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

Y. L. PICHUGINA,[a,b] R. M. BANTA,[b] T. BONIN,[a,b] W. A. BREWER,[b] A. CHOUKULKAR,[a,b] B. J. McCARTY,[a,b] S. BAIDAR,[a,b] C. DRAXL,[c] H. J. S. FERNANDO,[d] J. KENYON,[a,b] R. KRISHNAMURTHY,[d] M. MARQUIS,[b] J. OLSON,[a,b] J. SHARP,[e] AND M. STOELINGA[f]

[a] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[b] *NOAA/Earth System Research Laboratory, Boulder, Colorado*
[c] *National Renewable Energy Laboratory, Golden, Colorado*
[d] *University of Notre Dame, Notre Dame, Indiana*
[e] *Sharply Focused, LLC, Portland, Oregon*
[f] *Vaisala, Inc., Seattle, Washington*

ABSTRACT

Annually and seasonally averaged wind profiles from three Doppler lidars were obtained from sites in the Columbia River basin of east-central Oregon and Washington, a major region of wind-energy production, for the Second Wind Forecast Improvement Project (WFIP2) experiment. The profile data are used to quantify the spatial variability of wind flows in this area of complex terrain, to assess the HRRR–NCEP model's ability to capture spatial and temporal variability of wind profiles, and to evaluate model errors. Annually averaged measured wind speed differences over the 70-km extent of the lidar measurements reached $1 \, \mathrm{m \, s^{-1}}$ within the wind-turbine rotor layer, and $2 \, \mathrm{m \, s^{-1}}$ for 200–500 m AGL. Stronger wind speeds in the lowest 500 m occurred at sites higher in elevation, farther from the river, and farther west—closer to the Cascade Mountain barrier. Validating against the lidar data, the HRRR model underestimated strong wind speeds ($>12 \, \mathrm{m \, s^{-1}}$) and, consequently, their frequency of occurrence, especially at the two lowest-elevation sites, producing annual low biases in rotor-layer wind speed of $0.5 \, \mathrm{m \, s^{-1}}$. The RMSE between measured and modeled winds at all sites was about $3 \, \mathrm{m \, s^{-1}}$ and did not degrade significantly with forecast lead time. The nature of the model errors was different for different seasons. Moreover, although the three sites were located in the same basin terrain, the nature of the model errors was different at each site. Thus, if only one of the sites had been instrumented, different conclusions would have been drawn as to the major sources of model error, depending on where the measurements were made.

1. Introduction

Long-term mean profiles of the wind, those averaged over seasonal, annual, or longer time periods, are important for many areas of meteorology. Two major applications are wind energy (WE) and the validation and improvement of numerical weather prediction (NWP) forecast models. Wind energy's needs include, for example, searching for suitable sites for wind plants [resource assessment (RA)], characterization of flow types, shear, and turbulence environments in prospective locations, and operational forecasting for

existing wind plants (Schreck et al. 2008; Shaw et al. 2009; Marquis et al. 2011; Ahlstrom et al. 2013; Banta et al. 2013a; Fernández-González et al. 2018). Many wind farms are in complex terrain, complicating the acquisition of spatially representative wind profiles. NWP forecast model validation and improvement require accurate long-term measurements of meteorological quantities including wind, to get statistically reliable estimates of model error and to characterize conditions when large errors occur. These WE and NWP applications are interrelated (Banta et al. 2018a): WE's requirements could be addressed by more accurate NWP models. But the kinds of measurements now needed to validate and improve model boundary layer physics, traditionally cited as a major source of errors in models (Seaman 2000;

Corresponding author: Yelena L. Pichugina, yelena.pichugina@noaa.gov

Fernando and Weil 2010), are the same accurate profile data as those needed today by WE to address its needs—until such time as the more accurate, validated models become available. Long-term measurements of mean winds aloft have traditionally been a difficult problem.

Sources of vertical-profile measurements to address these needs include tall towers (e.g., Drechsel et al. 2012; Klaas et al. 2015; Fernández-González et al. 2018) and ground-based remote sensing, such as Doppler lidar. Doppler lidar can provide accurate profiles through a deeper layer of the atmosphere than towers, they can be easily moved, and they can be deployed in arrays to test the horizontal representativeness issues. The scanning or multibeam approaches to measuring the horizontal wind give estimates that are averaged spatially, rather than point measurements, similar in some respects to the spatially and temporally smoothed profiles provided by NWP models.

The issue of spatial variability of long-term wind profiles is important on many distance scales for both WE and NWP improvement, yet little quantitative information from the atmosphere exists. Here we investigate lidar-measured wind profiles in complex terrain, averaged over annual and seasonal time periods, at three sites separated by a total distance of 71 km. Such distances are relevant to decisions about where to site wind plants, so understanding the skill of models at determining wind speed spatial differences is important for assessing their usefulness for RA. For operational wind forecasting too, it is useful to understand the magnitude of the variability of the wind resource over a WE region, as well as the ability of models to simulate that variability. For model evaluation and improvement, it is important to know whether site-to-site differences exist in the nature of model errors and how *representative* these errors are of a region.

In this study wind properties are measured through the rotor layer and above by three scanning, pulsed Doppler lidars, sited along the Columbia River basin in eastern Oregon and Washington. These lidar locations are in the vicinity of several large wind farms, and the measurements provide a measure of spatial variability of winds over this region. Because of its ability to measure mean-wind characteristics sub-hourly at high precision ($<10 \, \text{cm s}^{-1}$) and resolution (vertical spacing of 10 m or less through the rotor layer), Doppler lidar has become an increasingly valuable technology for atmospheric studies (Banakh et al. 2010; Pichugina et al. 2008, 2012, 2017a; Pichugina and Banta 2010; Smalikho et al. 2013; Banta et al. 2013a, 2018a; Bonin et al. 2017) and model validation exercises related to WE (Krishnamurthy et al. 2011; Mann et al. 2017; Risan et al. 2018; Pichugina et al. 2017b;

Banta et al. 2018a; Fernández-González et al. 2018). Here we use lidar wind profile measurements in the basin over an 18-month period from two of the lidars and a 12-month period for the third to document these wind properties and their spatial and temporal variabilities, to compare these measurements with NWP output from the operational High-Resolution Rapid Refresh (HRRR) model run at NOAA/NCEP, and to evaluate several error statistics, such as bias and root-mean-square error (RMSE), as vertical profiles, time series, and functions of forecast lead time and hour of the day.

The lidar deployments were part of a larger field campaign, the Second Wind Forecast Improvement Project (WFIP2) from September 2015 through March 2017. The experiment location, the Columbia River basin, is not only the site of many large wind farms, but is also a major source of WE for the western United States. Overviews of the WFIP2 campaign describe the experiment and scientific objectives (Shaw et al. 2019), the instrumentation involved in the experiment, including some preliminary research findings (Wilczak et al. 2019), and the numerical forecast model experiments that were an integral part of the program (Olson et al. 2019).

A novel opportunity presented by this project is that a validation dataset has been provided from an NWP forecast model (HRRR) for which the dataset comprises high-quality vertical profiles every 15 min for the project duration from Doppler lidars taken at *three sites*. This triple deployment allows us to perform a comprehensive, long-term evaluation of model skill as well as an in-depth evaluation of the important issues of spatial variability and representativeness of the results. With the large dataset, we are able to separate the annual findings into seasons, to investigate the effects of wind direction, and to look at day-versus-night effects in a quantitative manner.

Section 2 of this paper describes the location and topography of the study region, focusing on the lidar locations, lidar technical parameters, measurement scanning sequence, and overall performance of each lidar along with a short discussion of the HRRR model used in this paper. Section 3 provides lidar-measured distributions and mean-profile values of wind speed and direction at each site for seasonal and annual time scales. This section also provides analyses of wind speed profiles averaged over westerly and easterly wind directions as well as for daytime and nighttime periods. Section 4 presents validation of the operational HRRR model by Doppler-lidar measurements at the three sites, and section 5 presents a brief summary and conclusions.

FIG. 1. (a) Map of the study area in the Columbia River valley; the locations of scanning Doppler lidars are denoted by gold circles. (b) A view of Mount Hood from Wasco County, showing also wind turbines and the complex land surface. (c) Photograph (provided through the courtesy of S. Wilson) of the Shepherds Flat Wind Farm in the area.

2. Study area, lidar locations, and HRRR model properties

A map of the study area located east of the Cascade Mountain Range is shown in Fig. 1a where orange circles mark the three scanning Doppler-lidar sites along the Columbia River valley. Synoptic-scale systems (often moving onshore from the Pacific Ocean), the Cascade Mountains with several high-elevation peaks, the Columbia River canyon, and local terrain features as well as the presence of numerous wind farms (Figs. 1a–c) all combine to create complex wind flows over the study domain. The imposing Cascade Mountain Range provides a natural barrier between the Pacific marine air to the west and the arid conditions over the wind-farm study area to the east (Figs. 1b,c). Wind flow through the narrow Columbia River Gorge, the only sea level gap through this range, has been studied by Sharp and Mass (2002, 2004), who point out that the gap winds are strongly controlled by the surface pressure gradient. Higher pressure offshore to the west of the mountains produces westerly gap flow through the Gorge, whereas higher pressure inland to the east generates easterly gap flow there.

a. Scanning Doppler lidars in WFIP2

Two scanning, pulsed Leosphere WindCube 200S Doppler-lidar systems (200S hereinafter) were deployed to sites near the Wasco and Arlington, Oregon, airports that are 40 km apart, providing real-time measurements from September 2015 to April 2017. The third scanning Doppler lidar, a HALO Streamline XR (HALO hereinafter), was deployed by the University of Notre Dame to a site near Boardman, Oregon, 31 km east-northeast of Arlington (Fig. 2, top) and was continuously operated from January through December 2016. The westernmost lidar at Wasco documented an inflow profile upstream of the wind farms during westerly flow, whereas wind profile shapes and magnitudes at the middle lidar at Arlington and the easternmost lidar at Boardman may be altered because of distance from the Cascade Range, other terrain, and upstream wind farms.

1) LIDAR SPECIFICATIONS

System parameters of these lidars are given in Table 1. The lidars operated at pulse repetition rates of 10–20 kHz, averaged to 2-Hz beams (for the 200S lidars) or 1-Hz
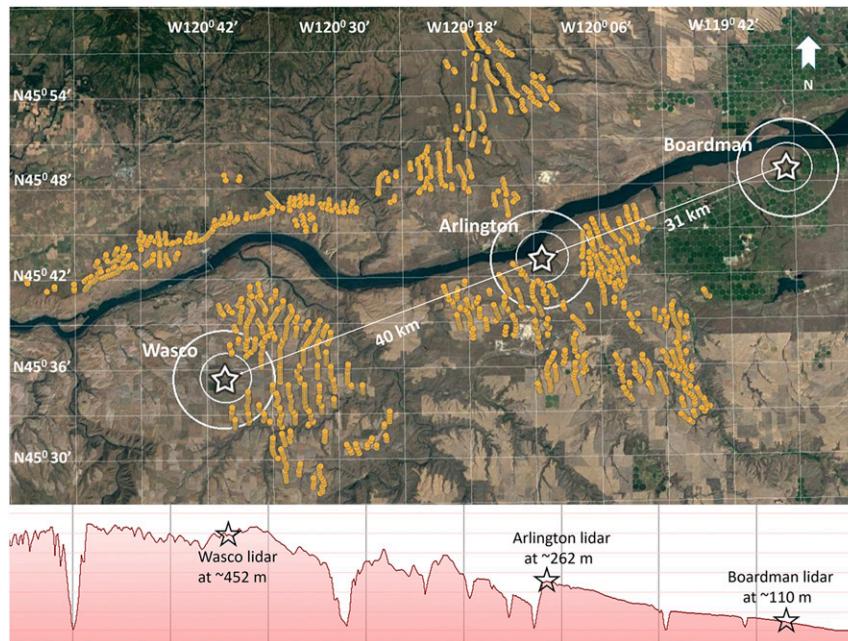
FIG. 2. (top) Google Earth (image credit: Landsat/Copernicus) map of the Eastern Gorge portion of the study area, including the locations of two NOAA Doppler lidars (200S) at the Wasco site at 452 m above mean sea level (MSL) and the Arlington site at 262 m MSL and the University of Notre Dame HALO lidar at the Boardman site at 110 m MSL. The lidar locations are shown by white-outlined stars. Two white circles around each lidar indicate measurement area up to 3 and 6 km. The white line indicates an ''East–West'' transect of the study region along prevalent wind directions observed from surface measurements and models during previous studies in this area (Sharp and Mass 2002, 2004). The surrounding wind turbines are indicated by the clusters of dark yellow circles. (bottom) Terrain elevation transect along this line.

beams (for HALO lidar), to provide range-resolved measurements of the radial or ''line of sight'' (LOS) velocity parallel to the beam u_r, at a range resolution of 50 or 100 m (Table 2) and an individual-beam instrumental precision of 20 cm s^{-1}.

2) SCANNING PROCEDURE

Lidar scan sequences were designed to measure sub-hourly profiles of the mean wind at three individual sites. For separation distances between lidars of more than 30 km, coordinated, overlapping scanning among several lidars as performed in projects such as Berg et al. (2015) was not possible. Our measurement routine included a 15-min sequence of multiple azimuthal (conical) scans, elevation (vertical slice) scans, and vertical staring of the lidar beam (vertical stare mode, to measure the vertical-velocity component w). Conical scans were performed at several elevation angles to provide high-resolution mean-wind profiles from near the surface through and above the atmospheric boundary layer. Elevation (vertical slice)

TABLE 1. System parameters of Doppler lidars.

Lidar parameters	WindCube 200S (50-m gate)	WindCube 200S (100-m gate)	Halo Streamline XR (100-m gate)
Wavelength (μm)	1.54	1.54	1.5
Pulse energy (mJ)	0.1	0.1	0.08
Pulse duration (ns)	200	400	350
Pulse repetition frequency (Hz)	20 000	10 000	10 000
Sampling frequency (MHz)	250	250	50
Accumulation time (s)	0.5/1	0.5/1	0.5/1
Range resolution (m)	50	100	48
Min range (km)	0.1	0.2	
Typical max range (km)	3–5	5–7	3–5

TABLE 2. Scan sequence continuously performed by lidars at three sites every 15 min.

Scan type	Elev (°)	Azimuth (°)	Duration (min)	Comments
		Wasco		
Conical (PPI)	3 and 30	0–360 scanning	6.5	Loaner lidar until 4 Nov 2015
	2.75, 6, 15, and 45	0–360 scanning	6.5	50-m gates until 21 Mar 2016
	2.75, 4, 6, 15, and 45	0–360 scanning	8.5	100-m gates after 21 Mar 2016
Elev (RHI)	0–30 scanning	0, 90, 180, 270	3.5	
Vertical stare	90	—	5 and 3	
		Arlington		
Conical (PPI)	1.75, 6, 15, and 45	0–360 scanning	6.5	50-m gates until 24 Feb 2016
	1.75, 3, 6, 15, and 45	0–360 scanning	8.5	100-m gates after 24 Feb 2016
Elev (RHI)	0–30 scanning	0, 90, 180, and 270	3.5	
Vertical stare	90	—	5 and 3	
		Boardman		
Conical (PPI)	1.75, 3, 6, 15, and 45	0–360 scanning	6.5	
Elev (RHI)	0–30 scanning	0, 90, 180, and 270	3.5	
Vertical stare	90	—	5	

scans were performed at fixed azimuth by scanning from the surface up to 30° elevation and back. These scans were repeated approximately toward the north, west, south, and east. For the last 3 min of each 15-min sequence, the lidar pointed vertically (90° elevation). Scan types and durations during each 15-min sequence are presented in Table 2 and illustrated in Fig. 3a,b. The scanning pattern, developed to meet lidar objectives during WFIP2 was synchronously repeated by each lidar every 15 min.

Our lidar mean-wind profile calculation consists of using the standard velocity–azimuth display (VAD) procedure (Lhermitte and Atlas 1961; Browning and Wexler 1968), modified to include all scans taken within the 12-min averaging period (Banta et al. 2002, 2015); vertically pointing data from the last 3 min were excluded. Data combined from all conical scans (four or five scans) and all four elevation (vertical-slice) scans are binned according to height AGL. Each bin thus represents a roughly disk-shaped horizontal layer containing concentric rings of data from the conical scans that intersect the layer plus four crossing lines of data from the elevation scans. Each data point in a given bin has an associated scan azimuth θ and horizontal radial-wind projection u_{rh}. Data in each bin, for all data points from all scans, are used to calculate the mean horizontal wind direction and speed using the modified VAD technique, which consists of finding the best-fit sine wave for the u_{rh} data as a function of θ.

An automated quality assurance (QA) procedure is then applied. First, data points suspected of second-trip contamination were filtered out using the procedure as in Bonin and Brewer (2017). Then the departure of the radial velocity from the mean-wind sine-wave value for each point (*residuals*) was calculated

as a function of θ, by subtracting the sine-wave value from the original data. Outliers exceeding 3σ, where σ is the standard deviation of the residuals, are removed from the original sample. In a second QA pass, a new mean-wind sine wave and new residuals are calculated, and outliers of $>3\sigma$ are again removed to yield a final sample. The mean winds used in this study were calculated from this final (θ, u_{rh}) sample for each bin in the vertical wind profile. This procedure removes spurious hard-target returns, sharp turbulent bursts, significant small-scale terrain effects, and other strong u_{rh} anomalies that are unrelated to the mean flow. A final QA step discards wind estimates if the standard deviation of the mean of the residuals is larger than 0.25 m s^{-1} at that particular level.

The result of this procedure is thus that in each layer in the vertical, winds are averaged over a sampling disk, which could be 10 km across, from scans that span a 12-min period. This averaging over time and space resembles the smoothing inherent in NWP models, because, as is well known in NWP modeling and recently recalled by Banta et al. (2018a, p. 1163), "numerical and explicitly modeled diffusion ensure that model variable fields are smooth enough to prevent numerical instability." That is, model gridpoint values represent a smoothed average over several model grid intervals in the horizontal, in the vertical, and in time, in a manner that has not been well characterized. The smoothing typically occurs over at least five grid intervals (Skamarock 2004 found perceptible effects up to seven intervals), which for HRRR would be 15 km. The multiscan lidar method described here may thus be the most appropriate measurement procedure currently available to approximate the innate smoothing in NWP model fields.
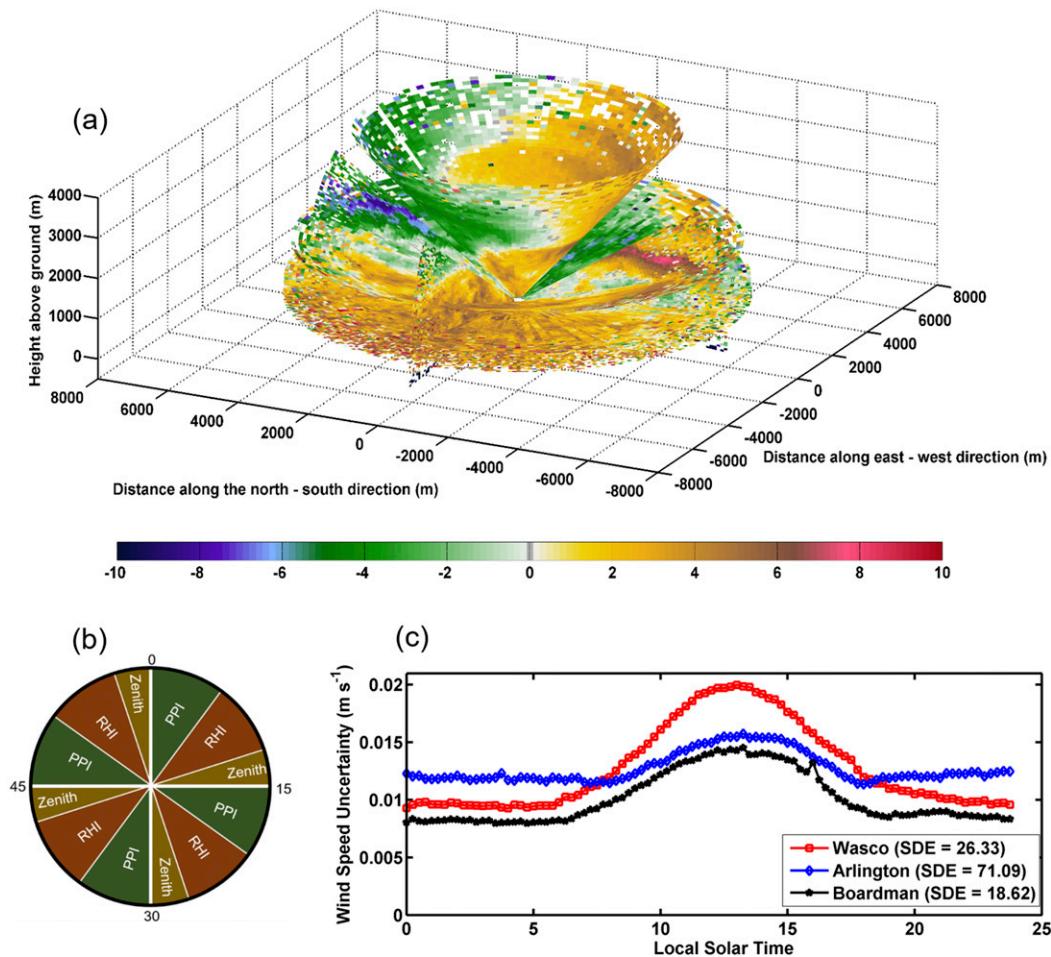
FIG. 3. (a) Illustration of lidar scanning sequence used during WFIP2, comprising several conical (PPI) scans, performed at low (2°, 3° or 4°, and 6°) and higher (15° and 45°) elevation angles, and four vertical-slice (RHI) scans performed at azimuth angles of 0°, 90°, 180°, and 270°. (b) Clockwise diagram of the 15-min lidar scanning pattern that was synchronously repeated by each lidar. (c) Average diurnal behavior of the measurement uncertainty over a full year, showing a strong diurnal signature as well as dependence on terrain complexity for each site, shown here as standard deviation of the terrain elevations (label SDE) within a 3-km radius of the instrument location (Ascione et al. 2008).

Mean winds calculated from remote sensing scans or multiple-beam measurements contain uncertainties from various sources, such as meteorological (turbulent eddies and other transient features), aerosol loading [signal-to-noise ratio (SNR)[1] is correlated with measurement uncertainty], terrain complexity (validity of homogenous-flow assumption), and others. Although studies have expressed concern that horizontal-flow inhomogeneities over individual scans may degrade the accuracy of the mean wind calculated over the scan footprint (Bingöl et al. 2009; Pauscher et al. 2016;

Mann et al. 2018), the results of interest in this study involve averaging over several months to a year, which comprise many thousands of profiles. Over those time periods and large sample sizes, random small-scale spatial-variability effects average out. In particular, they will not contribute incorrectly to calculated model errors, including biases.

This point was demonstrated in a complex-terrain setting by Klaas et al. (2015), who compared profiling-lidar and tall-tower data over an extended period to obtain sample sizes of 7000–10 000. They found that the long-term, lidar-tower differences (their Table 2: "mean absolute error") were less than 0.10 m s^{-1} at each of four vertical levels between 80 and 200 m AGL. This estimate is consistent with our own studies, using a formulation of

---

[1] Wide-band SNR as used here is the same as the carrier-to-noise (CNR) term used by the telecommunications industry.
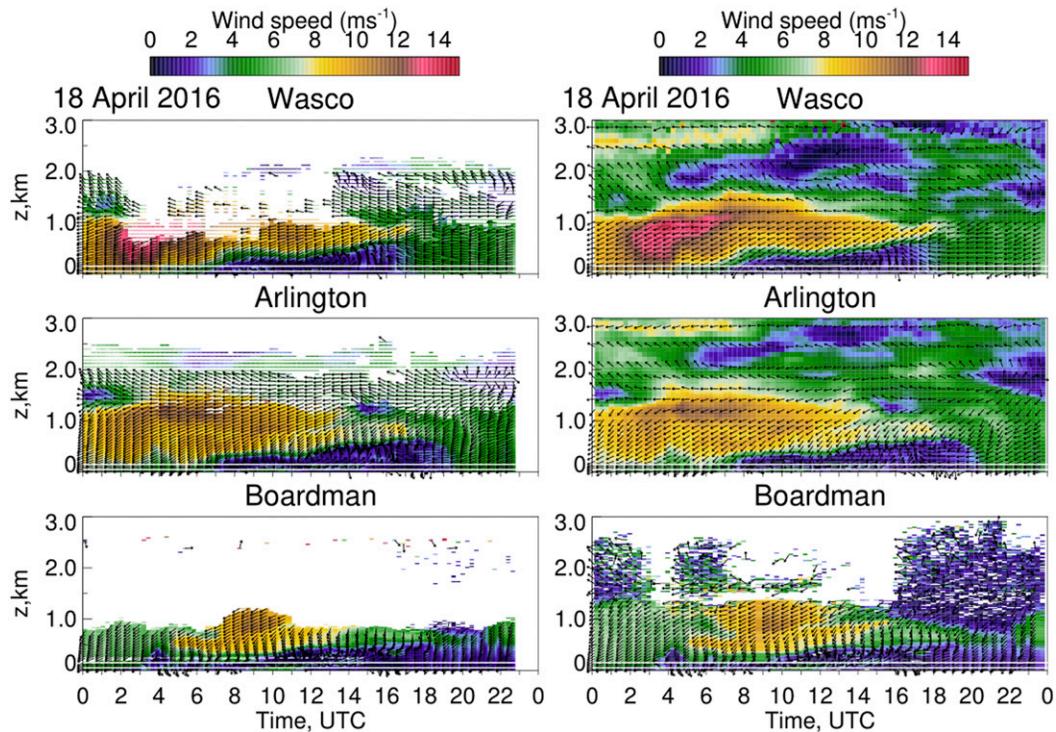
FIG. 4. Time–height cross sections of wind speed and wind direction at the (top) Wasco, (middle) Arlington, and (bottom) Boardman sites for 18 Apr 2016, showing data (left) processed in real-time using the VAD technique and (right) processed using a newly developed technique ("shifter") to increase lidar-backscatter signal range and vertical coverage (Bonin et al. 2018) and applying additional quality control procedures to the data from the HALO lidar.

Newsom et al. (2017), modified by adding a term representing SNR-weighted residuals of u_{rh} averaged over the year, to obtain the wind speed precision based on within-scan variability from all sources (Fig. 3c). These quantities show a strong diurnal signature as well as dependence on terrain complexity for each site. The wind speed estimates have an average precision of better than $0.02 \, \mathrm{m \, s^{-1}}$ within the rotor layer, mainly because of the large sample size of LOS measurements at each height over the year, producing high confidence in the mean-wind estimate. This magnitude of discrepancy, and the $0.10 \, \mathrm{m \, s^{-1}}$ found by Klaas et al. (2015), are not enough to affect any of the site-to-site lidar annual differences or annual model errors found in this study.

3) LIDAR-SCAN ADJUSTMENTS AND DATA AVAILABILITY

During certain periods of WFIP2, low aerosol concentrations or other environmental factors led to low lidar SNR and reduced lidar range, that is, the distance out from the lidar of useful signal. This happened most often in winter, when coastal storms could rain out much of the particulate matter on the west side of the Cascades. Stiff westerly flow through the Gorge would

then be cleansed of aerosol, at times resulting in lidar-range reductions. These reduced-range periods caused us to change our lidar settings and scan sequences (see Table 2).

We used two approaches to improve the data quality. First, we augmented vertical coverage by using a longer pulse width and increased range-gate length (from 50 to 100 m), thus enhancing SNR and lidar range. Along with this, adjustments to the elevation angles of the conical scans of the 200S lidars at Wasco and Arlington, as indicated in Table 2, were necessary to maintain the required vertical resolution and coverage. Second, during postprocessing, the profile data for the 200S lidars were reprocessed to achieve greater coverage, using a new NOAA lidar *shifter* spectral-processing technique (Bonin et al. 2018), which extracts mean-wind profile data from raw spectral data in low-SNR conditions, often extending the lidar range and, consequently, the depth of coverage significantly. Examples of the effectiveness of this technique are shown in the right panels of Fig. 4.

For the HALO lidar at Boardman, raw spectral data were unavailable and therefore the NOAA-lidar shifter technique could not be applied. However, the vertical coverage of winds was improved by accepting

TABLE 3. Availability (%) of lidar measurements of wind speed at 100 m AGL and the number of measurement days during each month of 2016. An asterisk indicates that the lidar was offline because of scanner replacement during the month.

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Availability of measurements												
Wasco	59	65	73	99	99	98	98	100	96	89	45	10[*]
Arlington	69	61	94	100	75	99	99	100	96	3[*]	65	54
Boardman	70	75	76	83	92	96	98	100	98	87	90	83
No. of measurement days												
Wasco	29	26	31	30	31	30	31	25	30	25	11[*]	4[*]
Arlington	30	27	31	30	24	30	31	25	30	3[*]	20	19
Boardman	27	29	29	26	29	30	31	25	30	23	24	23

lower-SNR data and weighting the data in the VAD fit by an uncertainty factor determined from the Cramer–Rao lower bound, a curve that relates the u_r uncertainty to SNR (Rye and Hardesty 1993).

Monthly availability of lidar measurements η at hub height, estimated as the ratio of the number of observations at a given height N_{OBS} to the expected (total) number N of 15-min profiles in the diurnal cycle ($\eta = 100\% \times N_{OBS}/N$), is given in Table 3. Overall, all three lidars performed better in March–September than in the winter months, partly because of the weather conditions as described (winter fog and precipitation; lower aerosol loading) and partly because of technical lidar hardware problems (e.g., malfunction and replacement of broken lidar scanners resulting in missing data in October 2016 at the Arlington site and in November and December 2016 at Wasco). Decreases of lidar performance with height, which invariably occur (Post et al. 1997; Banta et al. 2013b, p. 473), were mostly due to a drop in SNR resulting from the decrease of aerosol concentrations with height.

All analyses presented in this paper are based on the 12-min averaged profiles of wind speed and direction, which are available every 15 min, as just described. The QAed profile data were combined into daily time–height cross sections to demonstrate how the flows evolved on any given day, or were averaged to form seasonal and annual statistics. Two examples of individual-day time–height cross sections of the wind speed are shown in Fig. 5 for 26 February and 28 June 2016, a wintertime and a summertime day, respectively. The wind patterns on 26 February consisted of diminishing easterly gap flow due to a synoptic slackening of the pressure gradient across the Cascades, and the buildup of a cold-air-pool layer indicated by very weak wind speeds, as noted previously by Whiteman et al. (2001) and Zhong et al. (2001). The wind speed patterns on 28 June reflect the summertime diurnal cycle of winds due to daytime heating and nighttime cooling, which were especially strong east of the Cascades, including

over the Columbia River basin (Banta et al. 2018b, 2019, manuscript submitted to *Mon. Wea. Rev.*). These two meteorological patterns occurred several times during WFIP2 and are the subject of more in-depth studies of cold-air pools (McCaffrey et al. 2019, manuscript submitted to *J. Appl. Meteor. Climatol.*) and summertime diurnal westerly wind systems (Banta et al. 2019, manuscript submitted to *Mon. Wea. Rev.*).

b. HRRR–NCEP numerical weather prediction forecast model

This study uses forecast output from the operational HRRR as produced by NCEP during the WFIP2 field campaign. The HRRR is an hourly updated, convection-allowing NWP model with 3-km horizontal grid spacing that runs in a domain encompassing the continental United States. It is used by many wind-industry forecasting vendors in formulating predictions of turbine-height winds.

Since only a few studies have validated forecasts of wind profiles, one objective of this study is to quantify the skill of the HRRR model in predicting low-level (the first 500 m AGL) winds across different seasons. Moreover, in conjunction with—and informed by—the WFIP2 field campaign, a WFIP2 model-development effort was undertaken to improve the physical parameterizations in HRRR (Olson et al. 2019). Thus, a second objective of this study is to quantify HRRR forecast errors as a baseline against which the performance of future (upgraded) HRRR versions can be assessed.

The operational HRRR model analyzed in this study was upgraded on 23 August 2016, and this upgrade included changes to data-assimilation and model-physics components. With the exception of this upgrade, the HRRR model forecasts in this study were produced by a static (or "frozen") HRRR configuration. Benjamin et al. (2016) give a comprehensive description of the Rapid Refresh (RAP) and HRRR operational models. For data assimilation, both models use the NOAA Gridpoint Statistical Interpolation analysis system (Benjamin et al.
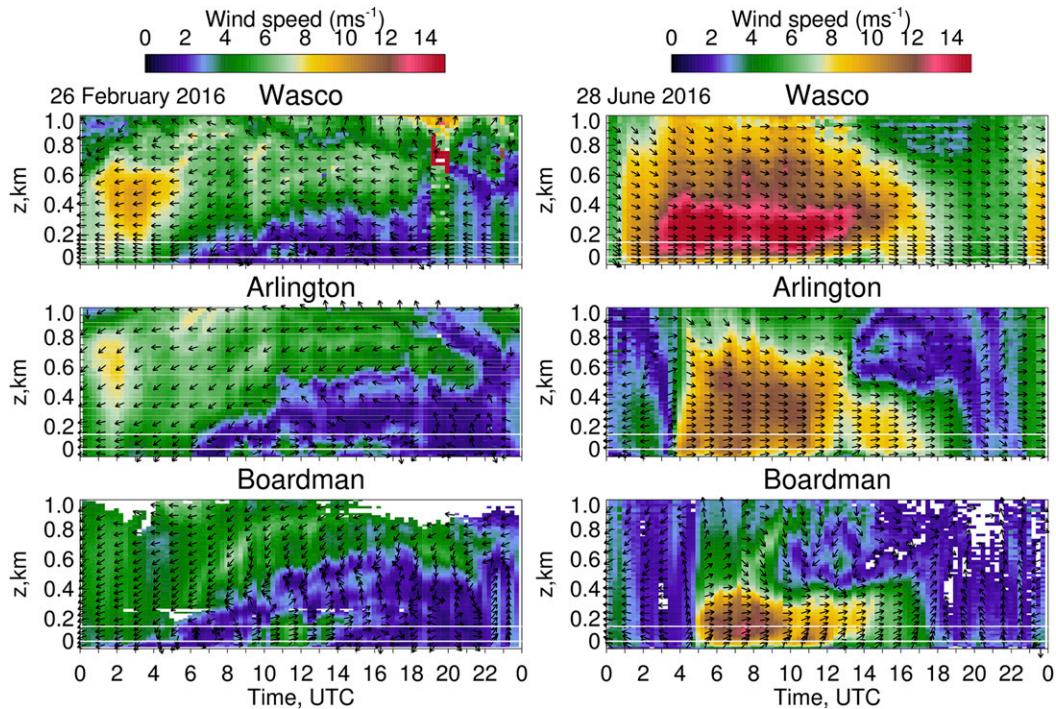
FIG. 5. Time–height cross sections of wind speed and wind direction at the (top) Wasco, (middle) Arlington, and (bottom) Boardman sites during (left) 26 Feb and (right) 28 Jun 2016, illustrating vertical, diurnal, and spatial (between sites) wind flow variability. Wind speed is color coded from 0 to 15 m s^{-1} according to the color scale. Black arrows show wind direction. Two white horizontal lines on each panel indicate heights of 50–150 m AGL.

2016, and references therein) with enhancements for hourly assimilation of data from a variety of sources, including surface sites, rawinsonde, radar, satellite, and onboard aircraft sensors (see Table 4 of Benjamin et al. 2016). One difference between the RAP and HRRR initialization procedure is in the post-data-assimilation rebalancing of the atmospheric state. The RAP uses a digital filter initialization (DFI; Peckham et al. 2016) to improve the balance of the initial condition and minimize the noise caused by gravity waves generated during the adjustment process. The current operational HRRR does not employ DFI after the data-assimilation process. Consequently, spurious gravity waves can produce artificial mixing, which can erode stable layers and potentially impact low-level winds. This problem was discovered during WFIP2, prompting the likelihood of adding DFI to the HRRR initialization procedure in future versions.

HRRR gridpoint values were available at the top of each hour. Doppler-lidar measurements were averaged for the 15-min value before and after the hour to provide an hourly dataset. The model grid values were interpolated bilinearly in the horizontal to the location of each lidar. This smoothing acts to reduce any phase or lag effects, and the long-term averaging performed in this study should also mitigate these effects.

The performance of NWP models has been evaluated using remote sensing measurements above the surface from various instruments during several short-term (mostly monthlong duration or less) field experiments (e.g., Zhong and Fast 2003; Fast and Darby 2004; Yang et al. 2017; Wilczak et al. 2015; Risan et al. 2018). Longer-term model validations have been performed using tall-tower measurements (Drechsel et al. 2012; Draxl et al. 2014; Krogsæter and Reuder 2015; Klaas et al. 2015; Fernández-González et al. 2018) and routine rawinsonde (Benjamin et al. 2016). Recent studies have validated simulated wind profiles from NOAA operational NWP models against lidar-measured profiles in an offshore setting. Validation of NOAA (RAP and HRRR) and NCEP [NAM Rapid Refresh (NAMRR) and NAMRR continental U.S. (CONUS) Nest] models were performed over the ocean off the U.S. East Coast, using ship-based, motion-compensated Doppler-lidar measurements (Pichugina et al. 2017b; Banta et al. 2018a). In general agreement with the other studies, these studies also found wind speed errors of 2–2.5 m s^{-1} in the lowest several 100 m of the atmosphere, with larger errors tending to occur during nighttime hours, often in the presence of low-level jets (LLJs).

These studies, along with Djalalova et al. (2016) also looked at the effects of model resolution by comparing the 13-km RAP with the 3-km HRRR. At lead times of less than 3 h, the HRRR generally showed lower errors than the RAP. But at lead times greater than 6 h, the coarser-resolution RAP often exhibited smaller errors than HRRR, contrary to expectation. Thus, model resolution affects model-error properties, but often in unexpected ways (Banta et al. 2018a). Unresolved variations of terrain elevation, roughness, and other surface properties also exist and contribute in an unknown way to model errors.

3. Wind flow properties and statistics measured by Doppler lidar

a. Hub-height and rotor-layer distributions of wind speed and wind direction

1) ANNUAL DISTRIBUTIONS

The mean 2016 annual distributions of wind speed in the rotor layer at 50–150 m (Fig. 6a) show the strongest winds at Wasco, and weakest at the lowest site, Boardman, where a large percentage of weak winds of 3 m s^{-1} or less produced a bimodal distribution to the wind speeds, which can also be seen at Arlington. The single-mode Weibull distribution often used for long-term distributions in WE (Kelly et al. 2014) was thus not a good fit for winds at these sites. Quantitatively, the 12-min wind speeds ranged from 0 to 20 m s^{-1} with mean values of 7.1, 6.7, and 5.8 m s^{-1} at Wasco, Arlington, and Boardman, respectively (Table 4).

Distributions of wind direction were not merely bimodal but nearly bidirectional, blowing from the west *or* the east but seldom from any other directions, especially at Wasco and Arlington. Even though the terrain is relatively open and basinlike in this area, the directional dichotomy indicates strong topographical channeling of the flow. The westerly and easterly flow directions had respective mean values of 272°, 261°, and 250° and 91°, 71°, and 71° at each site, where westerly flow by far dominated easterly for wind speeds greater than 4 m s^{-1} (Table 5).

The standard deviation σ of the wind directions[2] for westerly-component flow (Table 4) was only ~20° at Wasco and Arlington and 30° at Boardman, indicating a

---

[2] Because of the complete separation in the distributions of westerly and easterly flow directions, implying no 0° vs 360° ambiguities for either sample, we calculated a simple linear standard deviation for each westerly and easterly distribution of wind directions.
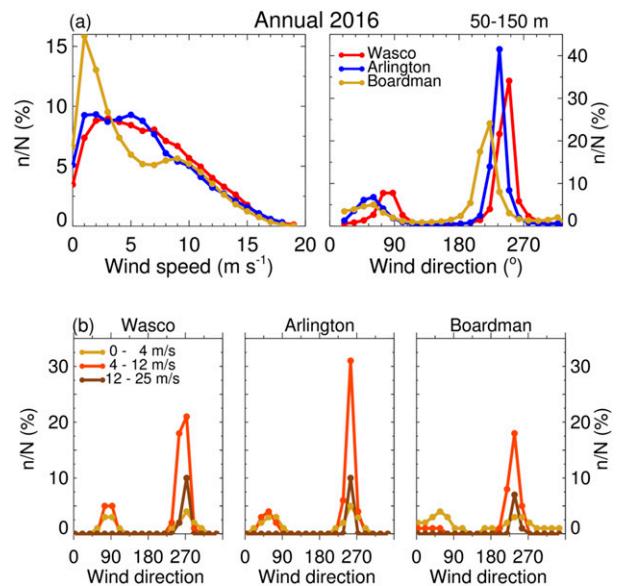

FIG. 6. (a) Annual distributions of rotor-layer wind speed and direction at the Wasco (red), Arlington (blue), and Boardman (goldenrod) sites. (b) Annual distribution of rotor-layer wind speed at each site, plotted as a function of wind direction. Colors indicate three wind speed categories as specified in the legend at the top of the leftmost panel.

narrow range. It is shown in Fig. 6b that these annual distributions were dominated by the statistics of the more frequent westerly flows. The speed distributions for the subsample of only westerly winds are also distinctly bimodal, especially at Arlington and Boardman (see Fig. 11, described in more detail below, for evidence). Thus, even for this narrow westerly sector, the Weibull distribution was not a good fit. Results were similar for the experiment-long (September 2015–March 2017) distributions from lidars at the Wasco and Arlington sites (Fig. A1, in the appendix). The westerly–easterly flow-direction dichotomy found by Sharp and Mass (2002, 2004) within the Gorge is thus shown here to extend well to the east into the Columbia Valley, at least as far as the site at Boardman (Fig. 2).

Annual distributions of 100-m wind speed as a function of wind direction are shown in Fig. 6b, where observed winds are categorized into three groups (0–4, 4–12, and 12–25 m s^{-1}) approximating those relevant to wind-turbine operations. The isolation of westerly versus easterly winds is obvious in these distributions. The differences in behavior between these two groups give insight into their relevance to WE and NWP improvement, so we will show their attributes separately as part of the analyses. We will also show daytime versus nighttime differences, to highlight the differences in effect between unstable and stable boundary layer physics.

TABLE 4. Annual mean and σ values of wind speed (m s^{-1}) and wind direction (°) distributions from 15-min lidar measurements.

	All		East		West		East		West	
	Speed	σ	Speed	σ	Speed	σ	Direction	σ	Direction	σ
					Annual 2016 for 50–150 m					
Wasco	7.1	4.3	4.5	3.1	8.0	4.2	91	28	272	20
Arlington	6.7	4.3	3.9	2.8	7.5	4.3	71	34	261	19
Boardman	5.8	4.2	3.1	2.2	7.0	4.3	71	45	250	30
					Annual 2016 for 0–1000 m					
Wasco	8.1	5.4	5.8	4.9	9.1	5.3	92	36	269	27
Arlington	7.6	5.3	5.2	4.8	8.6	5.1	76	39	259	24
Boardman	5.8	4.6	3.7	2.7	6.8	5.0	74	48	252	35

Westerly winds were more frequent than easterly in each wind speed category at all three sites (Fig. 6b, Table 5), and winds stronger than 12 m s^{-1} were almost exclusively from the west. The most frequent category was westerly winds of 4–12 m s^{-1} (57%, 55%, and 45%), and the least frequent, the strong 12–25 m s^{-1} easterly wind. Wind speeds stronger than 25 m s^{-1} from any direction were observed rarely, with only 11 (5) cases recorded at the Wasco (Arlington) site. This very small percentage of cases is negligible in the distributions, and winds stronger than 25 m s^{-1} have not been represented in Fig. 6b or Table 4.

2) SEASONAL DISTRIBUTIONS

Distributions of rotor-layer (50–150 m) wind speed and direction have been separated into seasons in Fig. 7, which shows stronger winds during summer months compared with winter (Table 6). Westerly wind directions were more frequent for spring, summer, and fall because of the seasonal tendency for offshore ridging, whereas during the winter season the occurrences of westerlies and easterlies were comparatively more equal. The increased occurrence of easterlies and weaker winds in winter resulted from frequent near-stagnant conditions beneath basin cold-air pools and frequent high pressure inland, as previously described. Boardman had a high percentage of weak winds in all seasons, with distinct bimodal distributions for all seasons except winter, when a large fraction of weak wind speeds formed a single mode at 2 m s^{-1}. This lower-elevation site farther from

the Gorge opening thus had a less favorable wind speed distribution for WE generation than the other two sites.

b. Vertical structure of wind flow—Mean profiles

1) ANNUAL

Figure 8 shows the annually averaged wind speed profiles at the three sites for all wind directions, and then for only the westerly wind and only easterly occurrences. Within each panel, the profiles are further divided into averages for approximate nighttime (0000–1200 UTC) and daytime (1200–0000 UTC) hours. The mean profiles for easterly flow indicate the strongest winds at Wasco, but these speeds at all sites varied little with height between 100 and 500 m. Among the three sites the easterly profiles differed by less than 1 m s^{-1}, with speeds of 3–4 m s^{-1} (daytime) and 4–5 m s^{-1} (nighttime).

For westerly flow, which also dominated the "All data" statistics by virtue of its greater frequency of occurrence, the westernmost Wasco site had the strongest winds, and easternmost Boardman, the weakest. The nighttime wind profile at Wasco was as much as 2 m s^{-1} stronger than Arlington, which was up to 1–2 m s^{-1} stronger than Boardman. The weaker daytime profiles showed a similar site-to-site trend but smaller differences. Westerly flow means reached 5 m s^{-1} stronger than those for easterly winds, and nighttime winds were stronger on average than daytime in all cases. LLJ structure was evident in the mean nighttime westerly profiles at Wasco and Boardman, as a

TABLE 5. Annual frequency of 100-m wind speed occurrence (%) in each wind speed category for all wind directions, and how each category is divided between easterly and westerly wind directions (% of all occurrences in the category). The last column shows the annual frequency NN (%) of winds above the cut-in speed (4 m s^{-1}) and available for power generation at the location of the lidars.

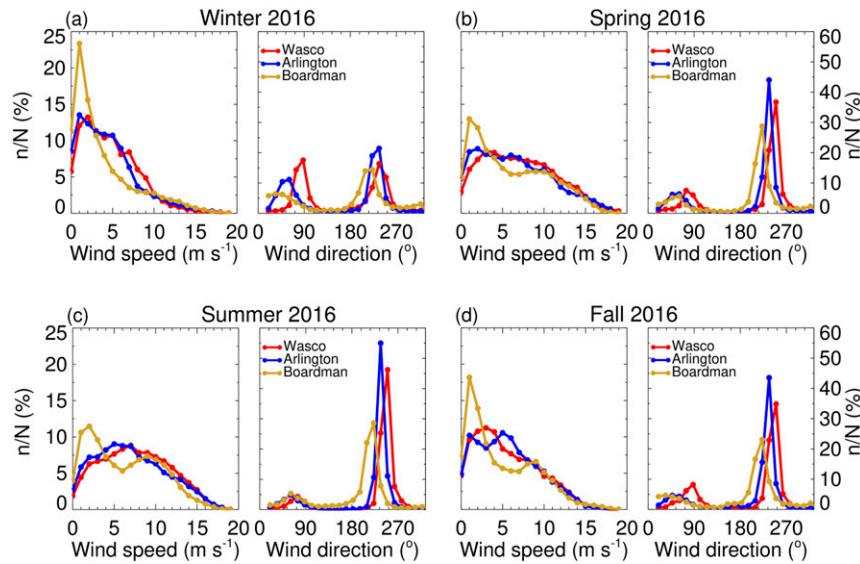|	0–4 m s^{-1}			4–12 m s^{-1}			12–25 m s^{-1}			NN %
	All	East	West	All	East	West	All	East	West	three groups
Wasco	31	46	54	57	26	74	12	0.5	99.5	69
Arlington	33	49	51	55	20	80	12	0.8	99.2	67
Boardman	46	50	50	45	19	81	9	0.4	99.6	54

FIG. 7. Seasonal distributions of rotor-layer wind speed and direction at three sites (color key).

result of gap flows in winter and nocturnal marine intrusions in summer, flows that were typically less than 600 m deep.

Averaging over annual periods (as here) produces smoothed wind speed profiles but large natural variability. The annual wind speed standard deviations were ~4–5 m s^{-1} for westerly and ~2–3 m s^{-1} (4 m s^{-1} at Boardman) for easterly winds (Table 4), with little difference in standard deviation between daytime and nighttime hours for both wind directions (not shown). Standard deviation values such as these can be useful

for verification of distribution parameters, such as the Weibull shape parameter (e.g., Kelly et al. 2014).

2) SEASONAL

Seasonally averaged profiles for westerly and easterly component wind speeds are shown in Fig. 9. Below ~400 m the smallest west-versus-east differences are seen in winter, because of the frequent cold-air pools and inversion layers. Wind speeds under these conditions tended to be weak, and directions, highly variable, with a higher frequency of easterly winds

TABLE 6. Seasonal mean and σ values of wind speed (m s^{-1}) and wind direction (°) distributions from 15-min lidar measurements.

|	All		East		West		East		West	
	Speed	σ	Speed	σ	Speed		Speed	σ	Speed	σ
Winter 2016 50–150 m										
Wasco	5.5	4.4	4.8	4.0	6.1	4.6	93	25	263	29
Arlington	5.3	4.4	4.1	3.6	6.2	4.7	68	30	258	23
Boardman	4.2	3.7	3.0	2.2	5.2	4.3	67	45	251	39
Spring 2016 50–150 m										
Wasco	7.5	4.2	4.8	2.6	8.4	4.3	85	29	271	21
Arlington	7.1	4.5	4.2	2.7	8.2	4.6	63	30	260	19
Boardman	6.5	4.4	3.5	2.3	7.7	4.5	63	38	253	31
Summer 2016 50–150 m										
Wasco	8.1	4.1	4.3	2.1	8.6	4.0	78	24	275	15
Arlington	7.5	4.2	3.8	2.2	8.3	4.0	66	28	263	15
Boardman	6.6	4.2	2.9	1.7	7.9	4.0	82	42	242	24
Autumn 2016 50–150 m										
Wasco	5.8	3.6	3.9	2.3	6.6	3.7	93	30	267	25
Arlington	6.1	3.8	3.5	2.8	8.9	3.7	69	34	260	21
Boardman	5.4	3.9	3.2	26	6.4	4.0	61	43	250	33
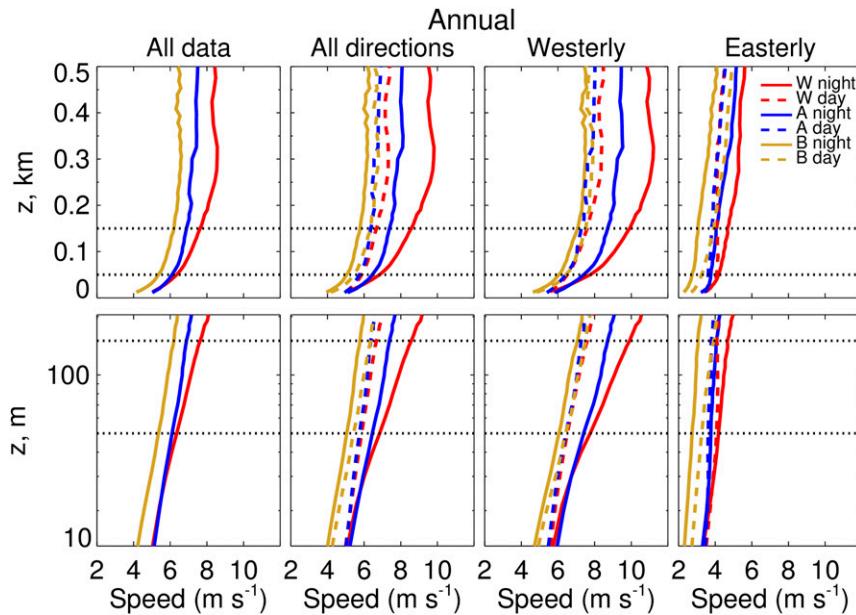
FIG. 8. Annually averaged wind speed profiles for (left) all data, (center left) all data segregated by day vs night, (center right) westerly wind directions, and (right) easterly wind directions. The vertical axis for the top panels is shown for a linear height scale from 0 to 500 m; in the bottom panels, heights are shown in a logarithmic scale from 0 to 200 m. Data are shown for lidar measurements at the Wasco (W; red), Arlington (A; blue), and Boardman (B; goldenrod) sites during 1 Jan–31 Dec 2016. In the day-vs-night panels, solid lines represent data averaged over nighttime (0000–1200 UTC) hours and dotted lines represent daytime (1200–0000 UTC) hours.

than in summer. The largest discrepancies between westerly and easterly flows happened in summer. The LLJ structure in the nocturnal westerly mean profiles seen in the annual profiles at Wasco and Boardman was most pronounced in spring and summer, reflecting the more frequent occurrences of marine intrusions and gap flows during those seasons.

4. HRRR model validation versus Doppler lidar at three sites

High-resolution lidar measurements are a valuable resource for NWP model-profile validation. To illustrate the nature of the data and HRRR model information being used, Fig. 10 shows lidar time–height cross sections for the three sites for the two individual days of Fig. 5, 26 February and 28 June, along with the corresponding HRRR model cross sections for forecast lead-time hours 1 and 3. The bottom rows show time–height cross sections of the model-minus-lidar wind speed differences (model bias errors) from which model error statistics presented here were calculated.

Cross sections such as Fig. 10 provide insight into the nature of the flows and associated model errors on individual days. For example, for 26 February, the diminishing

of the easterly wintertime gap flow and the buildup of the weak-wind layer at Arlington and Boardman as noted in Fig. 5 were simulated by the model, but at Wasco the stronger winds persisted in the model after 0600 UTC, indicating that the top of the weak-wind layer was predicted to be below the elevation of Wasco during this period (Fig. 10, left panels). The timing of the wind down-ramp and the vertical structure of the departing westerlies were major sources of model error through the lowest km. In the other example, the diurnal-wind cycle on 28 June (Fig. 10, right panels) was captured by the model, but the model timing of onset and the inability of the model to maintain the observed strong winds below 400 m AGL after 0600 UTC were sources of error on this day. Banta et al. (2018b, 2019, manuscript submitted to *Mon. Wea. Rev.*) show that this error pattern is typical for this type of wind flow.

a. Model wind speed distributions

Annual distributions of wind speed from HRRR–NCEP and lidar measurements at a height of 100 m, close to the hub height of many currently used wind turbines, are shown in the top three rows of Fig. 11 for the three sites. These plots address an important question for the RA aspect of WE, namely, how well do
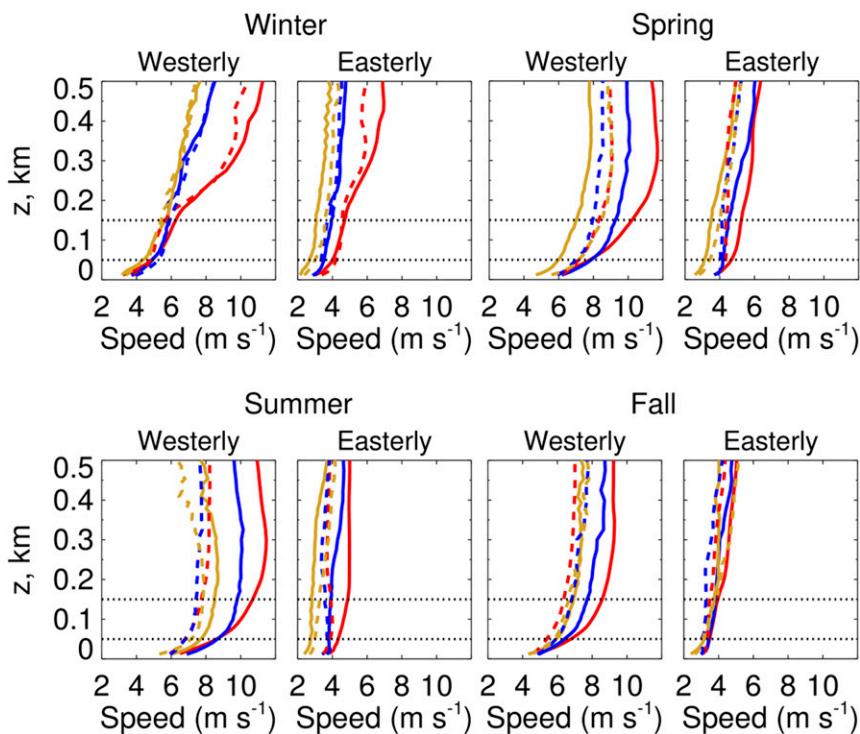
FIG. 9. Profiles of seasonally averaged wind speed in the first 500 m AGL, shown for westerly and easterly wind directions at three sites by lines and colors as in Fig. 8.

models predict the annual distribution of wind speeds, from which the projected annual energy production (AEP) is calculated. The HRRR plots show wind speeds from the 3-h forecast model output in blue superimposed on the red lidar-measured distributions at the three sites. The models captured the general shape of the annual wind speed distributions (Fig. 11, top left panels), including the bimodal distribution shapes at Boardman and Arlington, but underpredicted the occurrence of strong winds greater than $12 \, \mathrm{m \, s^{-1}}$ at the three sites. The bidirectional wind direction distribution was represented in the model for all sites; this was true for all forecast hours. Both modeled and measured westerly winds (Fig. 11, top right panels) were stronger and were over a broader range ($0–18 \, \mathrm{m \, s^{-1}}$) than the easterly winds, which were distributed over a range of $0–12 \, \mathrm{m \, s^{-1}}$ at all three sites. Site-to-site differences in model performance are most apparent as underpredictions of strong winds being largest at Arlington, underpredictions of the occurrence of weak ($<4 \, \mathrm{m \, s^{-1}}$) winds being largest at Wasco (mostly because of the winter season; Fig. 11, bottom panels), and overpredictions of weak winds ($<2 \, \mathrm{m \, s^{-1}}$) being largest at Boardman, also in winter.

Knowledge of the error in the hub-height wind speed forecasts is of critical economic importance for the calculation of energy produced by different types of

wind turbine. An example of power and AEP estimates of a hypothetical wind turbine are shown in Table 7. The power at each site was calculated using frequency distributions of measured and HRRR-simulated wind speeds at 80 and 100 m, and the reported power curve of a 1.5-MW General Electric (GE) wind turbine (GE Energy 2009), selected as one of the most widely deployed utility-scale turbines worldwide and in the study area.

See the appendix for annual distributions of computed power at the three sites (shown in Fig. A2). The last column in Table 7 shows normalized difference (%) between measured and modeled variables. The 100-m modeled wind speeds were too weak by 5.7%, 15%, and 17% at Wasco, Arlington, and Boardman, respectively, with similar values at 80 m (not shown). Because the model-lidar discrepancies can be significant over portions of the power curve that are sensitive to wind speed, the percentage power underpredictions can be larger than those for wind speeds: here they were about 10% too low at Wasco, 23% too low at Arlington, and 26% too low at Boardman. At Wasco the 5.7% differences in wind speed resulted from large HRRR overpredictions in winter compensated by underpredictions the rest of the year. The magnitudes were over such a range that the calculated power underpredictions were only about 10%. Thus the average wind speed errors, and even
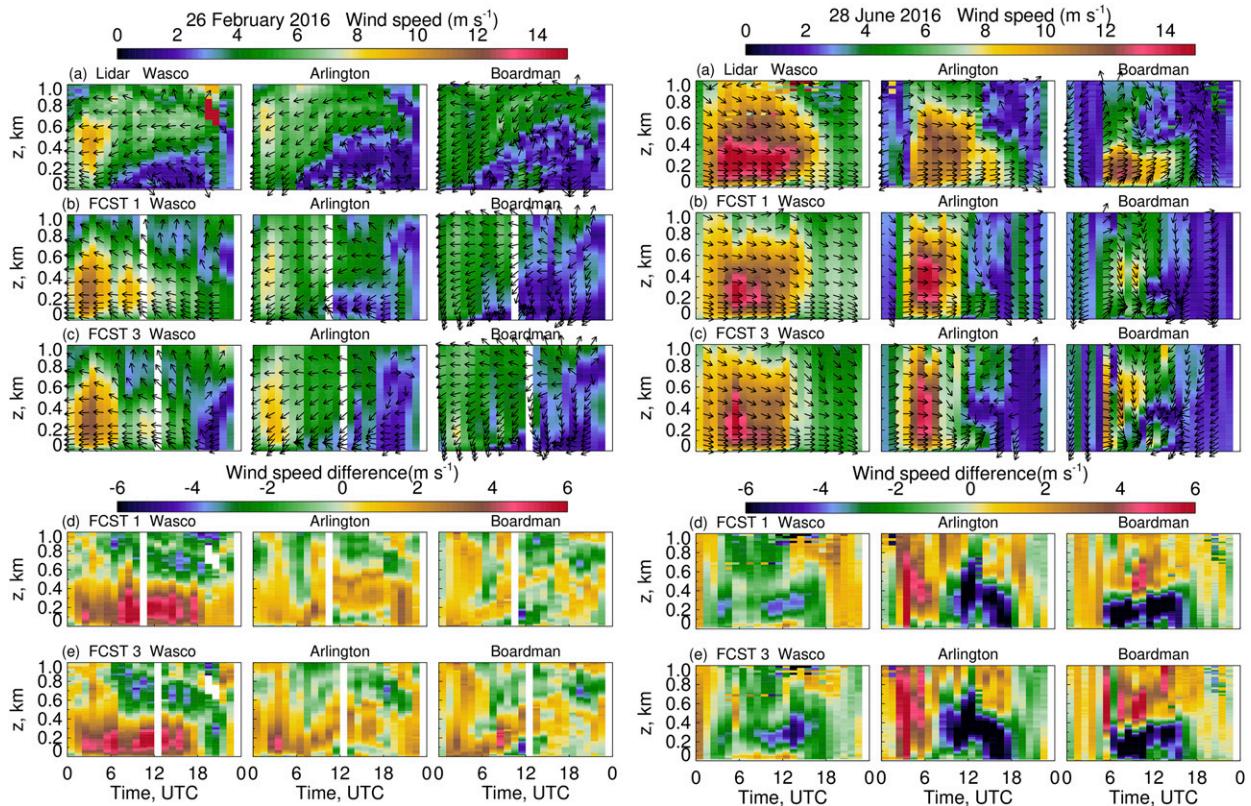
FIG. 10. Daily time–height cross sections of mean wind from (a) lidar and HRRR–NCEP forecast lead times (b) 1 and (c) 3 and wind speed difference for forecast lead-time hours (d) 1 and (e) 3 at each lidar site for (left) a cold-pool event on 26 Feb 2016 and (right) a diurnal-flow event on 28 Jun 2016.

more so the calculated AEP differences, depended critically on the location of the measurement in this complex-terrain setting, consistent with wind resource studies at much larger scales over several decades—for example, starting with Troen and Petersen (1989).

In the bottom three rows of Fig. 11 the distributions are given for each season for forecast hour 3. The underpredictions of strong winds seen in the annual distributions are shown to be mostly warm-season effects, the largest discrepancies occurring in summer. The underprediction of the occurrence of weak winds at Wasco, on the other hand, was mostly a wintertime effect, discussed further in the next section. The bidirectional distribution (not shown) was represented in the model for all seasons, as well as higher frequencies of the occurrence of the westerly-sector winds in the spring and summer seasons.

b. Validation of model mean-wind profiles

1) ANNUAL

The left column of Fig. 12 shows annually averaged wind speed profiles (black lines) through 500 m AGL

measured by the lidars at the three sites and those generated by the HRRR model at several forecast hours from 0 to 12 h. The 1- to 12-h predicted profiles did not differ much from each other, but the initial (red, hour 0) profiles differed from the others by as much as $1 \, \text{m s}^{-1}$, as might be expected because of imbalances in the initialized assimilation fields (Banta et al. 2018a); in general the initial profile had the largest discrepancies. The measured profile at Boardman up to 1 km (first 500 m shown here) indicated a broad LLJ structure that was not evident in the model mean profiles.

Vertical profiles of the HRRR error statistics for 2016 are shown in the other columns of Fig. 12. The panels in the second column of Fig. 12 show the annually averaged profiles of RMSE for wind speed through the lowest 500 m above the surface. The smallest RMSE values of $\sim 2 \, \text{m s}^{-1}$ occurred at the lowest points (approximately 20 m AGL), increasing with height to $4 \, \text{m s}^{-1}$ above 500 m. The remaining columns give the annually averaged bias, unbiased RMSE (RMSEub), and MAE of the model values versus the Doppler-lidar data, for reference. Biases tended to be negative with magnitudes less than $1 \, \text{m s}^{-1}$ below 500 m, except for a positive bias
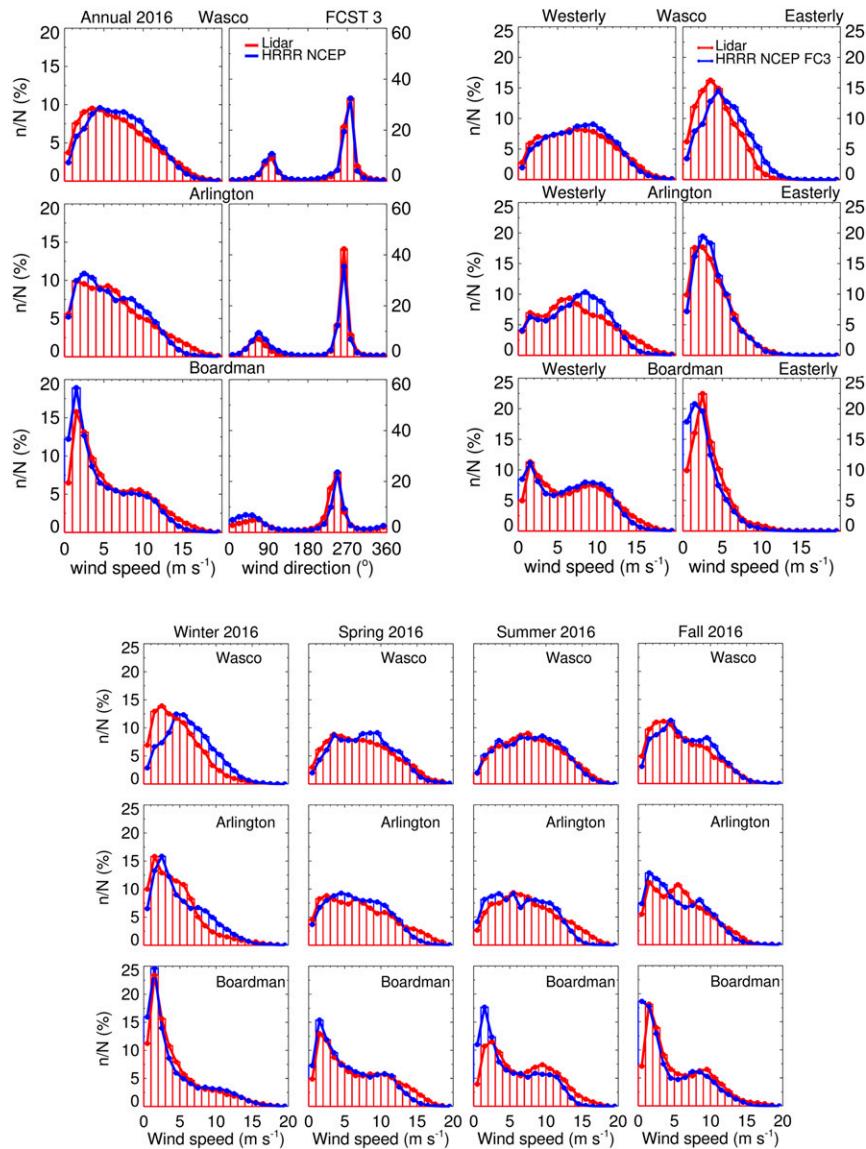
FIG. 11. Distributions of rotor-layer wind flow parameters for lead-time hour 3 at three sites (top left) for all of 2016 for all wind speeds and for all wind directions, (top right) for all of 2016 for wind speeds divided into westerly and easterly wind directions, and (bottom) by season for all wind speeds. Red color indicates 15-min lidar measurements at 98 m; blue color indicates hourly output from the HRRR–NCEP model at 100 m extracted at the location of each lidar by using a bilinear interpolation technique.

at Wasco, related to the overprediction of wintertime wind speeds noted in the seasonal distributions there. RMSEub were 1.5–2 m s^{-1} smaller than the total RMSE, reaching 2–2.5 m s^{-1} in the 200–500–m layer.

2) SEASONAL

Mean-wind and error profiles for the winter and summer seasons are shown in Figs. 13a and 13b, indicating the contributions of those seasons to the annual patterns. The underprediction of HRRR versus lidar-measured wind

speeds (low bias) was strongest in summer (Fig. 13b), especially at Arlington and Boardman, as a result of the underprediction of diurnal effects (Banta et al. 2018b, 2019, manuscript submitted to *Mon. Wea. Rev.*) and westerly gap flows. The mean measured profiles in summer show a mean LLJ structure not simulated by HRRR.

Model-predicted winds for winter (Fig. 13a) and fall (not shown) at Wasco were too strong below 400 m by up to 2 m s^{-1}. The cold-season overprediction of wind

TABLE 7. Mean values of annual distributions of wind speed from lidar measurements and HRRR–NCEP 3-h forecast, and error of annual power and annual energy production estimates at three sites.

Site		Lidar	Model	Diff between lidar and model	Normalized diff (%)
			Wind speed at 100 m		
Wasco	Wind speed (m s^{-1})	7.16	6.75	−0.41	5.73
	Power (MW)	558.88	503.34	−55.54	9.94
	AEP (MW h)	4895.81	4409.23	−486.58	9.94
Arlington	Wind speed (m s^{-1})	6.85	5.84	−1.01	14.74
	Power (MW)	517.67	398.63	−119.04	23.00
	AEP (MW h)	4534.78	3491.96	−1042.82	23.00
Boardman	Wind speed (m s^{-1})	5.89	4.92	−0.97	16.47
	Power (MW)	441.29	327.13	−114.16	25.87
	AEP (MW h)	3865.65	2865.67	−999.98	25.87
			Wind speed at 80 m		
Wasco	Wind speed (m s^{-1})	6.91	6.56	−0.35	5.07
	Power (MW)	529.73	473.99	−55.74	10.52
	AEP (MW h)	4640.45	4152.18	−488.27	10.52
Arlington	Wind speed (m s^{-1})	6.70	5.74	−0.96	14.33
	Power (MW)	498.88	380.85	−118.03	23.66
	AEP (MW h)	4370.21	3336.20	−1034.01	23.66
Boardman	Wind speed (m s^{-1})	5.67	4.66	−1.01	17.81
	Power (MW)	412.22	296.5	−115.72	28.07
	AEP (MW h)	3611.01	2597.37	−1013.64	28.07

speeds there, as noted in the previous section for the modeled wintertime distributions, thus occurred through a layer several hundred meters deep. A plausible explanation for these erroneously strong model winds is that the modeled depth of the wintertime cold pools was often too shallow. Predicting the cold-air-pool top to fall below the elevation of Wasco would expose that site to the stronger winds above the stagnant cold-pool layer, allowing higher momentum air from aloft to penetrate down to near the surface at Wasco but not at the two lower-elevation sites. The left panels of Fig. 10 show an example of these conditions (from 0600 to 0000 UTC). If this scenario occurred often enough, it would explain the winter peak in Wasco's modeled wind speed distribution (Fig. 11, third row from bottom) at 5–6 m s^{-1} instead of at 3 m s^{-1} as observed.

c. Validation by forecast lead time

Figure 14a shows the annually averaged model-verification statistics for wind speed at 100 m AGL as a function of forecast lead time. RMSE values were ~3 m s^{-1} for the first 2 h at this height, modestly increasing with lead time to ~3.5, 3.1, and 3.1 m s^{-1} at Wasco, Boardman, and Arlington, respectively. Thus, annually averaged model forecast skill did not degrade significantly with lead time, out to 15 h. This result is in contrast to findings over the Gulf of Maine, also using Doppler lidar and other instrumentation to validate

NCEP forecast models, including the HRRR-2012P version (Banta et al. 2018a; Pichugina et al. 2017b; Djalalova et al. 2016). These studies noted a distinct degradation of forecast skill as forecast lead times increased from 1 to 15 h.

The lack of degradation with forecast hour in the present study may result from constraints on the flow imposed by the complex topography. The flow is constrained by real topography in the atmosphere, but in the HRRR runs it is constrained by smoothed model topography, producing flow discrepancies that are more controlled by terrain misrepresentation on average than physics or numerical errors. Also, data assimilation may be adversely affected by unrepresentativeness of the measurement data being forced onto the model's terrain, contributing to errors in the early simulated fields. Such assimilation issues along with the topographical constraints could produce larger errors than would be seen over simple topography for early lead times, but the topographically controlled nature of the flows in both cases would not allow the errors to grow without constraint at longer lead times.

Bias plots indicate low biases in the winds at the lower two sites of less than 0.5 m s^{-1}, the largest magnitudes occurring at 5- and 6-h lead time. At Wasco, however, the speeds were biased high by approximately 1 m s^{-1}. The unbiased RMSE was similar in value to the total RMSE for all hours, consistent with the relatively small biases.
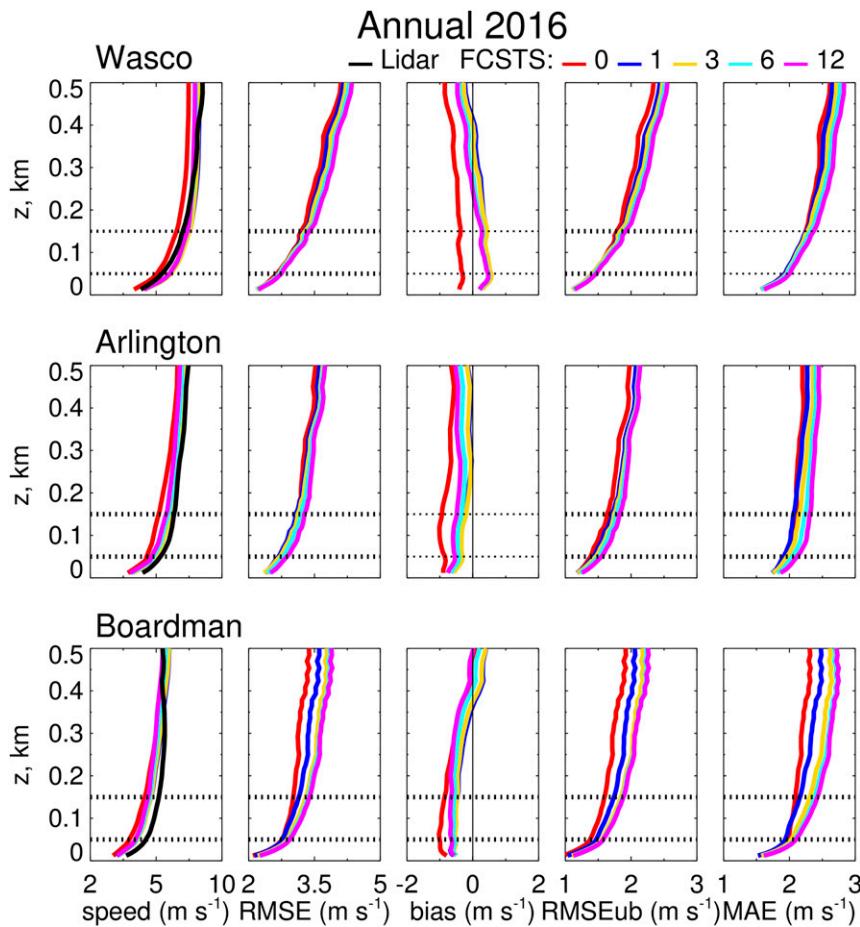
FIG. 12. Annually averaged profiles (black is lidar and colors are for HRRR–NCEP forecast hours 0–12) of (left) wind speed, (left center) RMSE, (center) bias, (right center) bias-corrected RMSE, and (right) MAE at (top) Wasco, (middle) Arlington, and (bottom) Boardman.

Figure 15 shows these values for the winter and summer seasons, which show the strongest contrasts. The most conspicuous results were the $2 \, \mathrm{m \, s^{-1}}$ positive biases at Wasco—2 times the annual value—concentrated in the winter (Fig. 15a) and autumn (not shown here) half year, and the $1 \, \mathrm{m \, s^{-1}}$ low biases at Arlington and Boardman in summer (Fig. 15c) and spring (not shown), again about 2 times the annual value. The large cold-season positive biases at Wasco are consistent with the other findings of wintertime winds being too strong in the model at Wasco.

RMSE's were smallest in winter, during which they varied from site to site and did increase somewhat with forecast lead time, from approximately 3.5 to $4 \, \mathrm{m \, s^{-1}}$ at Wasco and from 2.4 to $3.3 \, \mathrm{m \, s^{-1}}$ at Arlington and Boardman. The correlations R^2 were largest in summer (0.85–0.9) and spring (similar, not shown), indicating greater model skill in simulating warm-season phenomena, and were smallest in autumn and winter (0.55–0.65), when variations in wind speed resulting

from cold-season meteorological processes are of higher amplitude.

d. Validation by time of day

Figure 14b shows the annually averaged model error statistics as a function of hour of the day (midnight = 0800 UTC, and noon = 2000 UTC) for forecast hour 3, although all forecast hours behaved similarly. The positive bias at Wasco is evident during nighttime and early morning hours of 0300–1600 UTC. Negative biases at Arlington and Boardman were largest during the early morning hours between midnight and noon (0800 and 2000 UTC). The RMSEs were only weakly dependent on time of day, and R^2 correlation values were somewhat smaller during the morning transition (1500–1700 UTC).

Figures 15b and 15d show the time-of-day variations of model error for the winter and summer seasons. The high wintertime biases exceeding $1 \, \mathrm{m \, s^{-1}}$ at Wasco occurred mostly at night (0300–1500 UTC), accompanied

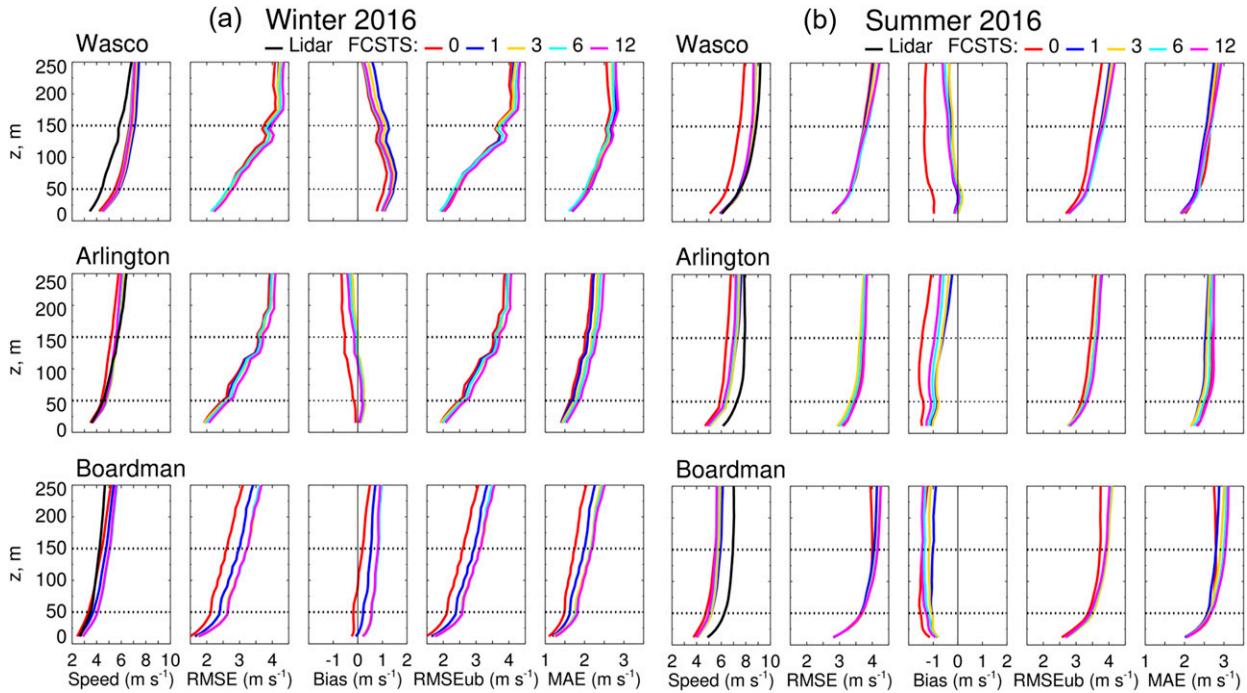FIG. 13. As in Fig. 12, but in the first 250 m AGL for the (a) winter and (b) summer seasons.

by large values of RMSE sometimes exceeding 4 m s^{-1}. The summertime negative biases at Arlington and Boardman had the largest magnitudes (at times greater than 2 m s^{-1}) at night, when diurnal westerly marine-intrusion-type flows would be occurring at those sites.

The diurnal behavior of model errors provides insight into the nature of those discrepancies. In winter, smaller daytime biases and other error values at Wasco may indicate a tendency for the cold pools in the atmosphere to become shallower during daylight hours, bringing model predictions into better agreement. This suggests
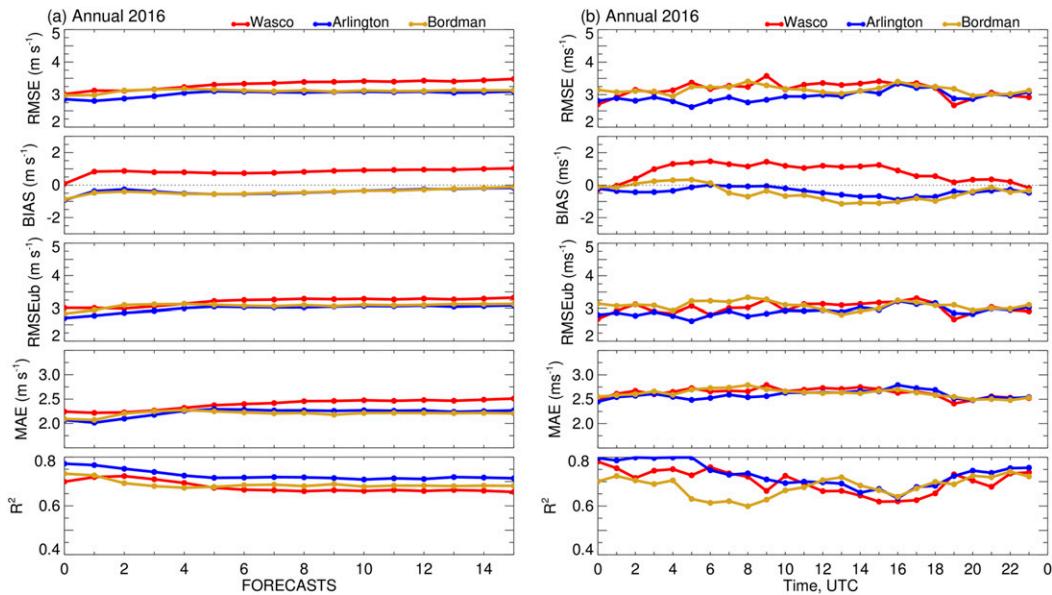
FIG. 14. Validation metrics between measured and modeled wind speed at three sites for the annual period January–December 2016 are shown as a function of (a) forecast hours and (b) time (UTC).
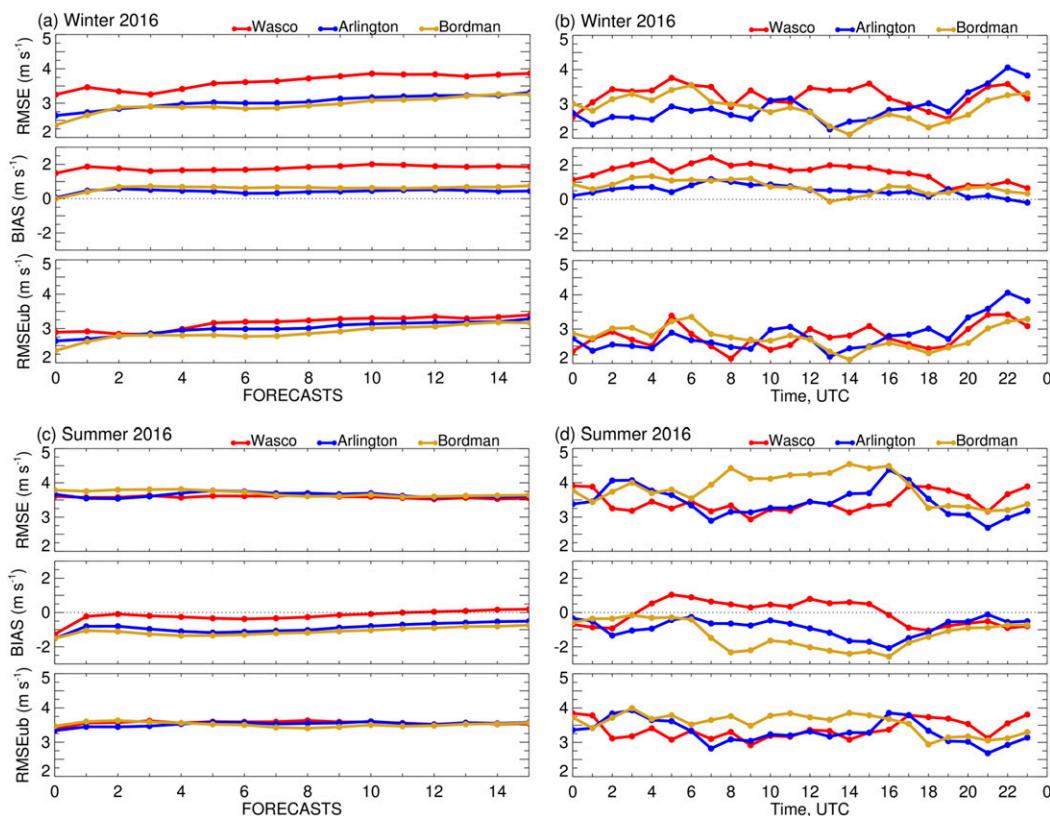
FIG. 15. As in Fig. 14, but for the (a),(b) winter and (c),(d) summer seasons.

that atmospheric processes leading to deeper cold pools tended to happen more often at night and were not matched by deeper predicted cold pools in the model. In summer, the large negative nocturnal biases at night at Arlington and Boardman are a consequence of model problems in reproducing flows that respond to the surface heating and cooling cycle. A study by Banta et al. (2019, manuscript submitted to *Mon. Wea. Rev.*) found that the models do predict the onset of these sea-breeze-type flows with some success, but fail to sustain the strong flows through the nighttime hours as observed, especially at Arlington and Boardman.

5. Conclusions

Fifteen-minute wind profiles from Doppler lidars deployed to three sites over a 71-km distance of similar terrain for a year allowed us to evaluate annually averaged wind characteristics at each site and differences in those characteristics among the sites. Seasonally, winds were weakest at all sites during winter.

Systematic differences were observed among the sites. The highest, westernmost site (Wasco) had the strongest winds, and the lowest, easternmost site (Boardman), the weakest, with a $1 \, \mathrm{m \, s^{-1}}$ difference

within the rotor layer, and $2 \, \mathrm{m \, s^{-1}}$ above 200 m AGL. Winds were strongest at night at two of the three sites. Histograms of wind speed showed significant differences among the sites, two of the sites exhibiting distinct bimodal distributions.

Data from the hourly operational HRRR runs were evaluated against the lidar data for the annual cycle. The model underestimated the annual mean 100-m wind speed, by $1 \, \mathrm{m \, s^{-1}}$ at two of the sites. Modeled wind speed distributions were also different from observed, underpredicting the strong winds ($>12 \, \mathrm{m \, s^{-1}}$) at all sites and weak winds ($<3 \, \mathrm{m \, s^{-1}}$) at Wasco. At a seasonal level, the underprediction of strong winds was strongest during summer and the underprediction of weak winds at Wasco was strongest in winter. Information from model error profiles, as well as inspection of data from individual days, such as lidar time–height cross sections, gives further insight into the model errors. Summertime underpredictions of wind speeds were associated with model errors in representing diurnal flows, and wintertime underestimates of weak winds at Wasco, which were most evident for easterly flow, were associated with errors in the depth of the persistent cold pools often present during the cold season. Annual model errors could thus be ascribed to seasonal errors, which
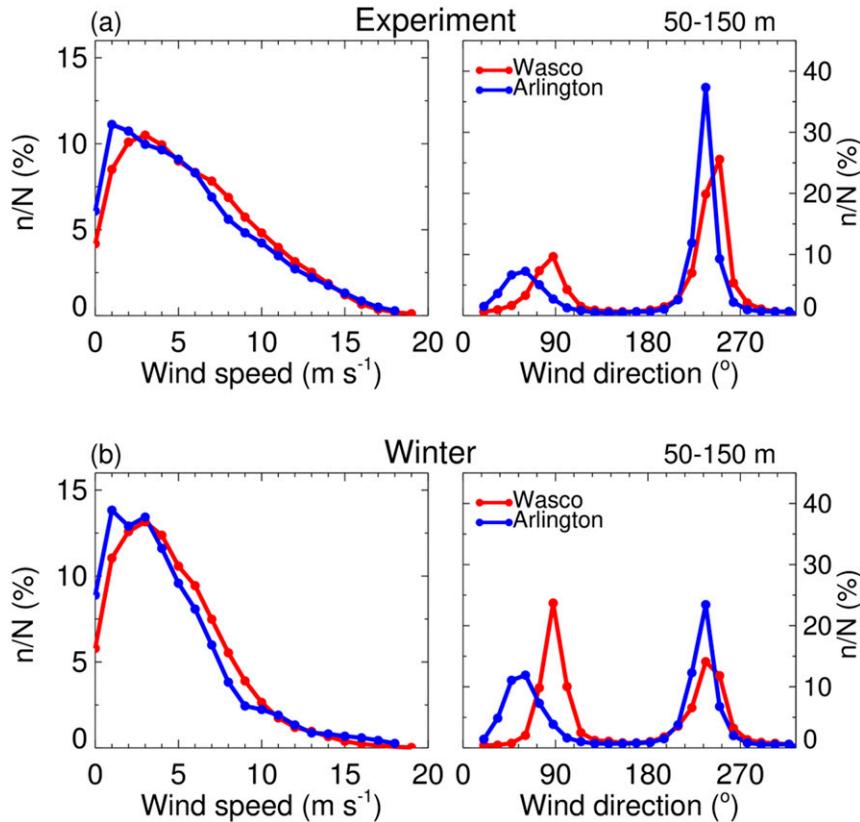
FIG. A1. (a) Experiment-long (September 2015–March 2017) and (b) winter-period (December 2015–February 2016) distributions of rotor-layer (50–150 m) wind (left) speed and (right) direction at the Wasco and Arlington sites.

in turn were traced to specific types of flow event that recurred during those seasons. These observations are potentially valuable clues as to where the models need improvement, here in the accuracy of simulated annual mean winds.

Model errors varied from site to site. Each site gave a different view as to where HRRR needs improvement. If only one of the sites had been instrumented for model evaluation, we would get a very different picture of the nature of the model errors in this region and where improvement was needed, depending on which of the sites was chosen to be instrumented. This finding argues for a high density of high-quality profile measurements in field programs designed to address WE and NWP improvement (Banta et al. 2013a), especially those in complex terrain.
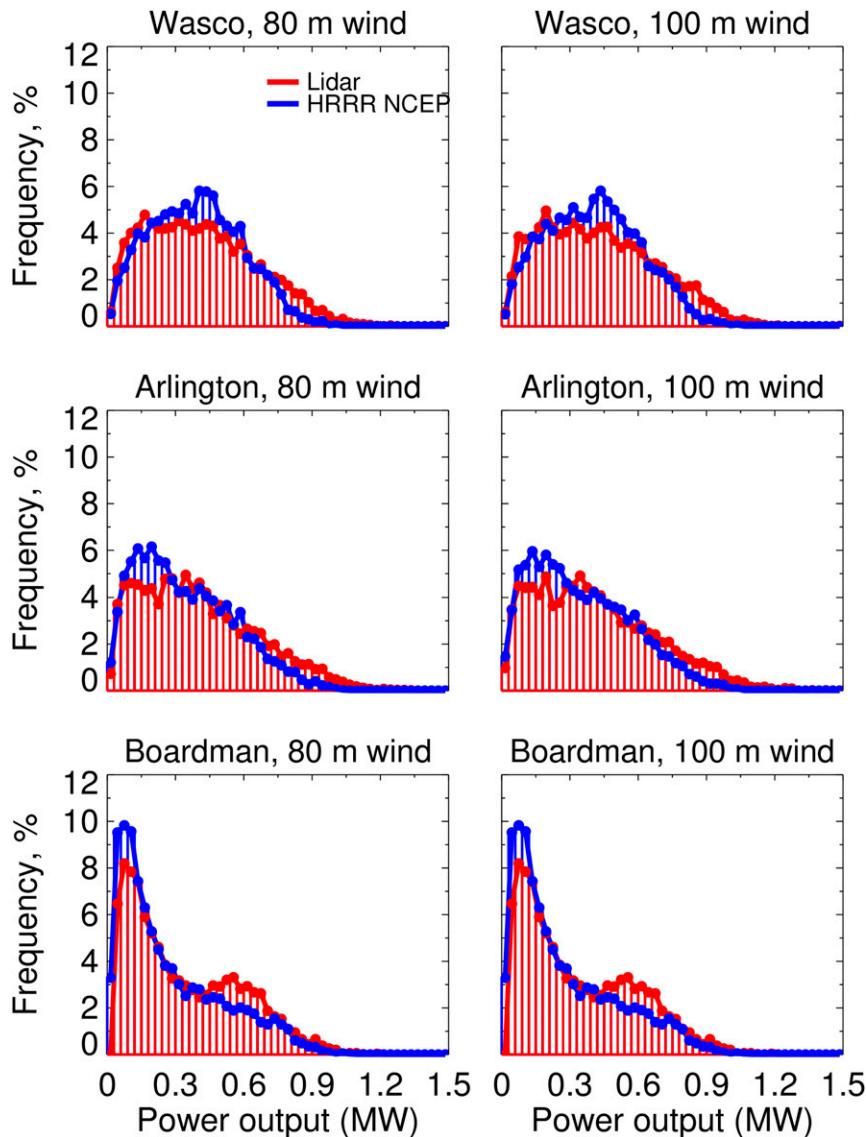
FIG. A2. Annual distributions of computed power at three sites using the power curve of a GE Model 1.5sle wind turbine and wind speed at (left) 80 and (right) 100 m. Red color indicates power from hourly averaged lidar-measured wind speed; blue color indicates hourly output from the HRRR–NCEP model for lead-time hour 3.

published form of this work or to allow others to do so, for U.S. government purposes.

APPENDIX

Additional Results

Distributions of rotor layer wind speed and direction for the duration of the experiment and winter season are shown in Fig. A1. Annual distributions of computed power from measured and modeled wind speed at three sites are shown in Fig. A2.

REFERENCES

Ahlstrom, M., and Coauthors, 2013: Knowledge is power. *IEEE Power Energy Mag.*, **11**, 45–52, https://doi.org/10.1109/MPE.2013.2277999.

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.

Banakh, V. A., W. A. Brewer, Y. L. Pichugina, and I. N. Smalikho, 2010: Measurements of wind velocity and direction with coherent Doppler lidar in conditions of a weak echo signal. *Atmos. Oceanic Opt.*, **23**, 381–388, https://doi.org/10.1134/S1024856010050076.

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 Meteor.*, **105**, 221–252, https://doi.org/10.1023/A:1019992330866.

——, Y. L. Pichugina, N. D. Kelley, R. M. Hardesty, and W. A. Brewer, 2013a: 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, 2013b: Observational techniques: Sampling the mountain atmosphere. *Mountain Weather Research and Forecasting*, F. K. Chow, S. De Wekker, and B. Snyder, Eds., Springer, 409–530, https://doi.org/10.1007/978-94-007-4098-3.

——, and Coauthors, 2015: 3D volumetric analysis of wind-turbine wake properties in the atmosphere using high-resolution Doppler lidar. *J. Atmos. Oceanic Technol.*, **32**, 904–914, https://doi.org/10.1175/JTECH-D-14-00078.1.

——, and Coauthors, 2018a: Evaluating and improving NWP forecast models for the future: How the needs of offshore wind energy can point the way. *Bull. Amer. Meteor. Soc.*, **99**, 1155–1176, https://doi.org/10.1175/BAMS-D-16-0310.1.

——, and Coauthors, 2018b: Observing and modeling recurrent diurnal summertime wind systems in the complex terrain of the Columbia River Basin during the Second Wind Forecast Improvement Program. *18th Conf. on Mountain Meteorology*, Santa Fe, NM, Amer. Meteor. Soc., 16.3, https://ams.confex.com/ams/18Mountain/webprogram/Paper346156.html.

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.

Berg, J., N. Vasiljević, M. Kelly, G. Lea, and M. Courtney, 2015: Addressing spatial variability of surface-layer wind with long-range WindScanners. *J. Atmos. Oceanic Technol.*, **32**, 518–527, https://doi.org/10.1175/JTECH-D-14-00123.1.

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.

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.

——, B. J. Carroll, R. M. Hardesty, W. A. Brewer, K. Hajny, O. E. Salmon, and P. B. Shepson, 2018: Doppler lidar observations of the mixing height in Indianapolis using an automated composite fuzzy logic approach. *J. Atmos. Oceanic Technol.*, **35**, 473–490, https://doi.org/10.1175/JTECH-D-17-0159.1.

Browning, K. A., and R. Wexler, 1968: The determination of kinematic properties of a wind field using Doppler radar. *J. Appl. Meteor.*, **7**, 105–113, https://doi.org/10.1175/1520-0450(1968)007<0105:TDOKPO>2.0.CO;2.

Djalalova, I. V., and Coauthors, 2016: The POWER experiment: Impact of assimilation of a network of coastal wind profiling radars on simulating offshore winds in and above the wind turbine layer. *Wea. Forecasting*, **31**, 1071–1091, https://doi.org/10.1175/WAF-D-15-0104.1.

Draxl, C., A. N. Hahmann, A. Pena, and G. Giebel, 2014: Evaluating winds and vertical shear from Weather Research and Forecasting model forecasts using seven planetary boundary layer schemes. *Wind Energy*, **17**, 39–55, https://doi.org/10.1002/we.1555.

Drechsel, S., G. J. Mayr, J. W. Messner, and R. Stauffer, 2012: Lower boundary wind speeds: Measurements and verification of forecasts. *J. Appl. Meteor. Climatol.*, **51**, 1602–1617, https://doi.org/10.1175/JAMC-D-11-0247.1.

Fast, J. D., and L. S. Darby, 2004: An evaluation of mesoscale model predictions of down-valley and canyon flows and their consequences using Doppler lidar measurements during VTMX 2000. *J. Appl. Meteor.*, **43**, 420–436, https://doi.org/10.1175/1520-0450(2004)043<0420:AEOMMP>2.0.CO;2.

Fernández-González, S., M. L. Martín, E. García-Ortega, A. Merino, J. Lorenzana, J. L. Sánchez, F. Valero, and J. S. Rodrigo, 2018: Sensitivity analysis of the WRF model: Wind-resource assessment for complex terrain. *J. Appl. Meteor. Climatol.*, **57**, 733–753, https://doi.org/10.1175/JAMC-D-17-0121.1.

Fernando, H. J. S., and J. C. Weil, 2010: Whither the stable boundary layer? A shift in research agenda. *Bull. Amer. Meteor. Soc.*, **91**, 1475–1484, https://doi.org/10.1175/2010BAMS2770.1.

GE Energy, 2009: 1.5MW wind turbine. General Electric Rep., 12 pp., http://geosci.uchicago.edu/~moyer/GEOS24705/Readings/GEA14954C15-MW-Broch.pdf.

Kelly, M., I. Troen, and H. E. Jørgensen, 2014: Weibull-*k* revisited: "Tall" profiles and height variation of wind statistics. *Bound.-Layer Meteor.*, **152**, 107–124, https://doi.org/10.1007/s10546-014-9915-5.

Klaas, T., L. Pauscher, and D. Callies, 2015: LiDAR-mast deviations in complex terrain and their simulation using CFD. *Meteor. Z.*, **24**, 591–603, https://doi.org/10.1127/metz/2015/0637.

Krishnamurthy, R., R. Calhoun, B. Billings, and J. Doyle, 2011: Wind turbulence estimates in a valley by coherent Doppler lidar. *Meteor. Appl.*, **18**, 361–371, https://doi.org/10.1002/met.263.

Krogsæter, O., and J. Reuder, 2015: Validation of boundary layer parameterization schemes in the weather research and forecasting model under the aspect of offshore wind energy applications—Part I: Average wind speed and wind shear. *Wind Energy*, **18**, 769–782, https://doi.org/10.1002/we.1727.

Lhermitte, R., and D. Atlas, 1961: Precipitation motion by pulse Doppler radar. Preprints, *Ninth Radar Meteorology Conf.*, Kansas City, MO, Amer. Meteor. Soc., 218–223.

Mann, J., and Coauthors, 2017: Complex terrain experiments in the New European Wind Atlas. *Philos. Trans. Roy. Soc.*, **A375**, 20160101, https://doi.org/10.1098/rsta.2016.0101.

——, R. Menke, N. Vasiljević, J. Berg, and N. Troldborg, 2018: Challenges in using scanning lidars to estimate wind resources in complex terrain. *J. Phys.: Conf. Ser.*, **1037**, 072017, https://doi.org/10.1088/1742-6596/1037/7/072017.

Marquis, M., J. Wilczak, M. Ahlstrom, J. Sharp, A. Stern, J. C. Smith, and S. Calvert, 2011: Forecasting the wind to reach significant penetration levels of wind energy. *Bull. Amer. Meteor. Soc.*, **92**, 1159–1171, https://doi.org/10.1175/2011BAMS3033.1.

Newsom, R. K., W. A. Brewer, J. M. Wilczak, D. E. Wolfe, S. P. Oncley, and J. K. Lundquist, 2017: Validating precision estimates in horizontal wind measurements from a Doppler lidar. *Atmos. Meas. Tech.*, **10**, 1229–1240, https://doi.org/10.5194/amt-10-1229-2017.

Olson, J., and Coauthors, 2019: Improving wind energy forecasting through numerical weather prediction model development.

Bull. Amer. Meteor. Soc., https://doi.org/10.1175/BAMS-D-18-0040.1, in press.

Pauscher, L., and Coauthors, 2016: An inter-comparison study of multi- and DBS lidar measurements in complex terrain. *Remote Sens.*, **8**, 782, https://doi.org/10.3390/rs8090782.

Peckham, S. E., T. G. Smirnova, S. G. Benjamin, J. M. Brown, and J. S. Kenyon, 2016: Implementation of a digital filter initialization in the WRF model and its application in the Rapid Refresh. *Mon. Wea. Rev.*, **144**, 99–106, https://doi.org/10.1175/MWR-D-15-0219.1.

Pichugina, Y. L., and R. M. Banta, 2010: Stable boundary-layer depth from high-resolution measurements of the mean wind profile. *J. Appl. Meteor. Climatol.*, **49**, 20–35, https://doi.org/10.1175/2009JAMC2168.1.

——, ——, N. D. Kelley, B. J. Jonkman, S. C. Tucker, R. K. Newsom, and W. A. Brewer, 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.

——, ——, W. A. Brewer, S. P. Sandberg, and R. M. Hardesty, 2012: Doppler lidar–based wind-profile measurement system for offshore wind-energy and other marine boundary layer applications. *J. Appl. Meteor. Climatol.*, **51**, 327–349, https://doi.org/10.1175/JAMC-D-11-040.1.

——, and Coauthors, 2017a: Properties of offshore low-level jet and rotor layer wind shear as measured by scanning Doppler lidar. *Wind Energy*, **20**, 987–1002, https://doi.org/10.1002/we.2075.

——, and Coauthors, 2017b: Assessment of NWP forecast models in simulating offshore winds through the lower boundary layer by measurements from a ship-based scanning Doppler lidar. *Mon. Wea. Rev.*, **145**, 4277–4301, https://doi.org/10.1175/MWR-D-16-0442.1.

Post, M. J., C. J. Grund, D. Wang, and T. Deschler, 1997: Evolution of Mount Pinatubo's aerosol size distributions over the continental United States: Two-wavelength lidar retrievals and in situ measurements. *J. Geophys. Res.*, **102**, 13 535–13–542, https://doi.org/10.1029/97JD00644.

Risan, A., J. A. Lund, C.-Y. Chang, and L. Sætran, 2018: Wind in complex terrain—Lidar measurements for evaluation of CFD simulations. *Remote Sens.*, **10**, 59, https://doi.org/10.3390/rs10010059.

Rye, B. J., and R. M. Hardesty, 1993: Discrete spectral peak estimation in incoherent backscatter heterodyne lidar. I. Spectral accumulation and the Cramer–Rao lower bound. *IEEE Trans. Geosci. Remote Sens.*, **31**, 16–27, https://doi.org/10.1109/36.210440.

Schreck, S., J. K. Lundquist, and W. Shaw, 2008: Research needs for wind resource characterization. U.S. Department of Energy Workshop/NREL Rep. TP-500-43521, 116 pp.

Seaman, N., 2000: Meteorological modeling for air-quality assessments. *Atmos. Environ.*, **34**, 2231–2259, https://doi.org/10.1016/S1352-2310(99)00466-5.

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/BAMS-83-12-1757.

——, and ——, 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., J. K. Lundquist, and S. J. Schreck, 2009: Research needs for wind resource characterization. *Bull. Amer. Meteor. Soc.*, **90**, 535–538, https://doi.org/10.1175/2008BAMS2729.1.

——, and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): General overview. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-18-0036.1, in press.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, https://doi.org/10.1175/MWR2830.1.

Smalikho, I. N., V. A. Banakh, Y. L. Pichugina, and W. A. Brewer, 2013: Accuracy of estimation of the turbulent energy dissipation rate from wind measurements with a conically scanning pulsed coherent Doppler lidar. Part II. Numerical and atmospheric experiments. *Atmos. Oceanic Opt.*, **26**, 411–416, https://doi.org/10.1134/S1024856013050151.

Troen, I., and E. L. Petersen, 1989: *European Wind Atlas*. Risø National Laboratory, 656 pp.

Whiteman, C. D., S. Zhong, W. J. Shaw, J. M. Hubbe, X. Bian, and J. Mittelstadt, 2001: Cold pools in the Columbia basin. *Wea. Forecasting*, **16**, 432–447, https://doi.org/10.1175/1520-0434(2001)016<0432:CPITCB>2.0.CO;2.

Wilczak, J. M., and Coauthors, 2015: The Wind Forecast Improvement Project (WFIP): A public–private partnership addressing wind energy forecast needs. *Bull. Amer. Meteor. Soc.*, **96**, 1699–1718, https://doi.org/10.1175/BAMS-D-14-00107.1.

——, and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): Observational field campaign. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-18-0035.1, in press.

Yang, B., and Coauthors, 2017: Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the weather research and forecasting model. *Bound.-Layer Meteor.*, **162**, 117–142, https://doi.org/10.1007/s10546-016-0185-2.

Zhong, S., and J. D. Fast, 2003: An evaluation of the MM5, RAMS, and Meso-Eta models at subkilometer resolution using field campaign data in the Salt Lake Valley. *Mon. Wea. Rev.*, **131**, 1301–1322, https://doi.org/10.1175/1520-0493(2003)131<1301:AEOTMR>2.0.CO;2.

——, C. D. Whiteman, X. Bian, W. J. Shaw, and J. M. Hubbe, 2001: Meteorological processes affecting the evolution of a wintertime cold-air pool in the Columbia basin. *Mon. Wea. Rev.*, **129**, 2600–2613, https://doi.org/10.1175/1520-0493(2001)129<2600:MPATEO>2.0.CO;2.