, the HRRR has higher hit rates, but also higher false alarm rates and ratios compared to the height gradient proxy. Thus, the same conclusion from the wind speed analysis regarding the care that must be taken when applying such proxies is applicable to wind gusts as well.

4. Forecast hour evaluation

The above analysis focused solely on the 6-h forecast time from the HRRR model. This was chosen as this forecast hour allows for adequate lead time for forecasters to warn events.

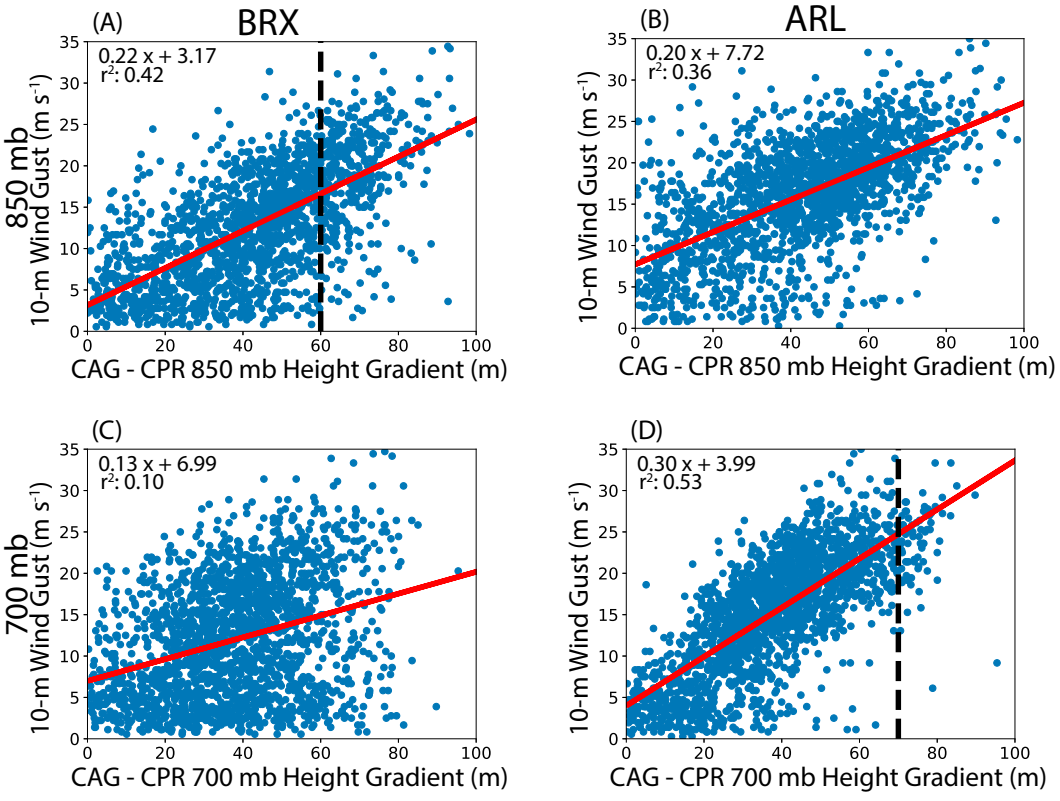

FIG. 13. As in Fig. 8, but for wind gusts.

However, the aforementioned wind speed biases could depend on the forecast time, with the hypothesis that the HRRR-forecasted wind speeds and gusts should exhibit reduced errors for shorter forecast lead times and increased errors for longer lead times.

We begin with a look into how the wind speed r^2 values change as the forecast hour increases, which is shown in Table 3. Clearly, for a forecaster, utilizing long lead times allows for more ample warning, but the question arises as to the cost of using guidance for longer lead times, especially given that issues with misses and false alarms discussed in section 3. As expected, the r^2 values decrease as the forecast hour increases. Forecast hour 3 has an r^2 value of 0.30, while the r^2 value of forecast hour 18 decreases to 0.19. However, these changes are marginal, and again low r^2 values are not uncommon in wind speed prediction. As such, the biases can provide more information regarding the performance of the model. For the dataset analyzed in this study, as the lead time increases, the magnitude of the wind speed biases for winds > 20 m s^{-1} increases in magnitude (from -13.10 m s^{-1} at forecast hour 3 to -14.70 m s^{-1} by forecast hour 18). We further analyze the hit rate, miss rate, false alarm rate, and ratio, to allow for a complete analysis of differing forecast hour performance. As the lead time increases, the hit rate remains very low (from 0.03 at forecast hour 3 to 0.02 by forecast hour 18), while the false alarm ratio increases (from 0.70 at forecast hour 3 to 0.83 by forecast hour 18). Therefore, regardless of the metric used to analyze the wind speed forecasts, the HRRR performs worse with increasing lead time.

TABLE 2. Comparison of verification statistics for HRRR-simulated 10-m wind gusts compared to observations, as well as the HRRR-simulated isobaric height gradient proxy. Dashes indicate that the calculation cannot be performed. Height gradients at BRX (60 m) are derived at 850 mb, while at ARL (70 m), they are for 700 mb.

	HRRR simulated 10-m wind gusts						HRRR-simulated height gradient					
	r^2	Total bias	Bias > 20 m s^{-1}	Hit (miss) rate	False alarm rate	False alarm ratio	r^2	Total bias	Bias > 20 m s^{-1}	Hit (miss) rate	False alarm rate	False alarm ratio
LAR	0.39	−2.45	−3.64	0.63 (0.37)	0.67	0.07	—	—	—	—	—	—
ARL	0.67	−1.20	4.01	0.91 (0.09)	0.20	0.08	0.53	—	—	0.14 (0.86)	0.01	0.02
BRX	0.50	−2.61	−8.57	0.64 (0.36)	0.11	0.12	0.42	—	—	0.12 (0.88)	0.08	0.32

TABLE 3. The r^2 values, biases for different forecast hours computed for observed wind speeds both greater than and less 20 m s^{-1}, hit rate, and false alarm rate.

Forecast hour	r^2	Bias $<$ 20 m s^{-1}	Bias $>$ 20 m s^{-1}	Hit rate (miss rate)	False alarm rate	False alarm ratio
F3	0.30	0.38	−13.10	0.03 (0.97)	<0.01	0.70
F6	0.29	0.31	−13.70	0.01 (0.99)	<0.01	0.81
F12	0.22	0.26	−14.26	0.02 (0.98)	<0.01	0.83
F18	0.19	0.27	−14.70	0.02 (0.98)	<0.01	0.83

We further conduct an analysis of the isobaric height gradient proxies with differing forecast hours (Fig. 14; only results at 850 mb for ARL and 700 mb for BRX are shown owing to the superior performance based on the above analysis). The r^2 values decrease with increasing forecast lead time, decreasing from 0.52 to 0.45 for 3–18-h lead times at 700 mb for ARL and 0.40 (at forecast hour 3 and 6) to 0.30 (at forecast hour 18) at 850 mb for BRX. However, these decreases in r^2 are again rather small, but the r^2 values remain larger than the values for all sites shown in Table 3. Moreover, the trend of lower hit rates and false alarm rates/ratios for the height gradient proxy is consistent across the different forecast lead times, and the results are similar for wind gusts (not shown).

5. Relation between wind forecast errors and complex terrain

Thus far, our analysis of errors in HRRR-predicted strong winds and gusts has considered all sites in the study domain or select sites conducive to strong winds and for which forecast proxies are commonly used to estimate the occurrence of strong winds. What we have yet to show is how the HRRR-predicted strong wind biases are related to the terrain in the study domain. To show this, we take the strong wind bias computed for each observational site and locate the sites exceeding the 90th percentile of the strong wind bias distribution, thus representing the sites with the largest underprediction of strong winds. These sites are mapped in Fig. 15a. Note that for this analysis, the threshold for strong winds was reduced from 20 to 10 m s^{-1} (as is also done in Part II) to ensure that there were a sufficient number of data points in the strong wind category for a robust analysis.

The results of this analysis indicate that the strong wind bias is not affected by oversampling of the Denver metro area (refer to Fig. 2 and the large number of observational sites in that are relative to the rest of the domain), with about half the sites exceeding the 90th percentile of the strong wind bias distribution in the Denver metro area and the other half being distributed elsewhere in the study domain. Moreover, a general trend indicated in Fig. 15b is the location of these sites on the leeward (eastern) side of steep topography. For example, the black (90th percentile and above) points along the Front Range of Colorado are west of the urban corridor, where the topography is quite steep. Evidence of the relation between large negative biases in the prediction of strong winds and elevation gradients is shown in Fig. 15b, which compares all-wind bias (blue) and strong wind bias (green and black) at each site with the corresponding topographical gradient. Here we quantify the gradient in the terrain simply as the elevation difference between each site and the maximum elevation in the east–west direction within a specified window (0.25°), with negative values indicating points on the eastern or leeward side of a topographical boundary, and vice versa. We restrict this analysis to the east–west direction owing to the prevailing wind direction being from the west in strong wind events. Corroborating the results from above, the all wind biases are generally positive, and the strong wind biases are generally negative. However, moving to increasingly more negative strong wind biases, sites generally exist below the 0-m elevation difference line, indicating that they are on the leeward side of the terrain. The specific locations of the sites with the largest strong wind bias magnitude are highlighted in Fig. 15a in yellow for reference.

This underperformance on the leeward side of steep topography suggests that the HRRR model struggles with simulating strong downslope windstorms, as suggested by Fovell et al. (2022) for the Marshall Fire in Boulder, Colorado, but here shown for a much longer time frame. Thus, it appears that the underestimate of downslope winds in the Marshall Fire is not an outlier but actually a systematic issue with the HRRR's forecasts of these events. The question remains though as to why the model struggles to accurately predict these strong winds, and a focus of Part II is to understand the sensitivity to the model to changes in resolution and model physics to better capture such events.

6. Conclusions

The first part of this study worked to evaluate the HRRR model with respect to strong winds (both wind speeds and gusts) over complex terrain, specifically focused on wintertime winds over Wyoming and Colorado. Moreover, the terrain itself can be a driver of these strong winds. This study aimed to understand the performance of the HRRR model regarding wind speed and gust forecasts as well as the potential applicability of proxies for estimating strong wind speeds used at local NWS offices, e.g., isobaric height gradients. The HRRR model was chosen as it is one of the highest-resolution operation NWP models available and is currently utilized by the NWS.

Wind speeds from the HRRR model were evaluated for six winter seasons (2016/17, 2017/18, 2018/19, 2019/20, 2020/21, and 2021/22). Generally, the HRRR slightly overestimates weak wind speeds but significantly underestimates strong wind speeds. This result is consistent across all winter seasons (indicating that wind speed forecast errors are not model version dependent), with a combined r^2 value of 0.26 and a bias of 0.29 m s^{-1}. Furthermore, of the almost 20 000 strong wind observations (>20 m s^{-1}) that

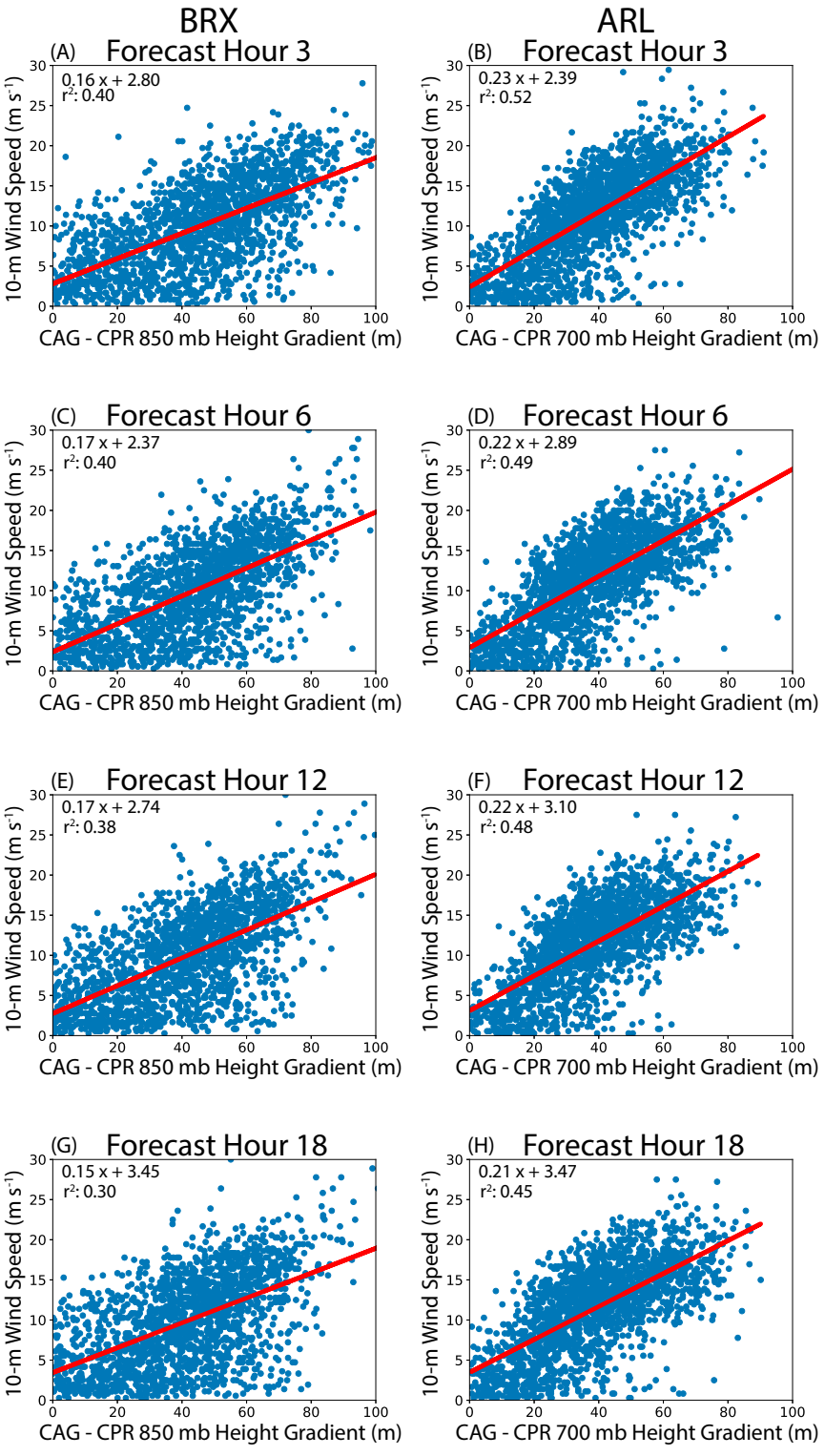

FIG. 14. As in Fig. 8, but for different forecast lead times. Forecast lead time increases from top to bottom. Data are shown for 700 mb at ARL and 850 mb at BRX.

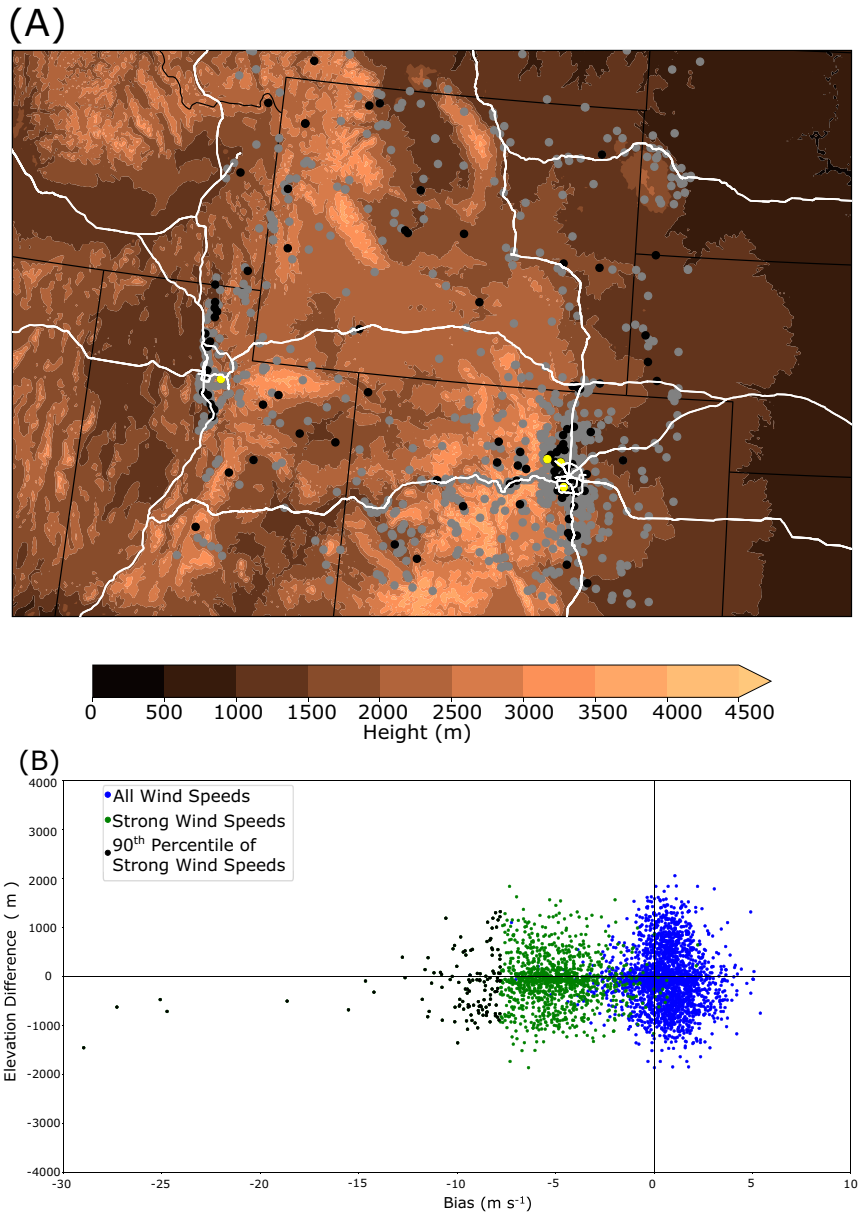

FIG. 15. (a) Map of sites exceeding the 90th percentile of the strong wind bias distribution (black; yellow indicates the outliers of the 90th percentile) across all seasons analyzed in this study, as well as the location of all sites in the 0th–90th percentiles (gray). The locations of the sites are overlaid on a contour map showing the surface elevation. (b) Site-specific wind speed biases relative to elevation difference around each site. Shown are the biases for all winds (black) and for strong winds only (green), with the sites exceeding the 90th percentile of the strong wind speed bias distribution highlighted (black). Horizontal and vertical lines denote no elevation gradient and no bias, respectively.

occurred during these six winter seasons in the study domain, the HRRR model only accurately forecasted 294 instances correctly, resulting in a very low hit rate (0.01). At the same time, the HRRR model is found to exhibit a high false alarm ratio and high miss rate (0.99).

Motivated by proxies already in use at some NWS offices (e.g., Hammer 2015; Finch 2015), attempts were made to identify other HRRR model fields that could be used as indicators of strong surface winds, including 700-mb wind speeds, PBL height, 80-m wind speeds, and isobaric height gradients. Only the latter, namely the 850- and 700-mb height gradients between Craig, Colorado, and Casper, Wyoming, provided viable results in terms of a proxy for strong surface winds at two wind-prone locales in SE Wyoming. With respect to wind speeds and wind gusts, BRX exhibits a larger r^2 value of 0.40 at 850 mb than 700 mb (0.09). While ARL exhibits a larger r^2 value of 0.49 at

850 mb than 0.34 at 700 mb. While these values are higher than the r^2 values for HRRR-simulated winds at all observation sites, they are in fact similar to the r^2 values at the respective locations. In general, the wind gust evaluation provided qualitatively similar results to that for wind speeds, albeit with a slight uptick in the performance of HRRR-forecasted wind gusts compared to wind speeds. Moreover, we find that for both wind speeds and wind gusts, the height gradient proxy leads to lower hit rates than the HRRR-simulated winds at these locations, although this comes with lower false alarm rates and ratios. As such, care should be taken when applying the height gradient metrics to forecast wind speeds and wind gusts because strong height gradients at 700 and 850 mb can occur in conjunction with below-advisory-level winds. However, the good correlation between the HRRR isobaric height gradients and wind speeds/gusts is motivation for an in depth analysis into why the HRRR is not accurately mixing high-momentum air in the boundary layer and thus misrepresenting strong downslope windstorms, which is the topic and focal point of Part II.

The final analysis of this study investigated the performance of different forecast hours/lead times. Longer lead times would provide forecasters with more time to provide more advanced warnings to the public. As expected, as the forecast hour increases from 3 to 18 h, there is a decrease in the r^2 values, and the magnitude of the bias for strong wind speeds increases. However, the isobaric height gradients were found to correlate well with strong surface winds at two wind-prone locations in SE Wyoming across all forecast lead times, albeit with a slight decrease in performance as the forecast lead time increases.

A potential uncertainty in this analysis, and as mentioned in section 2, is the anemometer height at each observational site in the MADIS dataset. While standards are in place for some data streams, this is not the case for every type of data contained in the dataset, e.g., state department of transportation sites may have varying heights, and anemometers may be closer to the ground at hydrometeorology sites. While errors in our assumptions of the heights of the anemometers would change the quantitative results of this work, we argue that the qualitative results would remain largely unchanged, especially for the strong wind biases, which were the focus of this work. For example, given the negative strong wind bias shown in this work, if the anemometer at an observational site was determined to be lower than assumed, this would indicate that the HRRR-predicted winds are even more strongly negatively biased when adjusted to the lower altitude (assuming a logarithmic wind profile that decreases toward the surface). Thus, we view the strong wind biases reported in this work as the lower limit in terms of magnitude. In addition, owing to the differences in anemometer heights and the complications this adds to model evaluation studies, increased consistency not across data types is encouraged in the future.

This study investigated the performance of HRRR wind speed and gust forecasts across Wyoming and Colorado (a region conducive to downslope and gap flow wind events). Our focus was on the systematic biases in these forecasts. Future work should focus on understanding how the wind biases differ for varying types of strong wind events and expand our understanding as to why model biases may vary under different synoptic conditions using techniques such as self-organizing maps (SOMs; e.g., Juliano and Lebo 2020). It is also important to note that due to differences in the mechanisms driving strong winds in the study domain (related to complex terrain) compared to other regions of the world, caution should be exercised when applying the results of this study to areas without downslope windstorms and gap flows, and future studies should expand the domain of interest to include other causes of severe winds. This is of particular importance given the finding that the HRRR model's largest underprediction of strong winds tends to occur on the leeward side of high mountains, suggesting poor performance when predicting downslope windstorms.

It is somewhat contradictory that the HRRR model has such large biases in strong wind speeds, yet the isobaric height gradients provide a reasonably good proxy for strong surface winds. This indicates that the reason behind the HRRR wind biases is inherently tied to the boundary layer and surface physics but could also be attributed to the 3-km grid spacing still being too coarse to fully represent the complex heterogeneous terrain of Wyoming and Colorado, as indicated by the most underperforming sites being located on the leeward side of the steep terrain in the study domain. Therefore, in Part II, we focus on addressing these potential factors influencing wind forecasts through winter-long sensitivity simulations.

Acknowledgments. EC, ZL, BG, SM, and RC acknowledge support from NOAA CSTAR Grant NA19NWS4680005. We also appreciate the advice and support from Trevor Alcott regarding the HRRR model.

Data availability statement. All data and HRRR model output used in this study are publicly available. The HRRR model output can be found at the following archive: https://registry.opendata.aws/noaa-hrrr-pds/. The MADIS dataset can be retrieved from https://data.eol.ucar.edu/dataset/9214. The ASOS data can be downloaded from https://mesowest.utah.edu/.

REFERENCES

Abatzoglou, J. T., B. J. Hatchett, P. Fox-Hughes, A. Gershunov, and N. J. Nauslar, 2021: Global climatology of synoptically-forced downslope winds. *Int. J. Climatol.*, **41**, 31–50, https://doi.org/10.1002/joc.6607.

Ashley, W. S., 2007: Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. *Wea. Forecasting*, **22**, 1214–1228, https://doi.org/10.1175/2007WAF2007004.1.

——, and T. L. Mote, 2005: Derecho hazards in the United States. *Bull. Amer. Meteor. Soc.*, **86**, 1577–1592, https://doi.org/10.1175/BAMS-86-11-1577.

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.

——, E. P. James, J. M. Brown, E. J. Szoke, J. S. Kenyon, R. Ahmadov, and D. D. Turner, 2020: Diagnostic fields developed for hourly updated NOAA weather models. NOAA Tech.

Memo OAR GSD-66, 58 pp., https://repository.library.noaa.gov/view/noaa/24212.

Bergen, W. R., and A. H. Murphy, 1978: Potential economic and social value of short-range forecasts of boulder windstorms. *Bull. Amer. Meteor. Soc.*, **59**, 29–44, https://doi.org/10.1175/1520-0477(1978)059<0029:PEASVO>2.0.CO;2.

Brennan, M. J., H. D. Cobb III, and R. D. Knabb, 2010: Observations of Gulf of Tehuantepec gap wind events from QuikSCAT: An updated event climatology and operational model evaluation. *Wea. Forecasting*, **25**, 646–658, https://doi.org/10.1175/2009WAF2222324.1.

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.

Carvalho, L., and Coauthors, 2020: The Sundowner Winds Experiment (SWEX) pilot study: Understanding downslope windstorms in the Santa Ynez Mountains, Santa Barbara, California. *Mon. Wea. Rev.*, **148**, 1519–1539, https://doi.org/10.1175/MWR-D-19-0207.1.

Collier, C. G., J. Dixon, M. S. J. Harrison, J. C. R. Hunt, J. F. B. Mitchell, and D. S. Richardson, 1994: Extreme surface winds in mid-latitude storms: Forecasting and changes in climatology. *J. Wind Eng. Ind. Aerodyn.*, **52**, 1–27, https://doi.org/10.1016/0167-6105(94)90036-1.

Collins, E., Z. J. Lebo, R. Cox, C. Hammer, M. Brothers, B. Geerts, R. Capella, and S. McCorkle, 2024: Forecasting high wind events in the HRRR model over Wyoming and Colorado. Part II: Sensitivity of surface wind speeds to model resolution and physics. *Wea. Forecasting*, **39**, 725–743, https://doi.org/10.1175/WAF-D-23-0037.1.

DeMaria, M., C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, https://doi.org/10.1175/BAMS-D-12-00240.1.

Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1.

Emanuel, K., 2000: A statistical analysis of tropical cyclone intensity. *Mon. Wea. Rev.*, **128**, 1139–1152, https://doi.org/10.1175/1520-0493(2000)128<1139:ASAOTC>2.0.CO;2.

Finch, Z. O., 2015: Utilizing gradients to forecast gap winds at Bordeaux, Wyoming. *40th National Weather Association Annual Meeting*, Oklahoma City, OK, National Weather Association, BP-28, https://nwas.org/annual-meeting-events/past-meetings/2015-agenda/.

Fovell, R. G., and A. Gallagher, 2020: Boundary layer and surface verification of the high-resolution rapid refresh, version 3. *Wea. Forecasting*, **35**, 2255–2278, https://doi.org/10.1175/WAF-D-20-0101.1.

——, and ——, 2022: An evaluation of surface wind and gust forecasts from the High-Resolution Rapid Refresh model. *Wea. Forecasting*, **37**, 1049–1068, https://doi.org/10.1175/WAF-D-21-0176.1.

——, M. J. Brewer, and R. J. Garmong, 2022: The December 2021 Marshall fire: Predictability and gust forecasts from operational models. *Atmosphere*, **13**, 765, https://doi.org/10.3390/atmos13050765.

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.

Hammer, C. L., 2015: Downslope windstorms: Improving forecast skill for Cheyenne, Wyoming. *40th National Weather Association Annual Meeting*, Oklahoma City, OK, National Weather Association, CP-16, https://nwas.org/annual-meeting-events/past-meetings/2015-agenda/.

Hannigan, A., and M. Godek, 2020: The utility of 1000–500 mb thickness and weather type as a rain-snow divide: A 30-year study at Albany, NY. *Atmos. Climate Sci.*, **10**, 372–391, https://doi.org/10.4236/acs.2020.103021.

He, C., F. Chen, M. Barlage, C. Liu, A. Newman, W. Tang, K. Ikeda, and R. Rasmussen, 2019: Can convection-permitting modeling provide decent precipitation for offline high-resolution snowpack simulations over mountains? *J. Geophys. Res. Atmos.*, **124**, 12 631–12 654, https://doi.org/10.1029/2019JD030823.

Iowa Environmental Mesonet, 2022: Iowa Environmental Mesonet. Iowa State University, accessed 15 September 2022, https://mesonet.agron.iastate.edu/plotting/auto/.

Juliano, T. W., and Z. J. Lebo, 2020: Linking large-scale circulation patterns to low-cloud properties. *Atmos. Chem. Phys.*, **20**, 7125–7138, https://doi.org/10.5194/acp-20-7125-2020.

Klemp, J. B., and D. R. Lilly, 1975: The dynamics of wave-induced downslope winds. *J. Atmos. Sci.*, **32**, 320–339, https://doi.org/10.1175/1520-0469(1975)032<0320:TDOWID>2.0.CO;2.

Kunz, M., S. Mohr, M. Rauthe, R. Lux, and C. Kottmeier, 2010: Assessment of extreme wind speeds from regional climate models—Part 1: Estimation of return values and their evaluation. *Nat. Hazards Earth Syst. Sci.*, **10**, 907–922, https://doi.org/10.5194/nhess-10-907-2010.

Martner, B. E., and J. D. Marwitz, 1982: Wind characteristics in southern Wyoming. *J. Appl. Meteor.*, **21**, 1815–1827, https://doi.org/10.1175/1520-0450(1982)021%3C1815:WCISW%3E2.0.CO;2.

McGovern, A., D. J. Gagne II, J. K. Williams, R. A. Brown, and J. B. Basara, 2014: Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. *Mach. Learn.*, **95**, 27–50, https://doi.org/10.1007/s10994-013-5343-x.

Mercer, A. E., M. B. Richman, H. B. Bluestein, and J. M. Brown, 2008: Statistical modeling of downslope windstorms in Boulder, Colorado. *Wea. Forecasting*, **23**, 1176–1194, https://doi.org/10.1175/2008WAF2007067.1.

Miller, P. A., M. F. Barth, L. A. Benjamin, R. S. Artz, and W. R. Pendergrass, 2007: MADIS support for UrbaNet. *14th Symp. on Meteorological Observation and Instrumentation*, San Antonio, TX, Amer. Meteor. Soc., JP2.5, https://ams.confex.com/ams/87ANNUAL/techprogram/paper_119116.htm.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

Oltmanns, M., F. Straneo, H. Seo, and G. W. K. Moore, 2015: The role of wave dynamics and small-scale topography for downslope wind events in southeast Greenland. *J. Atmos. Sci.*, **72**, 2786–2805, https://doi.org/10.1175/JAS-D-14-0257.1.

Pielke, R. A., and Coauthors, 1992: A comprehensive meteorological modeling system—RAMS. *Meteor. Atmos. Phys.*, **49**, 69–91, https://doi.org/10.1007/BF01025401.

Pokharel, B., B. Geerts, X. Chu, and P. Bergmaier, 2017: Profiling radar observations and numerical simulations of a downslope

wind storm and rotor on the lee of the Medicine Bow Mountains in Wyoming. *Atmosphere*, **8**, 39, https://doi.org/10.3390/atmos8020039.

Rahimi, S., W. Krantz, Y.-H. Lin, B. Bass, N. Goldenson, A. Hall, Z. J. Lebo, and J. Norris, 2022: Evaluation of a reanalysis-driven configuration of WRF4 over the western United States from 1980 to 2020. *J. Geophys. Res. Atmos.*, **127**, e2021JD035699, https://doi.org/10.1029/2021JD035699.

Reymann, M., J. Piasecki, F. Hosein, S. Larabee, G. Williams, M. Jimenez, and D. Chapdelaine, 1998: Meteorological techniques. AFWA Tech. Note AFWA/TN-98/002, 241 pp., https://apps.dtic.mil/sti/tr/pdf/ADA383417.pdf.

Taszarek, M., J. T. Allen, P. Groenemeijer, R. Edwards, H. E. Brooks, V. Chmielewski, and S.-E. Enno, 2020: Severe convective storms across Europe and the United States. Part I: Climatology of lightning, large hail, severe wind, and tornadoes. *J. Climate*, **33**, 10 239–10 261, https://doi.org/10.1175/JCLI-D-20-0345.1.

Wilks, D. S., 2020: *Statistical Methods in the Atmospheric Sciences*. Vol. 4. Elsevier, 818 pp.

Wyngaard, J. C., 2004: Toward numerical modeling in the "terra incognita." *J. Atmos. Sci.*, **61**, 1816–1826, https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2.

Zhong, S., and C. D. Whiteman, 2008: Downslope flows on a low-angle slope and their interactions with valley inversions. Part II: Numerical modeling. *J. Appl. Meteor. Climatol.*, **47**, 2039–2057, https://doi.org/10.1175/2007JAMC1670.1.