

Model Evaluation by Measurements from Collocated Remote Sensors in Complex Terrain

YELENA L. PICHUGINA,^{a,b} ROBERT M. BANTA,^{a,b} W. ALAN BREWER,^b J. KENYON,^{a,c} J. B. OLSON,^c D. D. TURNER,^c J. WILCZAK,^d S. BAIDAR,^{a,b} J. K. LUNDQUIST,^{e,h} W. J. SHAW,^f AND S. WHARTON^g

^a CIRES, University of Colorado Boulder, Boulder, Colorado

^b NOAA/Chemical Sciences Laboratory, Boulder, Colorado

^c NOAA/Global Systems Laboratory, Boulder, Colorado

^d NOAA/Physical Sciences Laboratory, Boulder, Colorado

^e National Renewable Energy Laboratory, Golden, Colorado

^f Pacific Northwest National Laboratory, Richland, Washington

^g Lawrence Livermore National Laboratory, Livermore, California

^h University of Colorado Boulder, Boulder, Colorado

(Manuscript received 3 January 2022, in final form 9 May 2022)

ABSTRACT: Model improvement efforts involve an evaluation of changes in model skill in response to changes in model physics and parameterization. When using wind measurements from various remote sensors to determine model forecast accuracy, it is important to understand the effects of measurement-uncertainty differences among the sensors resulting from differences in the methods of measurement, the vertical and temporal resolution of the measurements, and the spatial variability of these differences. Here we quantify instrument measurement variability in 80-m wind speed during WFIP2 and its impact on the calculated errors and the change in error from one model version to another. The model versions tested involved updates in model physics from HRRRv1 to HRRRv4, and reductions in grid interval from 3 km to 750 m. Model errors were found to be 2–3 m s⁻¹. Differences in errors as determined by various instruments at each site amounted to about 10% of this value, or 0.2–0.3 m s⁻¹. Changes in model skill due to physics or grid-resolution updates also differed depending on the instrument used to determine the errors; most of the instrument-to-instrument differences were ~0.1 m s⁻¹, but some reached 0.3 m s⁻¹. All instruments at a given site mostly showed consistency in the sign of the change in error. In two examples, though, the sign changed, illustrating a consequence of differences in measurements: errors determined using one instrument may show improvement in model skill, whereas errors determined using another instrument may indicate degradation. This possibility underscores the importance of having accurate measurements to determine the model error.

SIGNIFICANCE STATEMENT: To evaluate model forecast accuracy using remote sensing instruments, it is important to understand the effects of measurement uncertainties due to differences in the methods of measurement and data processing techniques, the vertical and temporal resolution of the measurements, and the spatial variability of these differences. In this study, three types of collocated remote sensing systems are used to quantify the impact of measurement variability on the magnitude of calculated errors and the change in error from one model version to another. The model versions tested involved updates in model physics from HRRRv1 to HRRRv4, and reductions in grid interval from 3 km to 750 m.

KEYWORDS: Data processing/distribution; Model evaluation/performance; Remote sensing; Renewable energy

1. Introduction

Numerical weather prediction (NWP) forecast models undergo continual changes and updates to the model physics, parameterization schemes, and horizontal grid spacing to improve model skill and forecast accuracy. Evaluating and quantifying the effects of these updates on model performance require accurate measurements of atmospheric variables such as wind speed, wind direction, and turbulence in various landscapes and atmospheric conditions.

To understand how model modifications may affect model skill, improve model physics, and increase the accuracy of short-term weather forecasts, the Second Wind Forecast

Improvement Project (WFIP2) was conducted in the Columbia River Valley between Oregon and Washington for 18 months from September 2015 to March 2017 (Olson et al. 2019; Shaw et al. 2019; Wilczak et al. 2019). A large array of measurement platforms, both in situ and remote sensing, was deployed to the study area to characterize atmospheric processes in the boundary layer and to support improvements in forecasting wind flow over complex terrain, which are challenging tasks. Wind flow in this area is complicated by mountainous terrain, coastal effects, and the presence of numerous wind farms. Project setup, instruments, and science goals are provided in overview papers (Shaw et al. 2019; Wilczak et al. 2019), and a detailed description of models used in this project can be found in Olson et al. (2019).

The project aimed to advance the forecasting skill of the HRRR model developed at the NOAA's Global Systems

Corresponding author: Yelena Pichugina, Yelena.Pichugina@noaa.gov

Laboratory (GSL). Refinements made to the model during WFIP2 were intended to improve how models represent complex terrain, the vertical mixing between the surface and the atmosphere, and the impact of turbulence in the horizontal as well as vertical directions (Olson et al. 2019).

Evaluation of parent and nested (HRRRNEST) versions of HRRR reforecast runs in the control (CNTR) and experimental (EXPR) configurations using sodars at 19 sites (Bianco et al. 2019) or scanning Doppler lidars at three sites (Pichugina et al. 2020) show how the updates in HRRR-EXPR affected model skill diurnally, season by season, over periods of atmospheric phenomena, and for differing terrain complexity. The results show that annually and seasonally averaged model evaluation statistics of 80-m wind speed, a typical hub-height of wind turbines on land in complex terrain, can vary significantly over the area, depending on mean-wind value and the location of instruments. Model errors vary over the diurnal cycle and in the vertical, with larger errors for all reforecast models found for stronger nocturnal winds and below 200 m AGL. The largest improvements in skill for both models resulted from reducing the horizontal grid spacing from 3 km (HRRR) to 750 m (HRRRNEST). Combining finer resolution with the updates in model physics of the experimental runs produced the best overall improvement statistics. The performance of these models, analyzed for different weather regimes observed in the Columbia basin (Bianco et al. 2019; Banta et al. 2020, 2021; Draxl et al. 2021; Olson et al. 2019; Pichugina et al. 2019, 2020; Wilczak et al. 2019), found larger errors for periods characterized by frequent occurrences of either wintertime cold pools or summertime marine intrusions.

In these previous studies, evaluation of the reforecast experimental runs against control runs of the HRRR and HRRRNEST were obtained using data from one type of remote sensing instrument at each site. To properly interpret the model evaluation results and relate error characteristics using the different remote sensing instrumentation, it is important to understand how the wind measurements, obtained using different sensors and procedures, different temporal and vertical resolution of profiles, and different criteria for data quality control, agree with one another, to provide insight into the calculated model errors as related to the measurement uncertainty of each instrument. The WFIP2 dataset provides a unique opportunity to compare wind measurements between multiple remote sensing instruments located at several research sites.

The present study uses measurements from all remote sensors at each of three WFIP2 sites to evaluate the HRRR versions in both domain configurations developed near the beginning (HRRRv1) and end (HRRRv4) of WFIP2 for the overlapping period (14–23 August 2016) of *reforecast* and *retrospective* runs, defined below. The second overlapping period (10–20 February 2016) was not considered in this paper because the two profiling lidars were not deployed until spring and a significant amount of data were missing for this period from the scanning HALO lidar at the Boardman site.

The paper is organized as follows: Section 2 describes the WFIP2 project, the sites selected for the study, the remote sensing instrumentation at each site, information on instrument characteristics and periods of operations, and HRRR

configurations used. Section 3 compares measurements of the annual and diurnal variability of hourly averaged 80-m (typical turbine hub-height) wind speeds available from each instrument in January–December 2016, including distributions of wind speeds and wind directions averaged over the overlapping period of operations. Wind-flow conditions for the period selected for model evaluation (14–23 August 2016) are given by data from scanning Doppler lidars in the first 1 km AGL with a summary of accompanying meteorological conditions. Section 4 compares HRRRv1 and HRRRv4 modeled winds within the first 1 km of the ABL and at 80 m during the periods selected for model evaluation. Section 5 presents time–height and time series analyses of measured and modeled winds and provides results on model improvements due to model physics refinements made during WFIP2, increased model horizontal grid resolution, and the combined impact of these factors. Finally, section 6 gives the conclusions.

2. Instrumentation and HRRR Model description

a. Collocated sensors at WFIP2 research sites

During WFIP2 several remote sensing instruments were collocated at each of three research sites separated by a distance of 40 and 71 km (Fig. 1a) from Wasco along a west-southwest/east-northeast line within a high-density (~2000 turbines) “wind-energy corridor” region of the Columbia River basin (Banta et al. 2020). Low-level winds were predominantly westerly through this region (Pichugina et al. 2019; Sharp and Mass 2002, 2004), placing Wasco upstream of the many wind farms clustered in this area; lidars there thus measured the flow undisturbed by the wind farms under these conditions (Fig. 1a).

Arlington was at a lower elevation, closer to the Columbia River, surrounded by wind farms. Boardman was at the lowest elevation, closest to Columbia River, farthest from Columbia Gorge, and well downstream of many wind farms during westerly flow conditions. Comparing westerly and easterly inflow profiles at the three sites thus provides information on how the wind profile shape and magnitude changed due to distance, terrain, and wind-farm effects (the analysis of wind-farm effects is in progress).

Table 1 presents the major information on remote sensing instruments used in this paper.

The table provides links to the quality-controlled data openly available to the public through the Data Archive and Portal (DAP) from the Atmosphere-to-Electrons (A2E) datasets supported by the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy (<https://a2e.energy.gov/data>). It also gives a list of the instrument owners responsible for the deployment and maintenance of the instrument, as well as the acquisition, quality control, and archiving of the data on the DAP in NetCDF format.

Instruments at each site were clustered at a close distance as illustrated in Fig. 1b for the Wasco site. Detailed information on all instruments over the WFIP2 study area including these three sites, as well as the instrument deployment strategy, can be found in Wilczak et al. (2019).

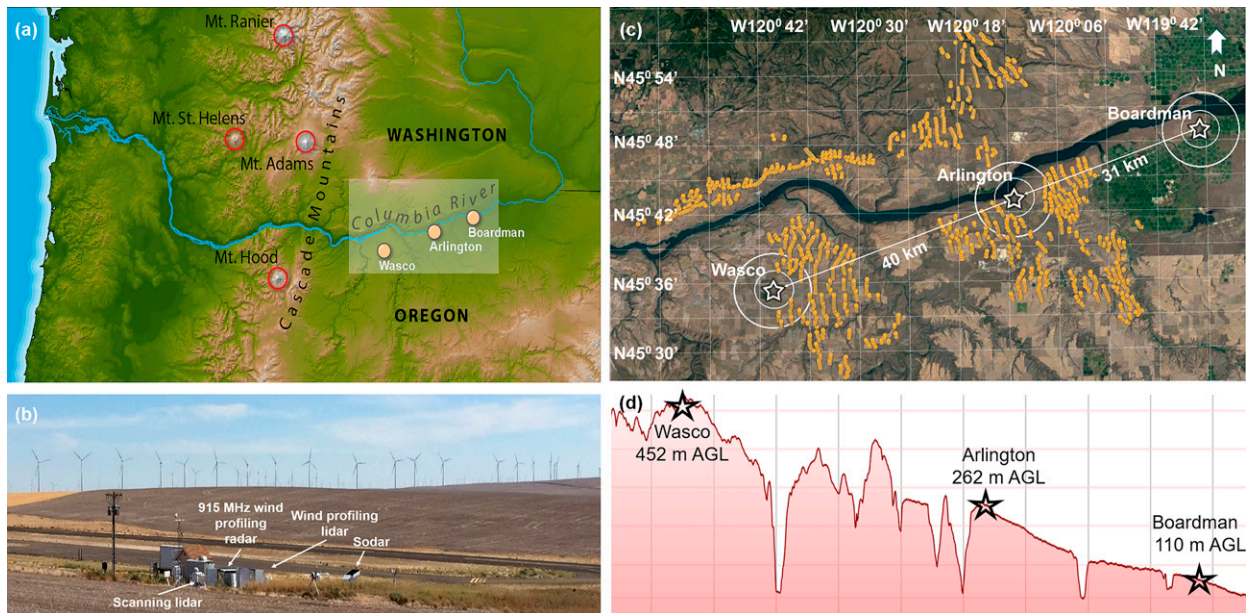


FIG. 1. (a) Google Earth map of the study area shows the terrain surrounding the research area and the location of three research sites Wasco, Arlington, and Boardman along the Columbia River Gorge. (b) Remote sensing instruments at the Wasco site (the picture credit to J. Sharp). (c) Detailed map of lidar locations. The white circles show the footprint (1- and 3-km radius) of the lidar horizontal scans at a 0° elevation angle. The surrounding wind turbines (total rated capacity of ~ 3800 MW) are indicated by clusters of dark yellow circles. The white line indicates the east–west transect of the study region along with the prevalent wind directions (Pichugina et al. 2019). (d) Terrain elevation along the white line in (c).

1) SCANNING DOPPLER LIDARS

Two NOAA scanning, pulsed Doppler lidars (Leosphere-200S) at the Wasco and Arlington sites provided real-time wind measurements from September 2015 through April 2017. The lidars are equipped with auxiliary systems to enable remote control and monitoring as well as real-time processing of the data. A postprocessing technique (Bonin and Brewer 2017; Bonin et al. 2017) allowed extending the vertical coverage of data profiles to heights of low lidar-backscatter signal, especially for wintertime cold-pool events, as illustrated in Pichugina et al. (2019). The third scanning Doppler lidar (HALO) operated continuously from the Boardman site during January–December 2016. All three Doppler lidars provided concurrent measurements of wind flow, each using similar scanning sequences and data-processing techniques.

The lidar measurement routine included a 15-min sequence of multiple azimuthal (conical, so-called “PPI”) scans, elevation (vertical slice, “RHI”) scans, and vertical staring of the lidar beam (vertical stare mode, to measure vertical velocity w). All data collected during both the conical and vertical-slice scans in the 15-min window were used in a velocity–azimuth display (VAD) analysis to obtain vertical profiles of mean wind variables (wind speed and direction) using line-of-sight (LOS) velocity measurements from the near-surface up to 3.5 km (Banta et al. 2002, 2020; Pichugina et al. 2008, 2019). The temporal (15 min) and vertical (~ 10 m below 200 m AGL and 25 m between 200 m and 1 km), resolutions of these profiles are well suited to understanding physical processes within the boundary layer (Banta et al. 2013). Data from scanning Doppler

lidars at each site are provided in A2E (2017a,b,c), respectively. A detailed description of WFIP2 scanning Doppler lidars including operational parameters, scanning patterns, and data quality control is provided in Pichugina et al. (2019). An assessment of the measurement precision for the 15-min winds in that paper showed diurnal variability with a peak value of ~ 0.02 m s $^{-1}$ around local noon, as well as dependence on terrain complexity for each site. Overall, wind speed measurement precision was determined to be small enough to quantify the accuracy of the NWP models to properly forecast challenging wind-flow conditions in WFIP2.

2) WIND PROFILING LIDARS

The WindCube profiling lidar samples line-of-sight velocities sequentially in four cardinal directions at a 28° angle from zenith (62° elevation angle) and a temporal resolution of 1 Hz per beam (Aitken et al. 2012; Bingöl et al. 2009; Lundquist et al. 2015; Rhodes and Lundquist 2013). These lidars provide estimates of wind speed, wind direction, and vertical velocity within the layer of 40–220 m (Bodini et al. 2019). The 2-min averages are based only on the 1-Hz LOS with CNR exceeding -22 dB; these datasets are posted to the DAP.

The ZephIR300 profiling lidar operated from the Arlington site (Wharton et al. 2015). The lidar uses coherent continuous-wave technology and a conical or VAD (1 scan per second, 50 measurements per scan, cone half-angle = 30°) to compute wind speed and direction at each programmed height. By adjusting the laser focus, winds may be sampled at 10 predetermined heights above ground level. One complete, multi-height

TABLE 1. Metadata of remote sensing instruments used in the paper. Organizations responsible for deployment and maintenance of the instruments: the Air Resources Laboratory (ARL), Department of Energy (DOE), the Chemical Science Laboratory (CSL), the University of Colorado Boulder (CU), the Lawrence Livermore National Laboratory (LLNL), the Physical Science Laboratory (PSL), and the University of Notre Dame (UND).

Sensor	DAP identifier	Instrument type	Owner	Min z (m)	Max z (m)	Resolution (vertical/temporal)	Period of operations
Lidar	https://a2e.energy.gov/data/wfp2/lidar.z03.b0	Leosphere WindCubeV1 profiling lidar	Wasco CU	40	220	20 m/2 min	Mar–Oct 2016
Lidar	https://a2e.energy.gov/data/wfp2/lidar.z04.b0	Leosphere 200S scanning lidar	NOAA CSL	10	>3000	10 m/15 min (<200 m); 25 m/15 min (>200 m)	Sep 2015–Apr 2017
Sodar	https://a2e.energy.gov/data/wfp2/sodar.z09.b0	ART Sodar Model VT-1	NOAA ARL	30	200	10 m/15 min	Oct 2015–Mar 2017
WPR 915	https://a2e.energy.gov/data/wfp2/radar.z04.b0	Vaisala LAP 3000	NOAA PSL	81	>5000	57 m/60 min	Oct 2015–Mar 2017
Lidar	https://a2e.energy.gov/data/wfp2/lidar.z05.b0	Leosphere 200S scanning lidar	Arlington NOAA CSL	10	>3000	10 m/15 min (<200 m); 25 m/15 min (>200 m)	Jan 2016–Mar 2017
Lidar	https://a2e.energy.gov/data/wfp2/lidar.z06.b0	ZephIR profiling lidar	DOE LLNL	(38) 40	200	10 m/10 min	Mar 2016–Mar 2017
Lidar	https://a2e.energy.gov/data/wfp2/lidar.z07.b0	HALO Photonics Streamline scanning lidar	Boardman UND	2	>3000	10 m/15 min (<200 m); 25 m/15 min (>200 m)	Jan–Dec 2016
Sodar	https://a2e.energy.gov/data/wfp2/sodar.z16.b0	ASC miniSODAR	NOAA ARL	30	200	10 m/15 min	Oct 2015–Mar 2017
WPR 915	https://a2e.energy.gov/data/wfp2/radar.z07.b0	Vaisala LAP 3000	NOAA ARL	124	>5000	55 m/60 min	Sep 2015–Mar 2017

TABLE 2. Days of missing measurements from each instrument during each month of the overlapping period of operations in March–October.

Month	Wasco				Arlington		Boardman		
	200S	WCubev1	Sodar	WPR	200S	ZephIR	HALO	Sodar	WPR
Mar	None	None	None	None	None	1–10	8, 9	None	None
Apr	None	1–19	None	None	None	None	27–30	None	None
May	None	12, 20–24	None	None	None	None	30–31	None	None
Jun	None	15	None	None	None	None	None	None	None
Jul	None	1–7	None	None	None	None	None	None	None
Aug	None	None	31	None	None	None	29	None	None
Sep	None	16, 17	None	None	24, 25	None	None	None	None
Oct	None	28–30	None	None	3–31	None	20, 21	10–31	31

scan over 360° takes place within fifteen seconds, and in the present configuration, three revolutions have been used for the derivation of the statistics at a single height. The backscattered signal comes mainly from the region close to the beam focus, where the signal intensity is at its maximum. The depth-of-focus increases nonlinearly as the beam is focused farther out from the lidar transmitter. Because of the scanning geometry, a 180° ambiguity in the wind direction was reconciled during internal data processing. The data are quality controlled by internal algorithms during collection which maximizes the signal-to-noise ratio and removes the effects of clouds. Similar to WindCube-v1, the ZephIR was programmed to provide measurements for 40–220 m AGL at vertical increments of 20 m plus an additional height at 38 m AGL. Unlike the WindCube the probe depth is not constant due to the increasing depth-of-focus with range, increasing with height such that the probe depth ranges from 1.5 m at the minimum height to 15.4 m at 100 m. The wind speed and wind direction data provided on the DAP at an averaging period of 10 min were linearly interpolated to 15 min to fit the model time resolution. Data from profiling Doppler lidars at Wasco and Arlington are provided in A2E (2017d,e), respectively.

3) SODARS

Nineteen sodar systems operated over the study area during WFIP2 (Wilczak et al. 2019). Sodars, being relatively inexpensive, can be widely distributed over the landscape to document the

spatial variability of the flow, but weak turbulence, weak stratification, or dry conditions can limit signal strength, at times leading to signal dropouts during evening transitions or above the nose of low-level jets. The sodar operated at the Wasco site was the ART (Atmospheric Research and Technology, LLC) VT1 model developed during the late 1990s. This model is a monostatic phased-array Doppler sodar system that includes a 48-element acoustical array. It provides 15-min vertical profiles of wind speed, wind direction, and *w* from 30 m up to 200 m at vertical intervals of 10 m.

An ASC (Atmospheric Sciences Corporation) MiniSodar at Boardman provided 10-min profiles of wind speed and direction at heights from 30 to 200 m AGL in 10-m range gates. After the initial quality control that rejected data when a pulse was below a specified signal/noise ratio or when the pulse was below a specified minimum amplitude, additional procedures were incorporated to flag odd-looking or unphysical data.

Only reliable data from both sodars (with flag values = 9) are used for the present analysis. More information about instrument setup and data filtering for the Wasco and Boardman sodars can be found in A2E (2017f,g), respectively.

4) WIND-PROFILING RADARS (WPRS)

The WFIP2 network of WPRs included three 449-MHz systems along the Pacific Coast and eight inland 915-MHz radar wind profilers also equipped with radio acoustic sounding

TABLE 3. The number (count) of 15-min wind speed and direction profiles available from each instrument during each month of the overlapping period of operations in March–October. The percent (%) of the available profiles is shown relative to the 15-min time intervals in each month. A 100% of hourly profiles availability from WPR at Wasco and Boardman sites, except the October 97% at Boardman (Table 2), are not shown in this table.

Month	Wasco						Arlington				Boardman			
	200S		Sodar		WCube v1		200S		ZephIR		HALO		Sodar	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Mar	2775	93	2878	97	2387	80	2855	96	1878	68	2634	89	1971	66
Apr	2839	99	2271	79	923	32	2866	100	2723	96	2764	96	2126	74
May	2954	99	2438	82	2968	100	2304	77	2873	97	2930	98	2462	83
Jun	2843	99	2584	90	2080	72	2869	100	2784	98	2848	99	2383	83
Jul	2871	96	2681	90	1881	63	2966	100	2880	100	2940	99	2604	88
Aug	2935	99	2435	82	2280	77	2953	99	2880	98	2842	95	1944	65
Sep	2812	98	2333	81	2192	76	2849	99	2784	99	2845	99	1762	61
Oct	2722	91	710	24	2131	72	193	6	2874	106	2786	94	2349	79

TABLE 4. The information on HRRRv1 reforecast and HRRRv4 retrospective runs in 2016.

Model	Reforecast control runs		Retrospective runs	
	HRRRv1	HRRRv1Nest	HRRRv4	HRRRv41Nest
Period of model run in 2016	The whole months of Apr, Aug, and Oct; about one week during the rest of the months		1) 10–20 Feb, 2) 14–23 Aug; the first initial hour on 14 Aug is 0900 UTC	
Horizontal grid	3 km	750 m	3 km	750 m
Output frequency	15 min	15 min	15 min	15 min
Forecast duration	0000–2400 UTC	0300–2400 UTC	0000–1800 UTC	0100–1800 UTC
Initialization frequency	0000, 1200 UTC	0000, 1200 UTC	0000, 0300, ..., 2100 UTC	0000, 0300, ..., 2100 UTC
Output height of profiles	20–200 m every 20 m; 200–500 m every 50 m; 500–1000 m every 100 m; additional height at 50 m			

system (RASS) temperature profiling capability. The 915-MHz WPRs, operated from Wasco and Boardman, provided hourly wind profiles up to 3–4 km through the planetary boundary layer in high (~60-m) and low (~120-m) vertical-resolution

modes. The first available height of measurements in the high-resolution mode at the Wasco site was 81 m AGL and at the Boardman site, 124 m AGL. WPR hourly data are used for distributions of wind speed and direction over the overlapping

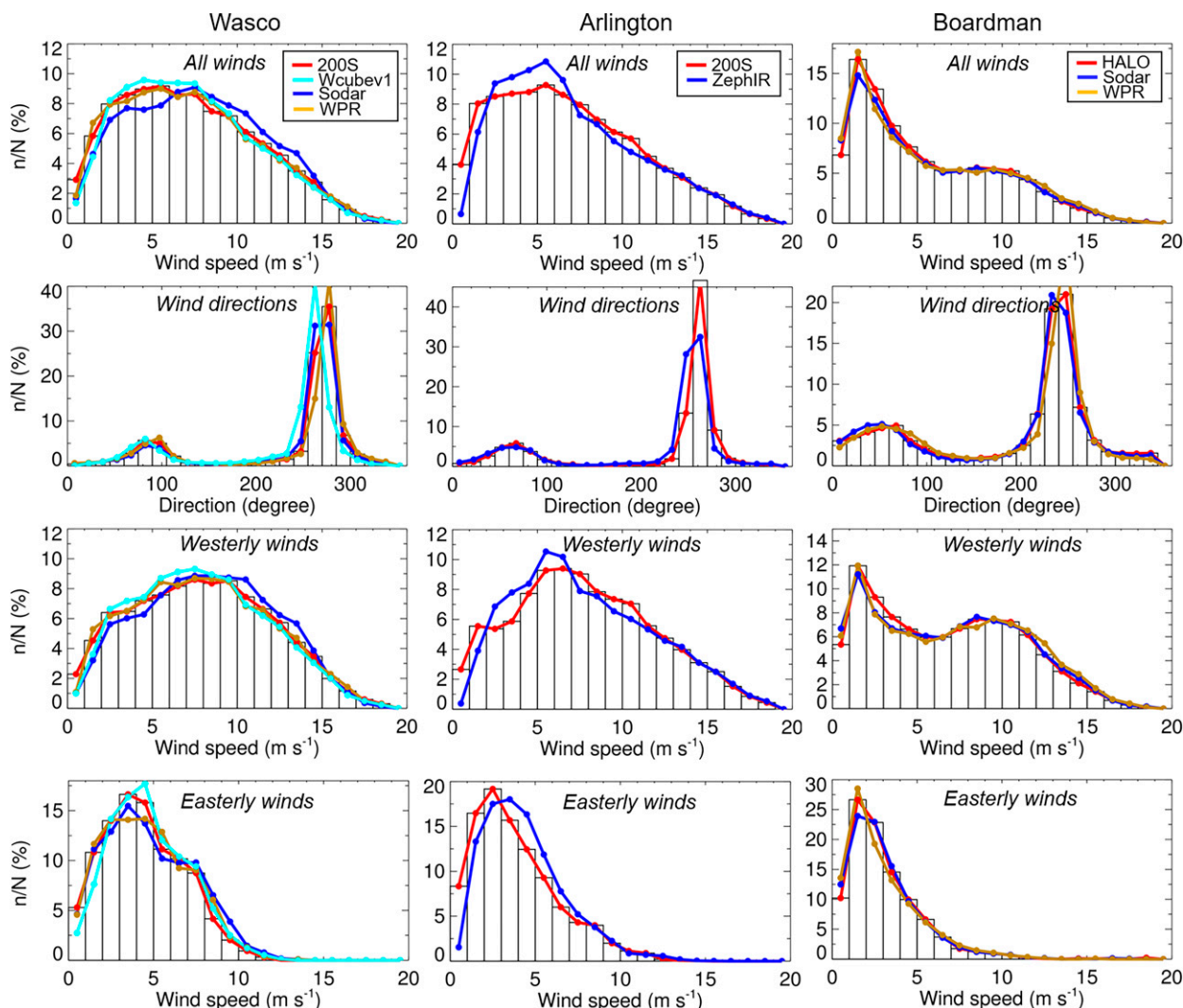


FIG. 2. Distributions of 80-m wind speed and direction from remote sensing instruments at Wasco, Arlington, and Boardman sites. Plots represent data available from each instrument in March–October 2016 and shown with the 15-min time resolution of all instruments except 1 h for WPRs. Note that WPR data at the Boardman site are used at the lowest measurement height of 124 m.

TABLE 5. Mean, standard deviation, and a number (count) of data points from distributions (Fig. 2) of wind speed and direction measured by each instrument at 80 m (except WPR at Boardman with the lowest height of 124 m).

Wasco				Arlington				Boardman			
Instrument	Mean	STD	Count	Instrument	Mean	STD	Count	Instrument	Mean	STD	Count
All wind speeds											
200S	6.6	4.0	30 244	200S	6.4	4.1	30 418	HALO	5.6	4.1	34 199
Sodar	7.3	4.7	33 450	ZephIR	6.8	4.1	23 001	Sodar	5.8	4.3	34 003
WPR	6.5	3.9	8640	—	—	—	—	WPR	5.7	2.9	7952
WindCube	6.9	3.9	19 026	—	—	—	—	—	—	—	—
Easterly wind speeds											
200S	4.2	2.3	8503	200S	3.7	2.4	8183	HALO	3.1	2.3	10 412
Sodar	6.3	6.0	9906	ZephIR	4.4	2.5	5722	Sodar	3.1	2.3	10 519
WPR	4.5	2.5	2504	—	—	—	—	WPR	2.9	2.2	2535
WindCube	5.0	4.7	4018	—	—	—	—	—	—	—	—
Westerly wind speeds											
200S	7.6	4.1	21 741	200S	7.4	4.2	22 235	HALO	6.7	4.2	23 787
Sodar	8.0	4.6	23 726	ZephIR	7.7	4.2	17 265	Sodar	7.0	4.5	23 430
WPR	7.3	4.1	6129	—	—	—	—	WPR	6.9	4.4	5413
WindCube	7.8	3.9	14 999	—	—	—	—	—	—	—	—
All wind directions											
200S	220.9	84.5	30 244	200S	209.6	89.0	30 418	HALO	194.1	92.5	34 199
Sodar	214.4	87.0	33 632	ZephIR	210.3	85.0	22 987	Sodar	189.8	93.2	33 949
WPR	220.6	85.4	8633	—	—	—	—	WPR	191.1	90.2	7948
WindCube	224.3	78.1	19 017	—	—	—	—	—	—	—	—
Easterly wind directions											
200S	91.2	28.0	8503	200S	67.7	29.9	8183	HALO	65.9	43.1	10 412
Sodar	87.4	34.8	9906	ZephIR	69.7	36.2	5722	Sodar	61.2	41.1	10 519
WPR	93.0	29.6	2504	—	—	—	—	WPR	68.5	41.2	2535
WindCube	80.8	32.9	4018	—	—	—	—	—	—	—	—
Westerly wind directions											
200S	271.6	21.7	21 741	200S	261.9	19.1	22 235	HALO	250.2	33.9	23 787
Sodar	267.5	25.8	23 726	ZephIR	256.9	21.4	17 265	Sodar	247.5	32.8	23 430
WPR	272.7	23.7	6129	—	—	—	—	WPR	248.5	28.6	5413
WindCube	262.8	21.1	14 999	—	—	—	—	—	—	—	—

period (March–October 2016) and are not used for evaluation of the 15-min outputs from models. Data from wind profiling radars at Wasco and Boardman are provided in A2E (2017h,i), respectively.

Days of missing measurements and the availability of 15-min wind speed profiles from each instrument are given in Tables 2 and 3 for each month of the overlapping period of operations (March–October).

b. HRRR Model configurations

The hourly updating, operational forecast models employ the Advanced Research version of the WRF Model (WRF-ARW; Skamarock et al. 2008; Benjamin et al. 2016). To support the goals of WFIP2 using the limited computational resources available to the project, a nonoperational version of the HRRR was used for selected reforecast periods during WFIP2 (Olson et al. 2019). This version of HRRR covered a

smaller domain than its operational counterpart but retained the standard 3-km horizontal grid spacing. Within the HRRR, a nested-domain version, “HRRRNEST,” was run at 750-m horizontal grid spacing. HRRR reforecasts employed a so-called cold start initialization, where initial conditions were supplied from the operational RAP without additional data assimilation or prior cycling, as described by Olson et al. (2019). Both models were run with a control (CNTR) physics configuration and again with an experimental (EXPR) physics configuration. The control configuration represents the state of HRRR physics at the beginning of WFIP2 (September 2015), which also corresponds to the physics of the operational HRRR-NCEP (HRRRv1), whereas the experimental configuration, represents physics developments made during WFIP2 (Olson et al. 2019), was similar to HRRRv3, as dates of code freezes were about 1 month apart.

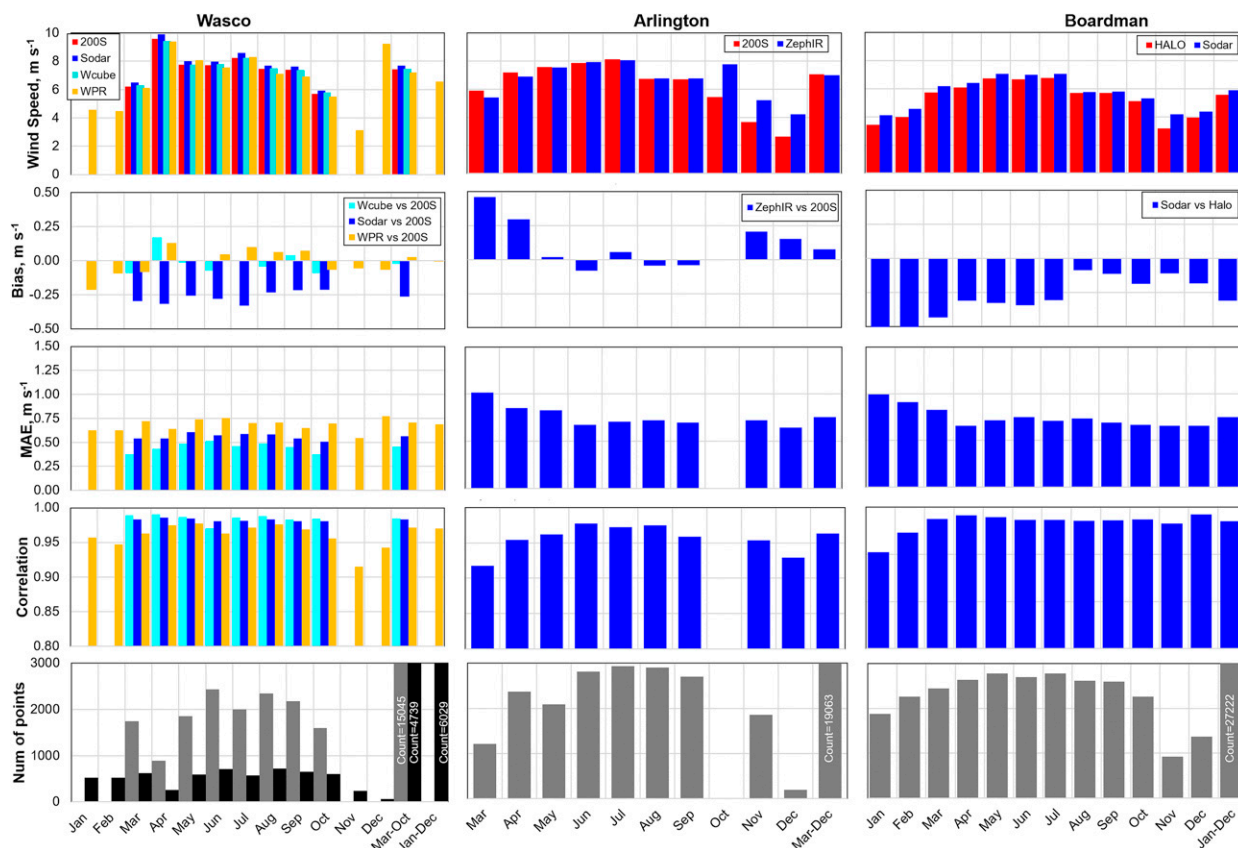


FIG. 3. Wind speed measurement-difference statistics (mean, bias, MAE, and R^2) between data from the scanning Doppler lidar and collocated instruments at Wasco, Arlington, and Boardman sites. Bar plots represent monthly mean values of 15-min (hourly for WPR at Wasco) winds at 80 m AGL for the period of the concurrent measurements (Table 5) at each site. The period-mean number of points shown in the bottom panels is scaled to fit the plot range and the actual values are shown for (gray) 15-min and (black) hourly data.

For the *reforecast* runs, this paper uses only a CNTR configuration of HRRR and HRRRNEST (hereafter HRRRv1 and HRRRv1 Nest). Both models were initialized twice daily at 0000 and 1200 UTC (the HRRRNEST lagged by 3 h), providing 24-h forecasts at 15-min output intervals.

In addition to the HRRR reforecast runs initialized without additional data assimilation (DA), a second Experimental (EXPR2) configuration of HRRR (hereafter HRRRv4) was run in fully cycled forecast-system mode (with DA) for both the RAP and HRRR. Referred to as *retrospective* runs, the RAP was cycled every hour, but the HRRRv4 was only cycled every third hour. The HRRRv4 with improved model physics and added wind-farm parameterization was run for two 10-day retrospective periods, 10–20 February and 14–23 August 2016 in two domain configurations: “parent” with (3-km grid) and “nested” version (750-m grid) (hereafter HRRRv4 and HRRRv4 Nest). Runs for each initial time consisted of 73 forecasts every 15 min, with the initialization lagged by 1 h for the nested version. Model and physics configuration for several versions of HRRR developed during the WFIP2 (James et al. 2022) are provided in the appendix (see Table A1). The information on HRRRv1 reforecast and HRRRv4 retrospective runs is given in Table 4.

The periods for reforecast and retrospective runs of models were selected as a compromise between model computational costs and representation of the major atmospheric events observed in the area. Cold pools in winter (McCaffrey et al. 2019) and westerly gap flows in summer (Banta et al. 2020) were two important weather types identified during the project (Olson et al. 2019; Pichugina et al. 2019, 2020; Wilczak et al. 2019).

3. Intercomparison of annual winds from collocated instruments

a. Distributions of wind speed and direction during the overlapping period of measurements from all sensors

During the spring–fall period of 2016, winds at the three research sites were blowing predominantly from the west or the east (Pichugina et al. 2019) indicating topographical channeling of the flow. It was shown that westerly wind directions were more frequent for this period due to the seasonal tendency for offshore ridging (Mass et al. 1986; Sharp and Mass 2004; Banta et al. 2020, 2021). Bimodal distributions of 80-m wind directions, with more frequent and stronger westerly winds, were found by all sensors at the three sites (Fig. 2) for the overlapping period of measurements (March–October 2016).

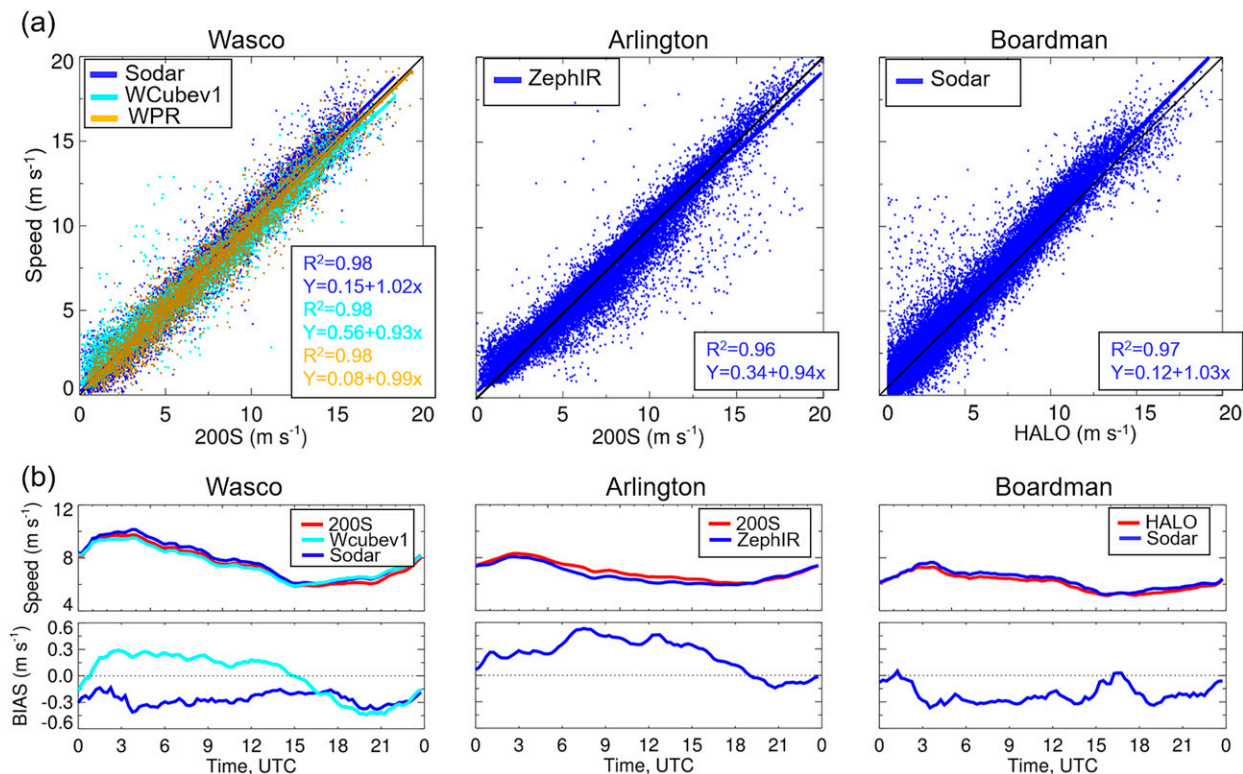


FIG. 4. (a) Scatterplots of wind speed at 80 m AGL for the period of the concurrent measurements from collocated instruments at each site vs scanning Doppler lidar. Each point represents 15-min data. Hourly data from the wind profiling radar (WPR) at the Wasco site are shown vs hourly averaged data from scanning Doppler lidar. (b) Time series of 80-m wind speed and bias (Doppler lidar–collocated instrument) averaged for concurrent measurements. All plots are shown for March–October at Wasco and Boardman sites. Data from Doppler lidar at the Arlington site are missing in October and plots are shown for the period of March–November.

The distributions of 80-m wind speed from all instruments at each site have similar shapes and mean values (Table 4). A large spread of the data about the mean values is observed for all instruments, with standard deviation (STD) values generally greater than half of the mean speed values. The dominant westerly winds were observed in 65%–80% of cases and are stronger from all instruments at Wasco (7.3–8.0 m s⁻¹) and Arlington (7.4–7.7 m s⁻¹) than at the lowest Boardman site (6.7–7.0 m s⁻¹) (Table 5). The easterly winds were half the strength of the westerlies at all sites, ranging between 4.2–6.3, 3.7–4.4, and 2.9–3.1 m s⁻¹ at Wasco, Arlington, and Boardman, respectively.

The mean westerly and easterly flow directions at all sites fell in the close range of 247°–272° and 61°–93°, respectively, with narrow differences between all instruments for westerly (10°, 5°, and 3°) and easterly (6°, 2°, and 5°) for Wasco, Arlington, and Boardman, respectively.

Good agreement in mean wind speeds between the scanning and profiling lidars at Wasco and Arlington [differences of 0.2–0.3 m s⁻¹ for westerly and slightly higher (0.8–0.6 m s⁻¹) for easterly winds], indicates the minimal influence of the measurement footprint on scanning lidar mean wind profiles in the complex terrain of the Columbia basin. The agreement between the scanning lidars and WPRs at Wasco and Boardman was 0.1–0.3 m s⁻¹ for both westerly and easterly winds.

b. Comparison statistics of data from all sensors versus scanning Doppler lidars

Each of the three sites had a scanning Doppler lidar, so these lidar data can be used as a reference to estimate wind speed measurement differences between data from the collocated sensors. Here we define the *measurement difference* in terms of bias, mean absolute “error” (MAE), and correlation coefficient R^2 between data from the scanning Doppler lidar and the indicated instrument at each site.

Plots of these variables computed for 15-min-averaged wind speeds at 80 m are illustrated in Fig. 3 for each month and the overall period of concurrent measurements at each site. Differences are calculated as lidar minus other instrument. Although we use the scanning Doppler lidars as a reference in this section, it is beyond the scope of this study to determine which of the instruments is most accurate.

We note that Pichugina et al. (2019) found a long-term agreement to be within 0.02 m s⁻¹ between 15-min mean winds measured by scanning Doppler lidar and those measured by a nearby tower at WFIP2, and Klaas et al. (2015) found a long-term agreement to within 0.10 m s⁻¹ between a profiling Doppler lidar and tower measurements at a complex terrain location.

The measurement-difference statistics (mean winds, bias, MAE, and R^2) and the number of concurrent data from

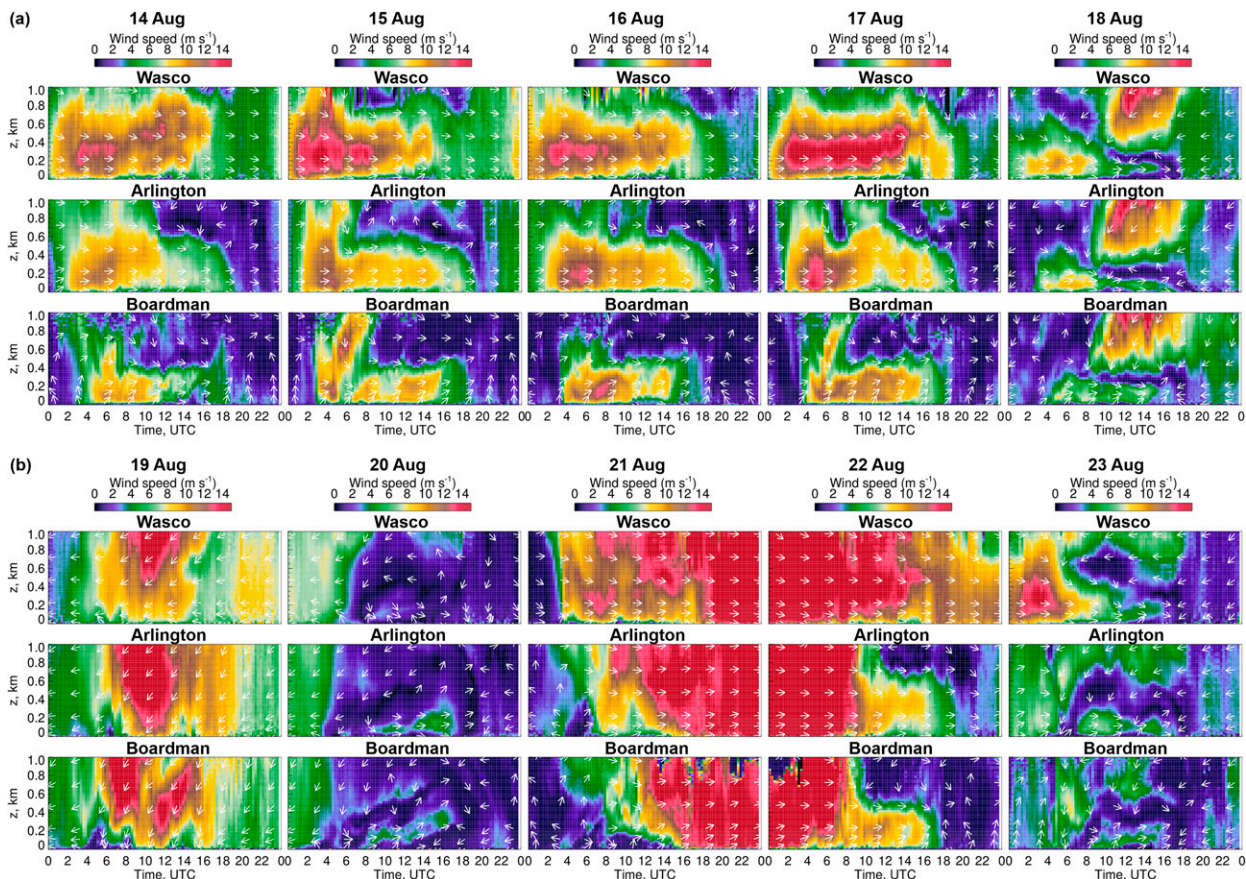


FIG. 5. Time–height cross sections of wind speed and wind direction from lidar measurements at 3 sites illustrate wind variability in the first 1 km AGL for each day from the period selected for evaluation of models. (a) The wind flow is shown for 14–18 Aug at Wasco, Arlington, and Boardman, shown from top to bottom, respectively. (b) As in (a), but for 19–23 Aug. Wind speed is shown by the color that filled each grid (15 min, 10 m) cell according to the color scales at the top of Figs. 5a and 5b, and wind direction is indicated by white arrows shown for each eighth grid cell to avoid too busy plots.

collocated instruments at each site show some month-to-month variability with larger data availability during warm months. Mean winds from sodars at Wasco and Boardman show stronger values compared to scanning lidar data (negative measurement bias). Measurement biases between the 200S Doppler lidars and the other instruments at Wasco (WindCube-v1, 915-MHz WPR) and the ZephIR-300 profiling lidar at Arlington, changed signs throughout the year, being negative mostly during cold months. At Wasco, bias magnitudes were generally $<0.3 \text{ m s}^{-1}$ for all instruments, showing larger values ($\sim 0.5 \text{ m s}^{-1}$) at Arlington and Boardman for some winter months. Scatterplots (Fig. 4a) show good agreement between instruments with a correlation coefficient of 0.98, 0.96, and 0.97 at Wasco, Arlington, and Boardman, respectively.

The measurement biases also showed a significant diurnal variability at each site (Fig. 4b). At Wasco, the largest differences ($0.33\text{--}0.46 \text{ m s}^{-1}$) between sodar and scanning lidar were observed during late-evening transition hours (0300–0500 UTC, 1900–2100 PST), and smaller values ($0.14\text{--}0.20 \text{ m s}^{-1}$) during the morning (1400–1800 UTC, 0600–1000 PST). Positive WindCube biases of $\leq 0.3 \text{ m s}^{-1}$ changed the sign to negative

values of comparable magnitude at 1700 UTC. Biases that change sign in this way would show smaller average values in calculating daily means. At Arlington, the ZephIR lidar showed a change in sign of the bias similar to the WindCube-v1 at Wasco albeit with a stronger change in magnitude. This effect may be stronger at Arlington compared to Wasco due to higher terrain complexity, resulting in stronger variations in wind speed measurement bias (from 0.53 m s^{-1} at 0730 UTC to -0.17 m s^{-1} at 2030 UTC). The sensitivity of the scanning Doppler-lidar wind measurements to vertical motions is expected to be negligible, due to the mix of lower-elevation scans used to measure the wind speed profile (Pichugina et al. 2019). At Boardman, sodar biases were negative during all 24 h, mostly showing magnitudes of $\leq 0.35 \text{ m s}^{-1}$ except for much smaller values during evening-transition (0000–0300 UTC) and morning (1500–1800 UTC) periods. The larger sodar biases at Wasco and Boardman may be attributable to both instruments operating in the field for a much longer period compared to wind-profiling lidars. All types of lidar were calibrated before the experiment and remotely monitored during WFIP2. The results of this study confirm that a good practice for any

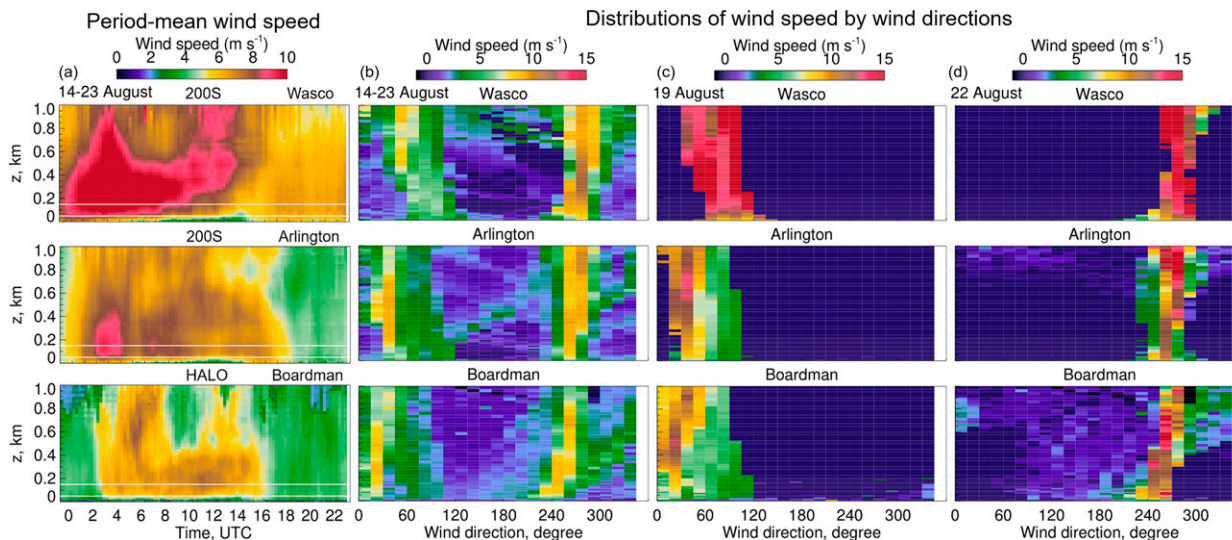


FIG. 6. (a) Period-mean wind speed from scanning lidars at three sites: Wasco, Arlington, and Boardman, shown from top to bottom, respectively. Distributions of wind speed by wind directions within the first 1 km AGL at three sites for (b) 14–23 Aug, (c) easterly winds on 19 Aug, and (d) westerly winds on 22 Aug.

long-term field campaign is frequent calibration and assessment of the instrument's ability to continuously operate and provide high-quality data.

c. Wind flow conditions during selected period (14–23 August)

The high temporal and vertical resolution of measurements from scanning lidars (Table 1) allowed us to analyze wind-flow conditions during 14–23 August in the first 1 km AGL, to better understand wind dynamics as shown by wind speed and predominant wind directions far above the maximum common height (~250 m) of data from all sensors.

Time–height cross sections of wind speed and direction (Fig. 5) illustrate day-by-day, site-to-site, and vertical variability during the selected period. Diurnal fluctuations in 80-m wind speed—strong increases or decreases of wind speed over a few hours, known as wind *ramp* events—are important for wind turbine operations. A significant ramping of wind speeds leads to a corresponding ramping in wind power production. Good agreement was shown (Wilczak et al. 2019, Pichugina et al. 2020) between wind power computed from 80-m west-southwest wind speed measured by scanning lidars at three sites and the fluctuations of total power generated over the Bonneville Power Administration (BPA) area, when large up or down power ramps reached 3 GW or more.

The beginning of this period (14–17 August) was dominated by westerly gap flow/marine intrusion winds (Banta et al. 2020, 2021), which brought strong nocturnal wind speeds to the area and a strong diurnal cycle of winds. A detailed evaluation of the model for the period of marine intrusions (15–17 August) is discussed in Banta et al. (2021). Subsequently, on 18 August a surface ridge passed through the study area and into the eastern part of the Columbia basin, causing a rare summertime easterly gap event with weaker

nighttime winds and a significant down ramp at the outset of the period, starting at ~1500 UTC at Wasco. The ridge and accompanying northeasterly winds persisted through 20 August. On 21 August a cold front swept through the study region from the northwest, attended by sharp surface pressure rises and westerly surges through the Gorge. A significant increase of 80-m wind speed on 21–22 August coincided with a >4-GW up-ramp of the BPA-generated power (not shown here). As the upper low behind the surface cold front progressed eastward, synoptic support for onshore flow diminished, resulting in a decrease of pressure gradient across the Cascades and the eastern Gorge and weakening winds on 22 August. An impressive down-ramp of winds occurred on this day after 0800 UTC. On 23 August pressure gradients continued to decrease, then reverse, as reflected by weakening wind speeds after ~0500 UTC and the shift to weak easterly flow after 1700 UTC.

Period-mean wind speeds were stronger (Fig. 6a) at Wasco with nighttime values > 10 m s⁻¹ below 500 m and weaker (~6 m s⁻¹ at 0300–1600 UTC) at Boardman.

Similar to the longer-term (Fig. 2) distributions of wind directions at the three sites, period-mean winds show two distinct modes (Fig. 6b), representing westerly and easterly flows as noted by Pichugina et al. 2019 for all seasons in 2016, although for some days winds were easterly (Fig. 6c) or westerly (Fig. 6d) through the 0–1 km AGL layer.

4. Modeled wind speed

Each of the model versions produced somewhat different results. Figures 7 and 8 show diurnal time–height, wind speed cross sections, and time series of the 80-m wind speed, respectively, composited for the 10 study days for each of the model versions at each measurement site.

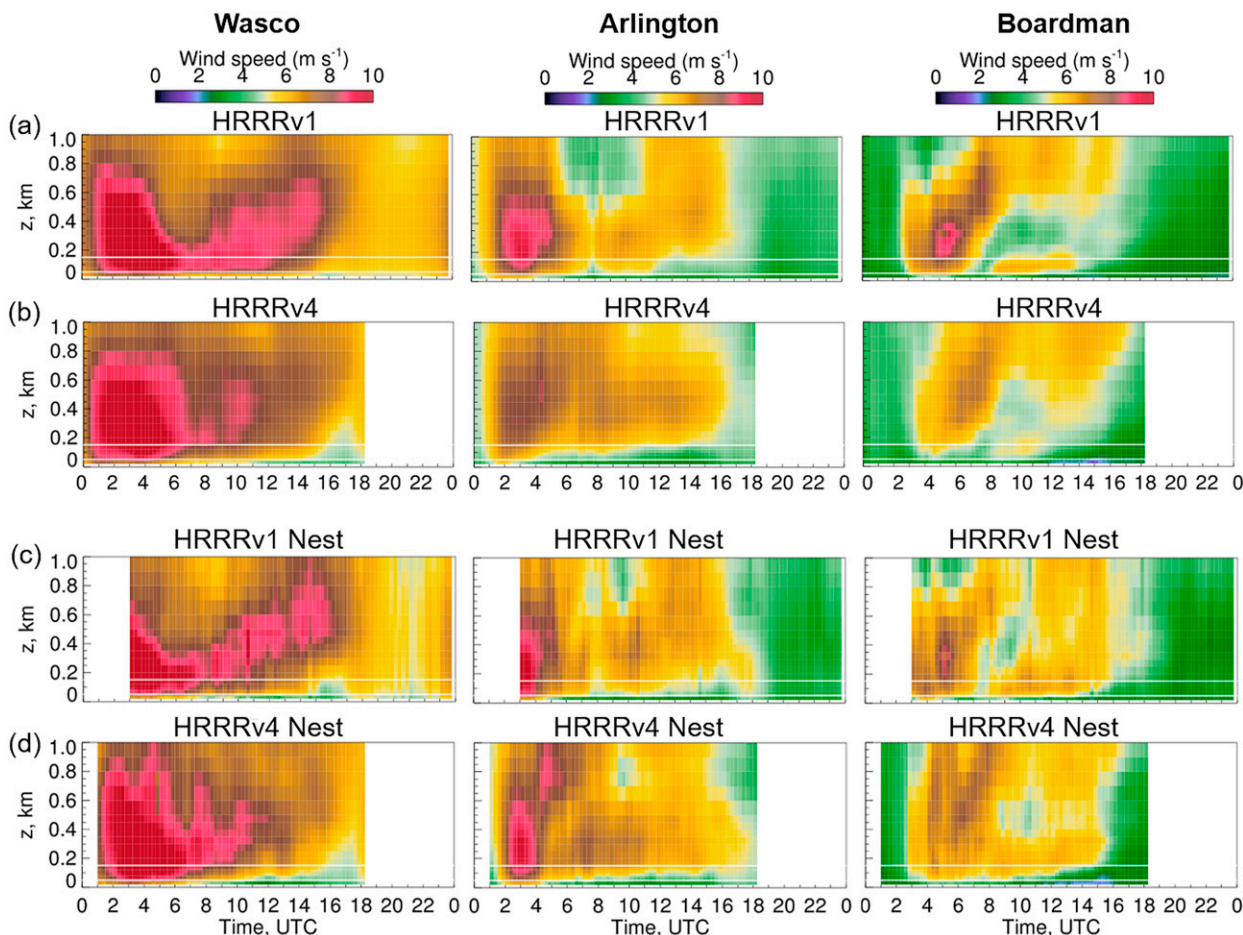


FIG. 7. Period-mean wind speed at three sites modeled by (a),(b) HRRRv1 and HRRRv4 runs in parent domains with the horizontal grid resolution of 3 km, and (c),(d) both models run in the nested domains with the horizontal grid resolution of 750 m. All models are shown for the initial time 0000 UTC.

Reforecast runs of HRRRv1 are available for 0000 and 1200 UTC initial times and the retrospective runs of HRRRv4 are available initialized every third hour. In this paper, wind speeds from all models were taken for model runs initialized at 0000 UTC (Fig. 7). As described, wind speed output for the nest runs was unavailable for the first hour (-v4) or 3-h (-v1) because of initialization spinup issues. Due to missing 0000 UTC retrospective (-v4) runs on 14 August, the period of 15–23 August is used for further analysis.

Figure 7 shows the vertical and diurnal variability among winds simulated by the different model versions at each site through the first 1 km AGL. Simulated 80-m winds from all models (Fig. 8) also varied during the overlapping (0300–1800 UTC) period, HRRRv4 tending to have weaker wind speeds than HRRRv1, sometimes by more than 1 m s^{-1} , as illustrated by the trends (Fig. 8a) and 15-h mean values (Fig. 8b). At the Arlington and Boardman sites the nested versions of both models show weaker mean wind speed compared to models in the parent domain. At Wasco, HRRRv1 simulated stronger speeds—by as much as $2\text{--}3 \text{ m s}^{-1}$ —compared both to HRRRv4 and to the two nested models. The wind speed

differences between models due to horizontal grid resolution, physics, or both are shown in Table 6.

5. Results: Modeled versus measured wind speed

Wind speeds from HRRRv1 and HRRRv4 in both domains were extracted at the Doppler lidar locations by using bilinear interpolation from model horizontal grid points (Fig. 9). Small distances ($<100 \text{ m}$) between collocated instruments at each site allowed the use of these model outputs for comparison with data from other instruments at the site. Previously Pichugina et al. (2020) showed that the bilinear interpolation technique and the nearest-gridpoint technique gave very similar results for each lidar location, with a correlation coefficient of 0.99 between the two extraction techniques and a difference in mean wind speed of $0.01\text{--}0.22 \text{ m s}^{-1}$. Further, the modeled values, which had vertical steps of 20 m, were linearly interpolated to the heights of measurement from each sensor (Table 1).

Measurements having a higher temporal resolution (Table 1) were averaged over 15-min intervals to fit that of the model output. The evaluation metrics, period-mean modeled and

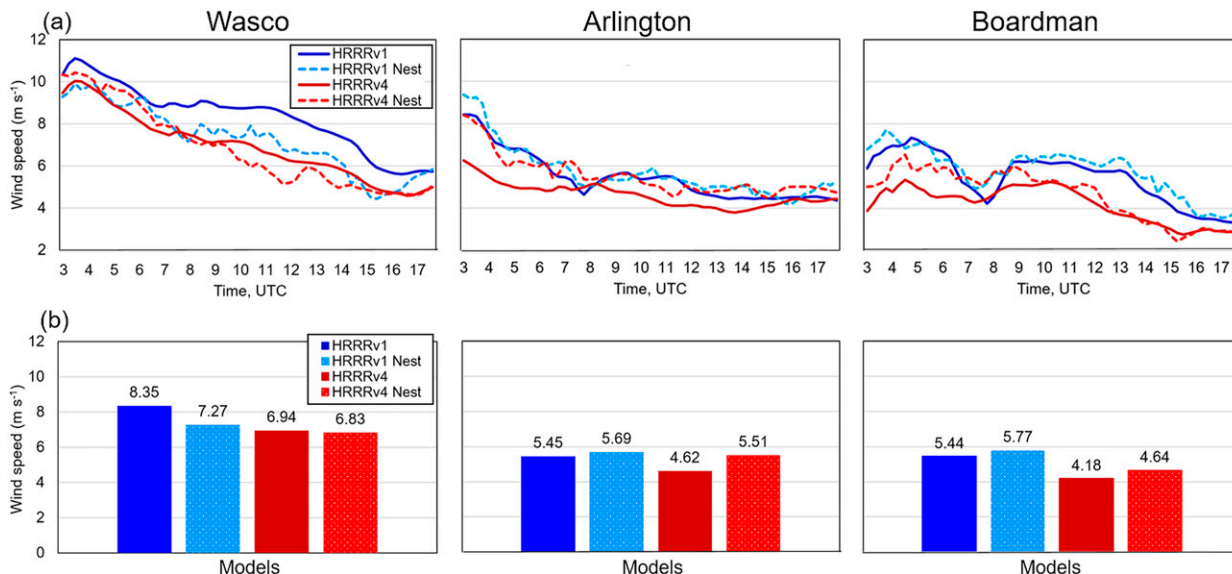


FIG. 8. (a) Temporal variability of 80-m winds from HRRRv1 and HRRRv4 runs in parent and nested domains are shown by colors according to the color scale at the top-left panel. (b) Mean values over 0300–1800 UTC for all models at three sites are shown by colors as indicated in the legend.

measured wind speeds, bias, and MAE, which now refers to model error, were analyzed as time–height cross sections up to 250 m, slightly above the lowest maximum-measurement height (200–220 m) of all instruments; previous error analyses using only scanning lidar data gave these kinds of cross sections up to 1 km AGL, so the cross sections here would represent only the lowest portion of those cross sections. Also shown here are time series of the 80-m wind speeds, the height saved in the model output. Data from the WPRs at Wasco and Boardman were excluded from this analysis due to their coarser vertical and temporal resolutions (Table 1), and their minimum-measurement height. Terrain complexity, expressed as the standard deviation of the terrain elevations (SDE) within a 3-km radius of each instrument location (Ascione et al. 2008), is largest for the Arlington site (71.09 m) followed by 26.33 m for Wasco, and 18.62 m at Boardman.

a. Time–height analysis

Period-mean wind speeds from all instruments and models (Fig. 10) are illustrated for the Wasco site, the site where a scanning lidar (200S), wind profiling lidar (WindCubev1), and sodar were collocated (Table 1). Again, only the most reliable data from the Wasco and Boardman sodars are used for the

analysis, producing differences among the cross sections. Similarly, wind speeds from the models were used only for the times of valid measurements from each instrument, thus also producing instrument-to-instrument differences in the appearances of the various model cross sections. All instruments showed the diurnal wind cycle, and models captured the stronger nighttime values through the layer of measurements. Differences in timing of onset and the intensity of the decelerations of the strong nocturnal flows were examples where the model values departed from the measurements.

Previous WFIP2 studies have used time–height analyses of error properties from earlier versions of HRRR for the understanding of the meteorological context of significant model errors and comparison of model versions to see how physics updates or reductions in grid interval affected these errors, (Pichugina et al. 2019, 2020; Bianco et al. 2019; Olson et al. 2019; Banta et al. 2020, 2021). Figure 11 shows an example of this kind of analysis using the more recent HRRR-version-4, illustrating the differences in MAE due to site and instrument selection, and for each site-instrument combination, the effects of reducing the grid interval to 750 m. Patterns in MAE were similar from instrument to instrument for the corresponding analyses, although those for sodar appeared somewhat

TABLE 6. The 15-h mean wind speed differences between models of different horizontal grid resolutions and parameterization schemes (physics), using data from Fig. 8.

Wind speed differences associated with		WAS	ARL	BOR
1	Horizontal grid resolution (HRRRv1 – HRRRv1 Nest)	1.08	–0.24	–0.33
2	Horizontal grid resolution (HRRRv4 – HRRRv4 Nest)	0.11	–0.89	–0.46
3	Physics (HRRRv1 – HRRRv4)	1.41	0.83	1.26
4	Physics (HRRRv1 Nest – HRRRv4 Nest)	0.44	0.18	1.13
5	Physics and Horizontal grid resolution (HRRRv1 – HRRRv4 Nest)	1.52	–0.06	0.8

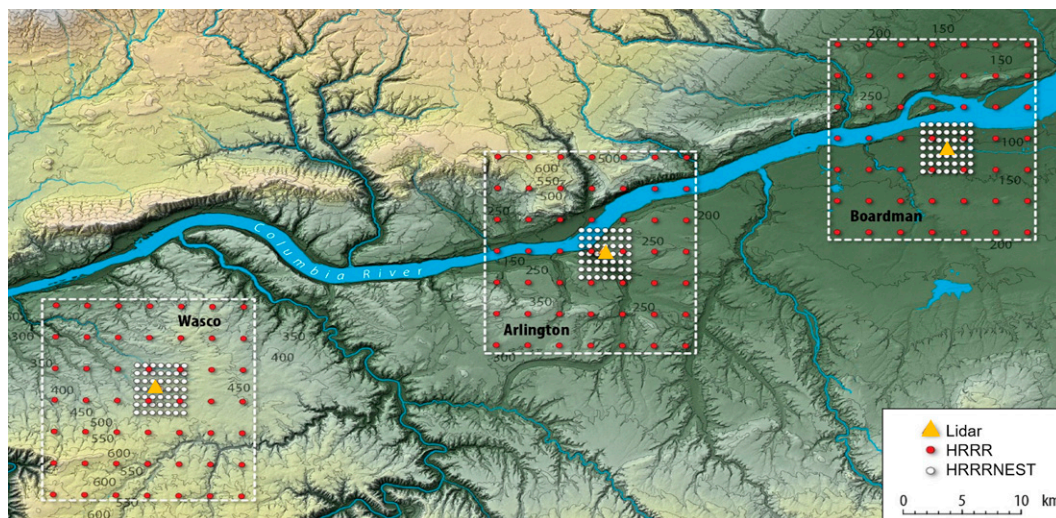


FIG. 9. Location of scanning Doppler lidars at the research sites Wasco, Arlington, and Boardman along the Columbia River Gorge relative to the model grid corners of (red circles) HRRR with a 3-km horizontal grid resolution and (white circles) HRRRNEST with a 750-m horizontal grid resolution.

more patchy and noisier. Errors at Wasco were mostly less than 2 m s^{-1} , irrespective of the instrument. MAEs at Arlington and Boardman appear larger than at Wasco, for many hours greater than 3 m s^{-1} . At Arlington, a large value of MAE occurred at ~ 0600 UTC followed by errors of $>3 \text{ m s}^{-1}$ until the morning transition, due largely to the premature decelerations of the nocturnal flow noted in other studies. In general, the MAE patterns for the nested versions of HRRRv4 indicated smaller MAE than the parent versions.

Figure 12 shows the error difference (Δ -MAE) between MAE of the HRRRv4 in the nested and the parent domain ($\text{MAE HRRRv4 Nest} - \text{MAE HRRRv4}$). Negative values indicate smaller errors for the higher-resolution (750-m) runs. The nested-domain runs show small-magnitude differences of both signs ($\pm 1 \text{ m s}^{-1}$) from the parent-model runs at Wasco, a general tendency for improvement in the 750-m runs at Arlington (especially above 50 m AGL), and short periods of larger magnitude differences ($\pm 2 \text{ m s}^{-1}$) at Arlington and Boardman. Of interest here is that the *patterns* of improvement/degradation appeared similar from instrument to instrument, although the magnitudes were often different.

b. Time–height model comparison

The comparison of HRRRv1 with the more recent version (HRRRv4) is summarized in Fig. 13 using wind speed MAE between each model and scanning lidar data. Similar results were found for other collocated instruments at each site since they showed mostly comparable MAE values (Fig. 12).

1) MODEL HORIZONTAL GRID RESOLUTION

Changes in model MAE (Δ -MAE) due to finer grid resolution are depicted in the top two rows (Fig. 13a for HRRRv4 and Fig. 13b for HRRRv1). As noted above for HRRRv4, finer resolution improved skill at Arlington but not consistently

at Wasco; the changes also produce inconsistent results for HRRRv1.

2) UPDATES IN MODEL PHYSICS

The third and fourth rows (Figs. 13c,d) show Δ -MAE between HRRRv4 and HRRRv1 for parent and nested models due to the *model physics* updates in the latest model version. Negative values here indicate smaller errors (“improvement”) for the updated HRRRv4.

The plots show the HRRRv4 errors compared to the HRRRv1, indicating smaller errors (improvement) for HRRRv4 above 150 m AGL, as found previously by Olson et al. (2019) and Pichugina et al. (2020) for the HRRR-EXPR comparisons. Skill improvement was also indicated below 150 m at Wasco after 0900 UTC, before which the changes were small in magnitude. At Arlington, some degradation in model skill (reaching $0.6\text{--}0.8 \text{ m s}^{-1}$) below 150 m is consistent with the findings of Pichugina et al. (2020) when comparing the HRRR-EXPR version against HRRRv1 (CNTR), and the largest degradation due to model physics updates occurred between 0300 and 0500 UTC probably associated with the evening transition. For the nested versions, Version-4 physics updates produced periods of improvement and periods of degradation, but a consistent diurnal pattern is not apparent.

3) OVERALL Δ -MAE

Changes in model MAE due to updates in both model physics and horizontal grid resolution are illustrated in Fig. 13e. Several periods of significant skill improvement can be seen at each site, but the timing was not consistent from site to site. Early evening degradation at 0500 UTC can be seen at each site, and a modest nighttime degradation of $<0.8 \text{ m s}^{-1}$ can be seen below 100 m at Arlington.

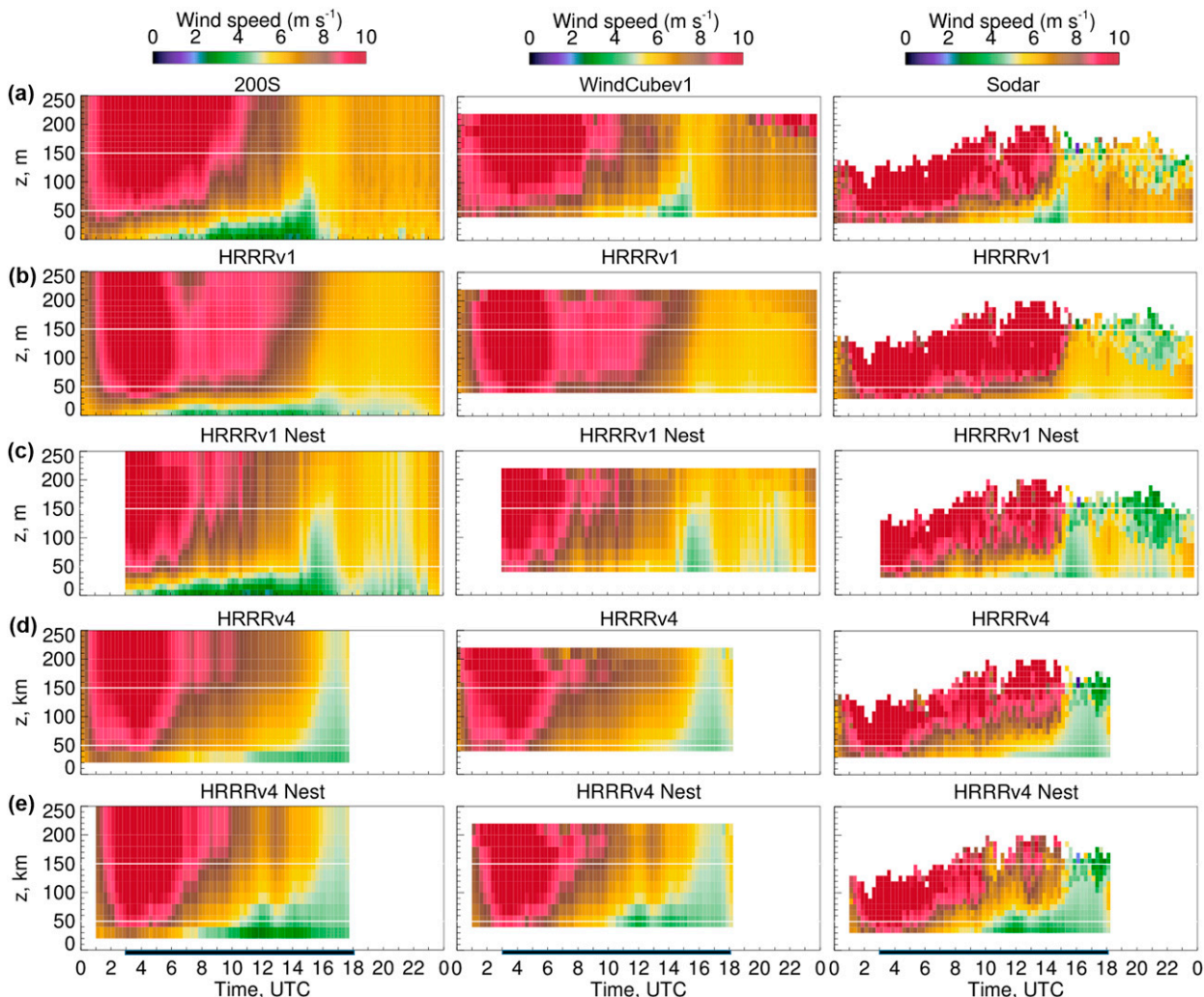


FIG. 10. Mean wind speed over 15–23 Aug at the Wasco site: (a) measurements from 200S, WindCubev1, and sodar, from left to right, respectively; (b),(c) HRRRv1 reforecasts and (d),(e) HRRRv4 retrospective runs in the parent (3-km horizontal grid) and nested (750-m horizontal grid) domains. All models are shown for the initial time 0000 UTC. Black lines at the bottom of the plot indicate the overlapping time (0300–1800 UTC) for all four models. A significant amount of sodar data above 80 m are flagged out at both sites, especially during the evening and morning transition hours, and only reliable data at Wasco and Boardman (flag values = 9) were used in the paper. White horizontal lines show the vertical layer of 50–150 m, which is important for the operations of modern wind turbines.

The color-scale time–height analysis in this section provides an overview of the variability of model MAEs and Δ -MAEs with height. This type of analysis is intended to further characterize model errors, suggesting where a deeper investigation is likely warranted.

c. 80-m winds

Knowledge of the forecast error for wind speed at the hub height of wind turbines is of critical economic importance for the calculation of energy produced by the wind resource. This section focuses on the evaluation of model skill in forecasting wind speed at 80 m AGL, at or near the hub height of most wind turbines in the surrounding WFIP2 area.

MEASUREMENTS OF 80-M WIND SPEED VERSUS HRRRv1 SIMULATIONS

Scatter diagrams (Fig. 14) of simulated 80-m winds for 15–23 August are plotted against the measured winds by each instrument, showing the significant spread of winds over the range of 0–20 m s⁻¹ for the 9 days.

The sample is shown for HRRRv1 and HRRRv1 Nest models to better cover the diurnal cycle, with 0000 UTC forecasts for 0000–2400 and 0300–2400 UTC, respectively. Both models underpredicted weak (<5 m s⁻¹) and overpredicted stronger (>12 m s⁻¹) speeds at the Wasco and Boardman sites. At the Arlington site, both models underpredicted all speeds.

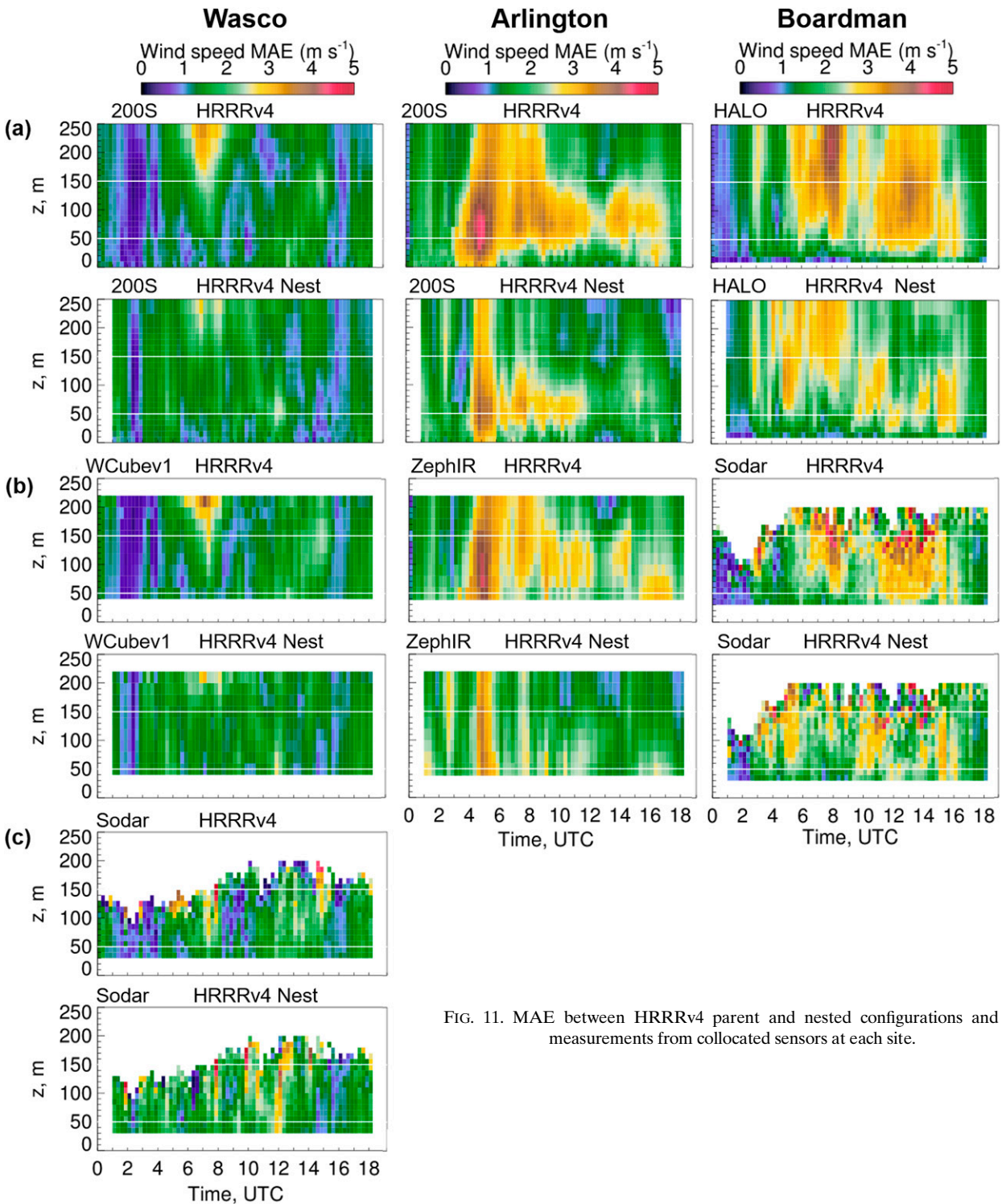


FIG. 11. MAE between HRRRv4 parent and nested configurations and measurements from collocated sensors at each site.

Table 7 summarizes the differences in the diurnally averaged wind speeds between HRRRv1 and each instrument. These differences are (-0.2 m s^{-1}) at the Wasco site (except $+0.3$ for sodar data), -0.6 m s^{-1} at the Boardman site, and even larger ($+1.5 \text{ m s}^{-1}$) at the Arlington site. It is

noteworthy that the model values at Wasco lie between the consensus lidar values and the sodar values, such that the magnitude of the model errors are similar for each instrument type, but the signs are opposite. Compared to the parent HRRRv1, the wind speed differences between instruments

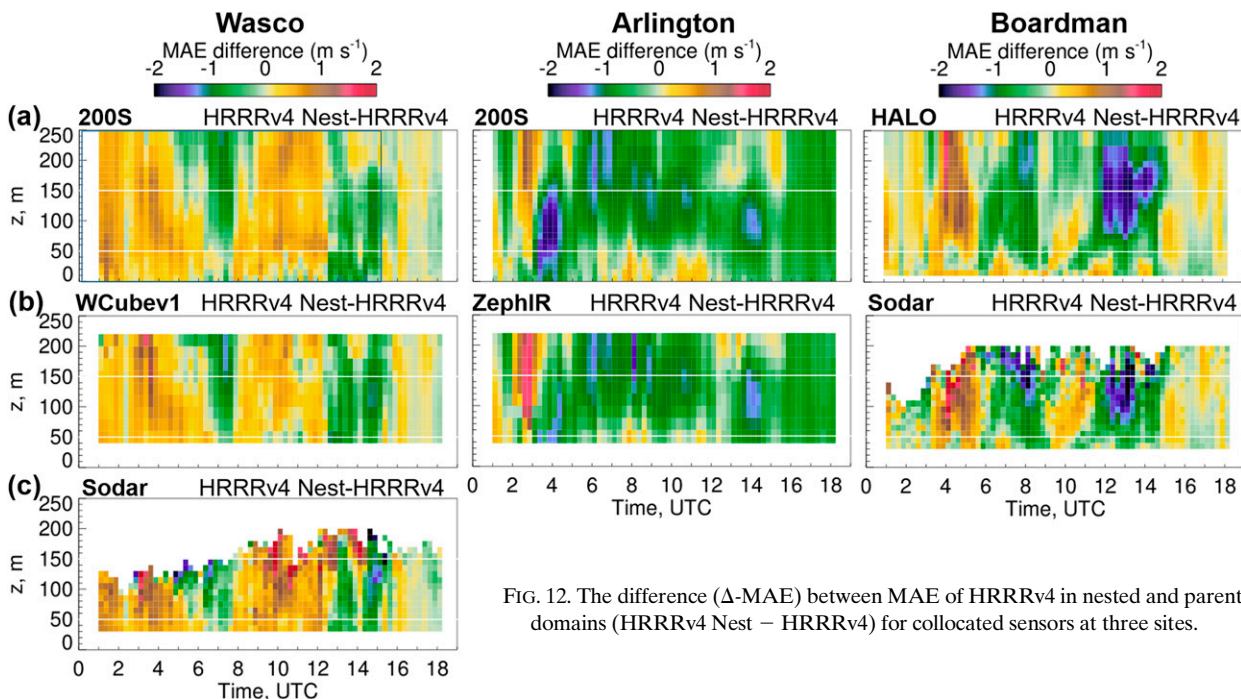


FIG. 12. The difference (Δ -MAE) between MAE of HRRRv4 in nested and parent domains (HRRRv4 Nest – HRRRv4) for collocated sensors at three sites.

and HRRRv1 Nest are of similar magnitude but the opposite sign at Wasco (except sodar data) and significantly smaller ($\sim 0.3 \text{ m s}^{-1}$) at Boardman. The large differences are still apparent at Arlington. Close values of wind speed between measurements and HRRRv1 Nest simulations can be explained by the shorter nighttime period (the HRRRv1 Nest 0000 UTC runs started from 0300 UTC) and finer horizontal grid resolution, which is important for the more complex terrain at the Arlington site.

Period-averaged wind speeds from scanning lidars are generally within $\sim 0.1 \text{ m s}^{-1}$ of the other sensors (except sodar at Wasco and the parent-version sample at Arlington). Correlation coefficients (Fig. 15) show good agreement between all models and all types of instruments. The correlation coefficient between measurements and all models is slightly larger at the Wasco and Arlington sites compared to the Boardman site. The high R^2 for scanning Doppler lidars at all 3 sites indicates a small impact of the scanning lidar measurement footprint.

Not much difference in mean wind speed values and correlation was found for the averaging over a smaller number of points (0300–1800 UTC) as shown in the parenthesis (Table 7).

d. Time series of 80-m wind speed from measurements and all models

Figure 16a shows the time series of period-mean 80-m wind speed from collocated sensors at each site. All sensors show similar diurnal variability and values mostly within 0.4 m s^{-1} of each other at each site, except for the nocturnal winds from the Wasco sodar. At Arlington, the ZephIR wind speeds were systematically stronger than the 200S during daytime and

weaker at night by as much as 0.4 m s^{-1} , consistent with the annually averaged behavior (Fig. 2).

Figure 16 shows the time series of (Fig. 16a) mean wind speed from each instrument and MAE for the HRRRv1 (Figs. 16b,c) and HRRRv4 (Figs. 16d,e) model runs for the parent and nested domains. At Wasco, the significant differences in wind measurements between the lidars and the sodar at night do not lead to significant differences in calculated model MAE in general, because the model-predicted winds lay between the two measurement values, as described in the previous section. At Arlington, the weaker nighttime winds measured by the ZephIR lidar are in better agreement with the model predictions than the 200S lidar, as indicated by the 0600–1600 UTC ZephIR-measured MAE being $\sim 0.4 \text{ m s}^{-1}$ smaller in general for each model version. Good agreement between sodar and HALO-lidar wind measurements at Boardman means the calculated MAEs for each, which ranged mostly between 2 and 3 m s^{-1} , showed similar values generally agreeing to within 0.1 m s^{-1} overall. Larger MAE values (up to 4 m s^{-1}) were found for HRRRv1 for short periods at Wasco (1400–1500 UTC) and Boardman (0500–0700 UTC).

The smaller calculated MAEs at night at Arlington, when using the ZephIR wind speed values as a reference, illustrate a potential consequence of measurement bias in calculating model-error statistics. The smaller MAEs resulted from systematic differences in measured wind speed between the ZephIR and 200S lidars. If we hypothesized for purposes of argument that the scanning 200S lidar was providing accurate measurements of the wind, and the ZephIR winds were exhibiting a nocturnal measurement bias, then that would mean that the errors calculated from the ZephIR winds

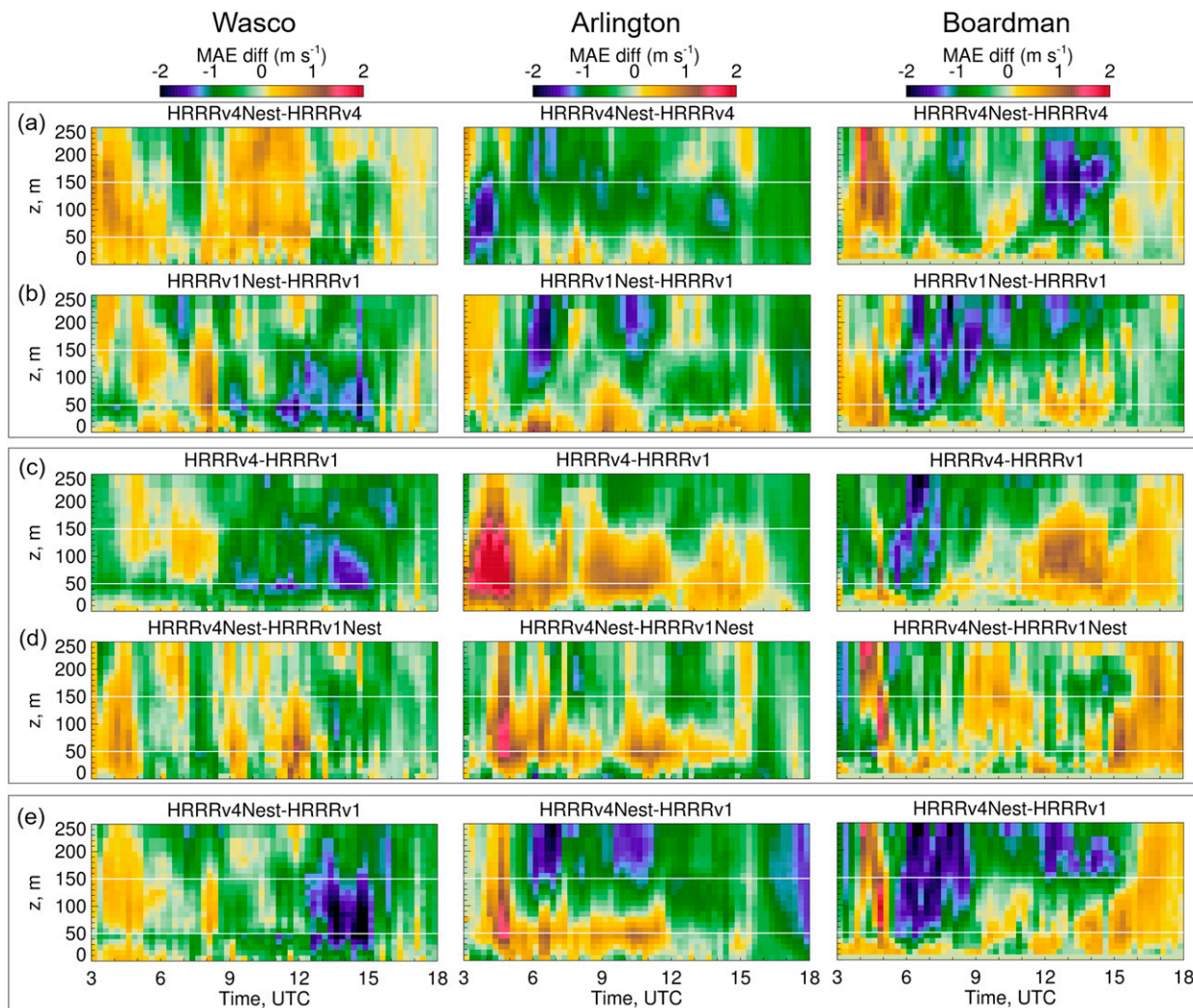


FIG. 13. MAE difference between scanning lidar measurements and four model configurations: (a) MAE HRRRv4 Nest – MAE HRRRv4, (b) MAE HRRRv1 Nest – MAE HRRRv4, (c) MAE HRRRv4 – MAE HRRRv1, (d) MAE HRRRv4 Nest – MAE HRRRv1 Nest, and (e) MAE HRRRv4 Nest – HRRRv1.

were erroneously small, because the true model errors (by hypothesis, those calculated as related to the winds from 200S lidar) were being partially offset by the measurement bias. Measurement accuracy—and in particular small measurement bias—is critical to useful model-error calculation.

Summary plots (Fig. 17a) of mean MAE values for the overlapping period of 03–18 UTC error differences between models and between sites tend to be somewhat larger than the error differences between the instruments used. The results are shown for the initial time 00 UTC since the analysis of the errors for different model forecast lead times is outside the scope of this study.

The values of these MAE differences (Δ -MAE) for all instruments and models at each site (Fig. 17b) generally indicate overall improvement (reduced errors, negative values), in some cases more than 0.5 m s^{-1} . Model grid-spacing effects for all instruments (light and dark gold colors) show the largest

improvements for both models at Arlington ($\sim 0.75 \text{ m s}^{-1}$) and more modest improvements at Boardman, but a small (0.03 – 0.13 m s^{-1}) degradation for the HRRRv4 Nest at Wasco.

The Δ -MAEs due to HRRRv4 physics updates (light-blue bars) show improvements to the parent 3-km HRRR model of 0.6 m s^{-1} at Wasco, but degradations at Arlington, with values of $+0.28$ and $+0.45 \text{ m s}^{-1}$ for the ZephIR and 200S lidars, respectively. At Boardman, these updates produced little change, but mostly small improvements. The nested versions (dark-blue bars) showed only minor changes in model skill due to model physics updates, amounting to 0.1 m s^{-1} or less and generally tending to conform to the sign of the Δ -MAE for the parent versions.

In general, the differences due to measurement by different instruments did not lead to contradictory conclusions about whether a given set of changes to the model improves or degrades skill; i.e., the signs of the Δ -MAEs are the same for all instruments at each site for each version-pair comparison,

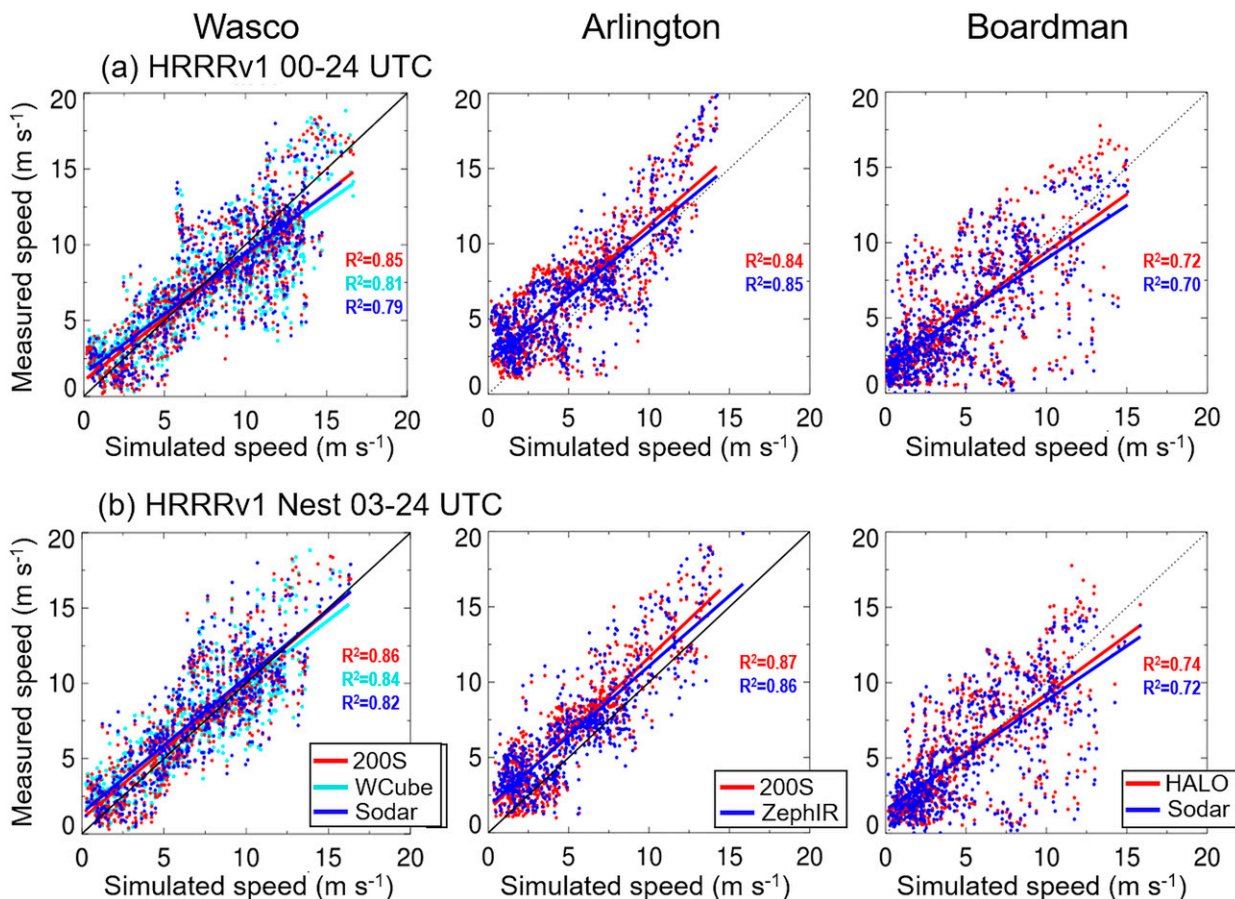


FIG. 14. Scatterplots of 80-m measured wind speed vs simulations from (a) HRRRv1 and (b) HRRRv1 Nest at three sites. Dots represent all days in the 15–23 Aug period and are shown by colors for each site instrument according to color scales at the bottom panels. The color solid lines indicate a linear fit and correlation coefficient values. The black solid lines indicate the best fit.

although the magnitudes may be quite different. However, for the physics comparison of the nested versions (dark blue: HRRRv4 Nest versus HRRRv1 Nest) at Boardman, the sodar shows that the model updates improved skill, but the HALO lidar shows a degradation in skill (increase in model error). Similarly at Wasco, the lidars led to different conclusions from the sodar. Although the differences are small, these examples illustrate that conclusions about which version generates smaller errors can depend upon which instrument is used as a reference value for calculating model error.

These results indicate that in general, the evaluation of the model-resolution and model-physics impacts was not strongly dependent on the instrument used but was often site-dependent for this study period. For example, the degraded skill of some models such as HRRRv4 was noted at the Arlington site, located in the most complex terrain of the three research sites. The site-dependent nature of model errors was previously pointed out by Pichugina et al. (2019), Bianco et al. (2019), Banta et al. (2020), and Pichugina et al. (2020) noted that changes in model skill resulting from updated model physics and increased resolution also varied from site to site.

6. Conclusions

For this study, we had three sites where more than one type of remote sensing measurement system provided wind profiles over periods of several months. Variations were noted among the mean-wind values from the sensors at each location, these variations differed by month, time of day, and measurement location, even though the instruments were separated by less than 100 m. Each site had a different mix of instruments, but all three sites had a scanning Doppler lidar, so these lidars were used as a reference against which the other measured values were compared. Differences in measured wind speed were typically 0.1–0.2 m s⁻¹, but at times exceeded 0.4 m s⁻¹.

These measurement systems have been used in previous studies where model errors were calculated to evaluate model skill above Earth's surface. Because the measured values of wind speed and direction vary from instrument to instrument, an important question is, what is the impact of this variability on the calculated model errors, and on the comparison of error statistics from one model or model version to another—for example, to see whether a version having updated model physics produces forecasts having smaller errors and thus improved

TABLE 7. Mean wind speed statistics from scatterplots (Fig. 14) between collocated sensors at each site and four model configurations. Data averaged over 0000 UTC forecasts from each model and (in parentheses) over 0300–1800 UTC, a common period for all models.

Site	Sensor	Measured		Modeled		$Y = A + Bx$		Correlation	
		Mean	STD	Mean	STD	A	B	R^2	Count
HRRRv1 0000–2400 UTC (0300–1800 UTC)									
WAS	200S	7.44	3.99	7.66	4.10	1.03	0.83	0.85 (0.82)	864 (743)
	WCube	7.45	3.73	7.66	4.10	1.72	0.74	0.81 (0.78)	864 (743)
	Sodar	7.96	3.84	7.66	4.10	1.48	0.79	0.80 (0.79)	743 (743)
ARL	200S	6.76	3.97	5.19	3.58	1.93	0.93	0.85 (0.83)	858 (855)
	ZephIR	6.54	3.75	5.19	3.58	2.02	0.87	0.87 (0.83)	858 (855)
BOR	HALO	5.21	3.89	4.57	3.64	1.70	0.76	0.72 (0.73)	851 (810)
	Sodar	5.13	3.68	4.57	3.634	1.76	0.70	0.70 (0.71)	823 (810)
HRRRv1 Nest 0300–2400 UTC (0300–1800 UTC)									
WAS	200S	7.57	3.85	7.39	3.48	0.99	0.92	0.86 (0.86)	753 (650)
	WCube	7.58	3.58	7.39	3.48	1.55	0.84	0.84 (0.82)	744 (650)
	Sodar	7.78	3.79	7.39	3.48	1.42	0.89	0.82 (0.83)	658 (650)
ARL	200S	6.61	3.95	5.15	3.37	1.29	1.04	0.87 (0.87)	748 (748)
	ZephIR	6.55	3.67	5.15	3.37	1.75	0.93	0.86 (0.86)	748 (748)
BOR	HALO	5.39	3.79	5.06	3.58	1.44	0.78	0.74 (0.74)	740 (740)
	Sodar	5.24	3.60	5.06	3.58	1.57	0.72	0.72 (0.72)	740 (740)

skill. In this study, we have quantified the impact of instrument measurement variability during WFIP2 on the magnitude of calculated errors and the change in error from one model version to another. The model versions tested involved updates in model physics from HRRR-v1 to HRRRv-4, and reductions in grid interval from 3 km to 750 m.

In general, model errors were found to be 2–3 m s^{-1} , in agreement with previous WFIP2 studies. Differences in the errors as determined by the various instruments at each site amounted to

about 10% of this value, or 0.2–0.3 m s^{-1} . The magnitudes of the changes in model skill due to physics or grid-resolution updates also differed depending on which measurement was used to determine the errors, where most of the instrument-to-instrument differences were $\sim 0.1 \text{ m s}^{-1}$, but some were as large as 0.3 m s^{-1} . In most cases, all instruments at a given site showed consistency in the sign of the change (Δ -MAE) in error, but two examples were found where the sign changed. Although the magnitude of the changes was small, these two examples

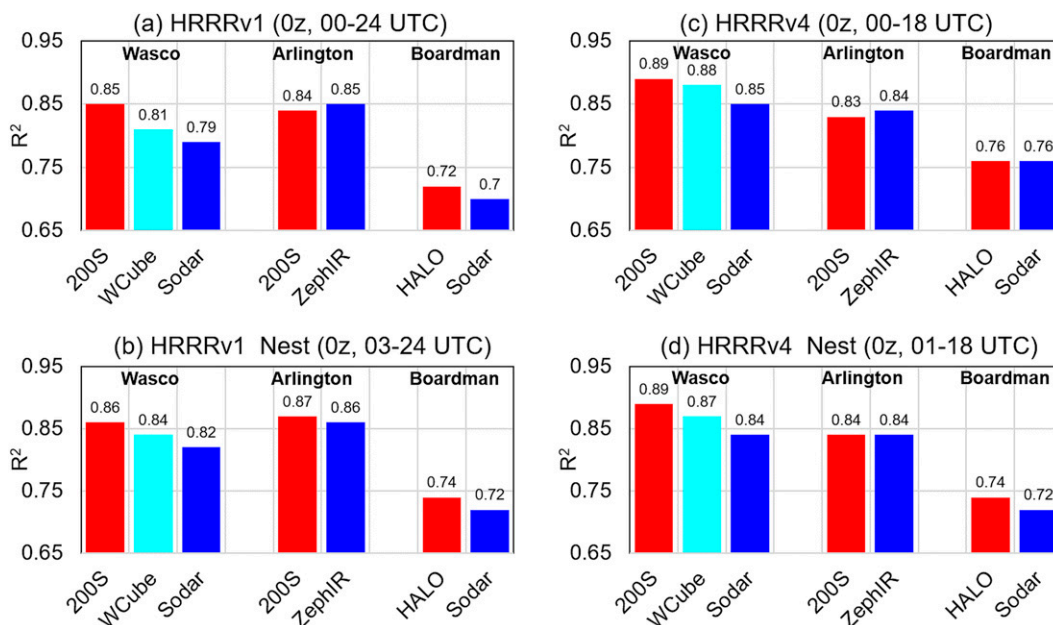


FIG. 15. Period-mean 80-m correlation coefficients between data from collocated sensors and four models are shown for three sites. The name of the model and the average period is indicated at the top of each panel.

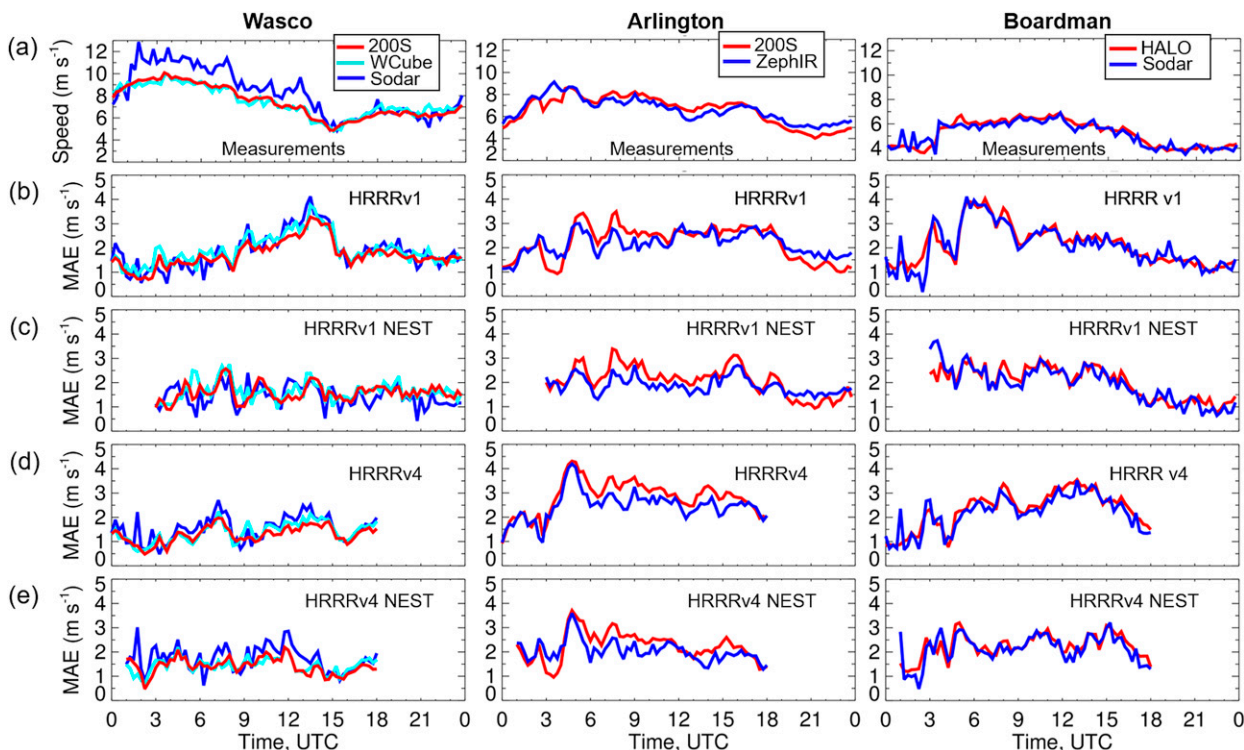


FIG. 16. Time series of 80-m mean (a) wind speed; (b),(c) MAE between measurements and HRRRv1 in parent and nested domains; and (d),(e) MAE between measurements and HRRRv4 in parent and nested domains are shown for collocated sensors at each site. Data from each instrument and the corresponding MAE are colored according to the color scales at the top panels. The MAEs for all four models are shown for the initial time 0000 UTC.

illustrate that a consequence of the differences in measurements is the possibility that errors determined by using one instrument may show improvement in model skill, whereas errors determined for values measured by another instrument may indicate degradation. This possibility underscores the importance of having accurate measurements to determine the model error.

Improvement of NWP models of all scales will continue to be an important endeavor. Improving models consists of creating a new version of a model by updating model physics routines, numerics, initialization schemes, or other aspects, then running both versions and comparing their output against available measurements. Model skill is deemed to have improved if the new version agrees better with the measurements than the old version, i.e., if the model-measurement differences (errors) are smaller for the new version. Using measurements for model improvement is thus critical to advancing model skill, but uncertainties in those measurements and how they might affect the outcome of validation results have not been significant areas of concern as important aspects of advancing model skill.

It is unusual to have collocated sensors at each of several spatially separated sites, to be able to see how differences in measurement uncertainties among instruments can affect model-error evaluation. As just summarized, it is clear from our results here that discrepancies exist among wind values measured by different instruments. As a result, from a given model run, the errors calculated for one instrument will differ

from those from another instrument. Calculated error based on one instrument will thus be less than error based on the other instrument, and this could affect significance testing, in that the smaller “measured” error may pass a null-hypothesis test, whereas the larger one may fail it, obviously leading to different conclusions about whether there is a statistically significant difference between the two model results. Even more concerning, if the model results fall between the data from two instruments, that means that one instrument would show “improvement,” and the other would show “degradation.”

Here we have used a dataset of opportunity from WFIP2 to demonstrate that these kinds of measurement issues exist and should be of concern, not only to measurement specialists but to those who would use measurements to determine relative skill among model versions as a part of model improvement efforts. More careful and detailed field studies against known reference measurements, such as tower-based, should be performed in a variety of times and locations to determine by which instrumentation and under what conditions sufficiently accurate values can be provided to use for model evaluation between two model versions. In absence of such clarity, model improvement studies are in danger of making faulty conclusions about the effects of model updates from one model version to the next.

Acknowledgments. The authors thank the WFIP2-experiment participants who aided in the deployment and the collection of lidar data. A special thanks go to our colleagues

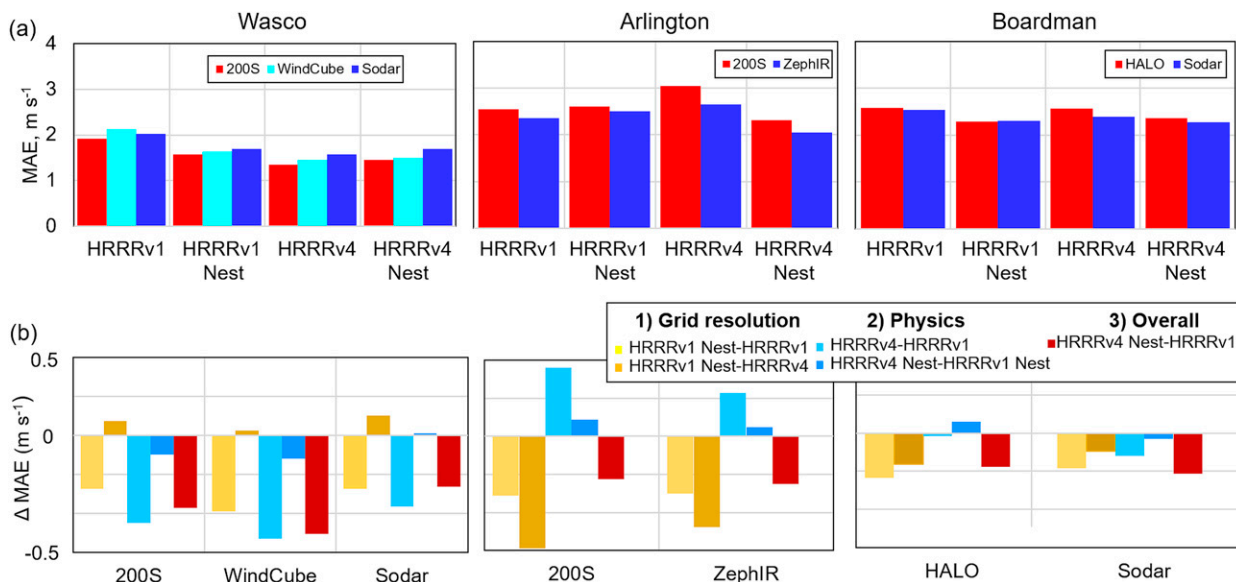


FIG. 17. Period-mean of 80-m wind speed (a) MAE between measurements from collocated instruments at each site and simulated by HRRRv1 and HRRRv4 run in the parent and nested domains and (b) Δ MAE for different model configurations. Colors indicate improvements (negative values) in HRRRv4 models made during WFIP2 due to the better horizontal grid resolution, parameterization scheme (physics), and both (overall).

Scott Sandberg and Ann Weickmann from NOAA/CSL and C. Hocut, Army Research Laboratory, for their tremendous work in calibrating and deploying lidars to the research sites. From NOAA/PSL we thank Clark King for communications with landowners and obtaining site licenses and Aditya Choukulkar (CSL) for the deployment of lidars, monitoring lidar data in real-time, and providing quality controlled lidar data to the data archive portal (DAP). We thank Joel Cline (DOE), Melinda Marquis (NOAA), and Jim McCaa (Vaisala) for their effort to propose, design, and lead the WFIP2. We appreciate the efforts of Jim Bickford, Clara St. Martin, Joseph Lee, and Rochelle Worsnop for the deployment of the WindCube v1 profiling lidar. This work was sponsored by the U.S. Department of Energy Wind Energy Technologies Office, and by the NOAA Atmospheric Science for Renewable Energy Program. The views expressed in the article do not necessarily represent the views of NOAA, DOE, or the U.S. government.

Data availability statement. The information on instrument metadata such as location, dates of deployment, data-processing methods including time averaging, whether the data were transferred to the MADIS and whether the data were assimilated in real-time into developmental versions of the Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) Models run at NOAA/ESRL can be found at https://madis.noaa.gov/support_overview.shtml. The data that support the findings of this study are openly available in Data Archive and Portal (DAP), <https://a2e.energy.gov/data>. The DAP establishes a sustained data management structure with protocols and access to assure massive datasets resulting from A2e (Atmosphere to Electrons) efforts will have the quality needed for scientific discovery and portals required to make data available to a broad

stakeholder group. The DAP will collect, store, catalog, process, preserve, and disseminate all significant A2e data—and ultimately all historical wind data supported by the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy—with state-of-the-art technology while conforming to or defining new industry data standards. Real-time and quality-controlled data from scanning Doppler lidars are available on DAP and per request from the authors. During WFIP2 two scanning, pulsed Leosphere WindCube 200S Doppler lidar systems continuously from September 2015 to April 2017, and the third scanning Doppler lidar, a Halo Streamline XR continuously operated from January to December 2016. In addition to DAP, real-time lidar measurements, as well as data processing products, can be found at <https://www.esrl.noaa.gov/csd/groups/csd3/measurements/wfip2/>. Power generation data are available on request due to privacy/ethical restrictions. The data that support the findings of this study are available from the Bonneville Power Authority (BPA) balancing area. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors upon reasonable request and with the permission of BPA. Data on wind power generation within the BPA can be found here: <https://transmission.bpa.gov/Business/Operations/Wind/twndbspt.aspx>.

APPENDIX

Model and Physics Configuration for Several Versions of HRRR Developed during the WFIP2

In Table A1 we present model and physics configuration for several versions of HRRR developed during the WFIP2 (James et al. 2022).

TABLE A1. Model and physics configuration for HRRRv1 – HRRRv4 (James et al. 2022).

System	HRRRv1	HRRRv2	HRRRv3	HRRRv4
Model	WRF-ARWv3.4.1+	WRF-ARWv3.6+	WRF-ARWv3.8.1+	WRF-ARWv3.9.1+
Domain	CONUS	CONUS	CONUS, Alaska	CONUS, Alaska
Initial frequency	1 h	1 h	1, 3 h	1, 3 h
Map projection	Lambert conformal	Lambert conformal	Lambert conformal (CONUS), polar stereographic (AK)	Lambert conformal (CONUS), polar stereographic (AK)
Grid points (x, y)	1800 × 1060	1800 × 1060	1800 × 1060, 1300 × 920	1800 × 1060, 1300 × 920
Grid spacing	3 km	3 km	3 km	3 km
Vertical layers	51	51	51	51
Pressure top	20 hPa	20 hPa	20 hPa	15 hPa
Lateral boundary conditions	RAP	RAP	RAP	RAP
Initial conditions	RAP post-DFI	RAP post-DFI plus 1-h spinup, 3-km GSI with GDAS	RAP post-DFI plus 1-h spinup, 3-km GSI with GDAS	HRRRDAS mean plus 1-h spinup, 3-km GSI with HRRRDAS
Vertical coordinate	Sigma	Sigma	Hybrid sigma–terrain-following	Hybrid sigma–terrain-following
Horizontal/vertical advection	Fifth-order upwind	Fifth-order upwind	Fifth-order upwind	Fifth-order upwind + IEVA
Scalar advection	Positive definite	Positive definite	Positive definite	Positive definite
Large time step	20 s	20 s	20 s	20 s
Upper-level damping	Rayleigh, dampcoef = 0.2 s ⁻¹ , zdamp = 5000 m	Rayleigh, dampcoef = 0.2 s ⁻¹ , zdamp = 5000 m	Rayleigh, dampcoef = 0.2 s ⁻¹ , zdamp = 5000 m	Rayleigh, dampcoef = 0.2 s ⁻¹ , zdamp = 5000 m
Computational horizontal diffusion	None	Sixth-order (0.25)	Sixth-order (0.25), horizontal only (not on slopes), applied to all variables	Sixth-order reduced to 0.04 for tracers, including water vapor and hydrometeors, and to 0.12 for other model variables
Run frequency	Hourly	Hourly	Hourly, 3h	Hourly, 3h
Forecast duration	15 h	18 h	36 h every 6 h, otherwise 18 h	48 h every 6 h, otherwise 18 h
Radiation	RRTMG	RRTMG	RRTMG	RRTMG (modified with SGS cloud)
Land surface, including No. of layers	RUC LSM, 9 soil levels, 2-layer snow (v3.5+)	RUC LSM, 9 soil levels, 2-layer snow, reduced wilting point (v3.6+)	RUC LSM, 9 soil levels, 2-layer snow (v3.8+)	RUC LSM, 9 soil levels, 2-layer snow (v3.9+)
Land use	30 ^{yr} MODIS	30 ^{yr} MODIS	30 ^{yr} MODIS,	15 ^{yr} MODIS, BNU soil type (via WRF), switch to MODIS albedo
Planetary-boundary and surface layer	Mellor–Yamada–Nakanishi–Niino (v3.5+)	Mellor–Yamada–Nakanishi–Niino (v3.6+)	Mellor–Yamada–Nakanishi–Niino (v3.8+)	Mellor–Yamada–Nakanishi–Niino (v3.9+)
Subgrid-scale (SGS) clouds	None	MYNN RH-based (Benjamin et al. 2016; App. B)	MYNN prognostic SGS cloud fraction, cloud water	MYNN removed limit to SGS cloud water, reduced radii
Orographic drag	None	None	Small-scale orographic drag	Turbulent drag from subgrid orography (CONUS only)
Cloud Microphysics	Thompson (v3.4.1)	Thompson–Eidhammer “aerosol-aware” (v3.6.1)	Thompson–Eidhammer “aerosol-aware” (v3.8)	Thompson–Eidhammer “aerosol-aware” (v3.9)
Microphysics tendency limit	0.07 K s ⁻¹	0.07 K s ⁻¹	0.07 K s ⁻¹	None needed due to IEVA

REFERENCES

- Aitken, M. L., M. E. Rhodes, and J. K. Lundquist, 2012: Performance of a wind-profiling lidar in the region of wind turbine rotor disks. *J. Atmos. Oceanic Technol.*, **29**, 347–355, <https://doi.org/10.1175/JTECH-D-11-00033.1>.
- Ascione, A., A. Cinque, E. Miccadei, F. Villani, and C. Berti, 2008: The Plio-Quaternary uplift of the Apennine chain: New data from the analysis of topography and river valleys in Central Italy. *Geomorphology*, **102**, 105–118, <https://doi.org/10.1016/j.geomorph.2007.07.022>.
- A2E, 2017a: wfp2/lidar.z04.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 19 December 2017, <https://doi.org/10.21947/1418023>.
- , 2017b: wfp2/lidar.z05.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 19 December 2017, <https://doi.org/10.21947/1418024>.
- , 2017c: wfp2/lidar.z07.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 29 March 2018, <https://doi.org/10.21947/1402036>.
- , 2017d: wfp2/radar.z04.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 15 November 2018, <https://doi.org/10.21947/1412526>.
- , 2017e: wfp2/lidar.z06.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 23 November 2018, <https://doi.org/10.21947/1349273>.
- , 2017f: wfp2/sodar.z09.b0. Maintained by A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 26 October 2021, <https://doi.org/10.21947/1356333>.
- , 2017g: wfp2/sodar.16.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 25 March 2019, <https://doi.org/10.21947/1356340>.
- , 2017h: wfp2/radar.z04.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 19 November 2018, <https://doi.org/10.21947/1412526>.
- , 2017i: wfp2/radar.z07.b0. A2e Data Archive and Portal for U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, accessed 19 November 2018, <https://doi.org/10.21947/1412529>.
- Banta, R. M., R. K. Newsom, J. K. Lundquist, Y. L. Pichugina, R. L. Coulter, and L. Mahrt, 2002: Nocturnal low-level jet characteristics over Kansas during CASES-99. *Bound.-Layer Meteorol.*, **105**, 221–252, <https://doi.org/10.1023/A:101992330866>.
- , Y. L. Pichugina, N. D. Kelley, R. M. Hardesty, and W. A. Brewer, 2013: Wind energy meteorology: Insight into wind properties in the turbine-rotor layer of the atmosphere from high-resolution Doppler lidar. *Bull. Amer. Meteor. Soc.*, **94**, 883–902, <https://doi.org/10.1175/BAMS-D-11-00057.1>.
- , and Coauthors, 2020: Characterizing NWP model errors using Doppler-lidar measurements of recurrent regional diurnal flows: Marine-air intrusions into the Columbia-River basin. *Mon. Wea. Rev.*, **148**, 929–953, <https://doi.org/10.1175/MWR-D-19-0188.1>.
- , and Coauthors, 2021: Doppler-lidar evaluation of HRRR-model skill at simulating summertime wind regimes in the Columbia River basin during WFIP2. *Wea. Forecasting*, **36**, 1961–1983, <https://doi.org/10.1175/WAF-D-21-0012.1>.
- 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>.
- Bianco, L., and Coauthors, 2019: Impact of model improvements on 80 m wind speeds during the second Wind Forecast Improvement Project (WFIP2). *Geosci. Model Dev.*, **12**, 4803–4821, <https://doi.org/10.5194/gmd-12-4803-2019>.
- Bingöl, F., J. Mann, and D. Foussekis, 2009: Conically scanning lidar error in complex terrain. *Meteor. Z.*, **18**, 189–195, <https://doi.org/10.1127/0941-2948/2009/0368>.
- Bodini, N., J. K. Lundquist, R. Krishnamurthy, M. Pekour, L. K. Berg, and A. Choukulkar, 2019: Spatial and temporal variability of turbulence dissipation rate in complex terrain. *Atmos. Chem. Phys.*, **19**, 4367–4382, <https://doi.org/10.5194/acp-19-4367-2019>.
- Bonin, T. A., and W. A. Brewer, 2017: Detection of range-folded returns in Doppler lidar observations. *IEEE Geosci. Remote Sens. Lett.*, **14**, 514–518, <https://doi.org/10.1109/LGRS.2017.2652360>.
- , and Coauthors, 2017: Evaluation of turbulence measurement techniques from a single Doppler lidar. *Atmos. Meas. Tech.*, **10**, 3021–3039, <https://doi.org/10.5194/amt-10-3021-2017>.
- Draxl, C., and Coauthors, 2021: Mountain waves impact wind power generation. *Wind Energy Sci.*, **6**, 45–60, <https://doi.org/10.5194/wes-6-45-2021>.
- James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, <https://doi.org/10.1175/WAF-D-21-0130.1>.
- Lundquist, J. K., M. J. Churchfield, S. Lee, and A. Clifton, 2015: Quantifying error of lidar and sodar Doppler beam swinging measurements of wind turbine wakes using computational fluid dynamics. *Atmos. Meas. Tech.*, **8**, 907–920, <https://doi.org/10.5194/amt-8-907-2015>.
- Mass, C. F., M. D. Albright, and D. J. Brees, 1986: The onshore surge of marine air into the Pacific Northwest: A coastal region of complex terrain. *Mon. Wea. Rev.*, **114**, 2602–2627, [https://doi.org/10.1175/1520-0493\(1986\)114<2602:TOSOMA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2602:TOSOMA>2.0.CO;2).
- McCaffrey, K., and Coauthors, 2019: Identification and characterization of persistent cold pool events from temperature and wind profilers in the Columbia River basin. *J. Appl. Meteor. Climatol.*, **58**, 2533–2551, <https://doi.org/10.1175/JAMC-D-19-0046.1>.
- Olson, J. B., and Coauthors, 2019: Improving wind energy forecasting through numerical weather prediction model development. *Bull. Amer. Meteor. Soc.*, **100**, 2201–2220, <https://doi.org/10.1175/BAMS-D-18-0040.1>.
- Pichugina, Y. L., S. C. Tucker, R. M. Banta, W. A. Brewer, N. D. Kelley, B. Jonkman, and R. K. Newsom, 2008: Horizontal-velocity and variance measurements in the stable boundary layer using Doppler lidar: Sensitivity to averaging procedures. *J. Atmos. Oceanic Technol.*, **25**, 1307–1327, <https://doi.org/10.1175/2008JTECHA988.1>.
- , and Coauthors, 2019: Spatial variability of winds and HRRR–NCEP model error statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia River basin. *J. Appl. Meteor. Climatol.*, **58**, 1633–1656, <https://doi.org/10.1175/JAMC-D-18-0244.1>.

- , and Coauthors, 2020: Evaluating the WFIP2 updates to the HRRR model using scanning Doppler lidar measurements in the complex terrain of the Columbia River Basin. *J. Renewable Sustainable Energy*, **12**, 043301, <https://doi.org/10.1063/5.0009138>.
- Rhodes, M. E., and J. K. Lundquist, 2013: The effect of wind-turbine wakes on summertime U.S. Midwest atmospheric wind profiles as observed with ground-based Doppler lidar. *Bound.-Layer Meteor.*, **149**, 85–103, <https://doi.org/10.1007/s10546-013-9834-x>.
- Sharp, J., and C. Mass, 2002: Columbia Gorge gap flow: Insights from observational analysis and ultra-high-resolution simulation. *Bull. Amer. Meteor. Soc.*, **83**, 1757–1762, <https://doi.org/10.1175/1520-0477-83.12.1745>.
- , and C. F. Mass, 2004: Columbia Gorge gap winds: Their climatological influence and synoptic evolution. *Wea. Forecasting*, **19**, 970–992, <https://doi.org/10.1175/826.1>.
- Shaw, W. J., and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP 2): General overview. *Bull. Amer. Meteor. Soc.*, **100**, 1687–1699, <https://doi.org/10.1175/BAMS-D-18-0036.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Wharton, S., J. F. Newman, G. Qualley, and W. O. Miller, 2015: Measuring turbine inflow with vertically-profiling lidar in complex terrain. *J. Wind Eng. Ind. Aerodyn.*, **142**, 217–231, <https://doi.org/10.1016/j.jweia.2015.03.023>.
- Wilczak, J. M., and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): Observational field campaign. *Bull. Amer. Meteor. Soc.*, **100**, 1701–1723, <https://doi.org/10.1175/BAMS-D-18-0035.1>.