09 January 2024 21:36:56

1 **Evaluating the WFIP2 updates to the HRRR model using scanning Doppler lidar**

2 **measurements in the complex terrain of the Columbia River Basin**

3 Yelena L. Pichugina[1, 2], Robert M. Banta[1, 2], W. Alan Brewer[2], L. Bianco[1, 4], C. Draxl[5], J. Kenyon[1, 3],

4 J. K. Lundquist[6, 7], J.B. Olson[1], D.D. Turner[3], S. Wharton[11], J. Wilczak[4], S. Baidar[1, 2],

5 L.K. Berg[5], H.J.S. Fernando[8], B.J. McCarty[1, 2], R. Rai[5], B. Roberts[6],

6 J. Sharp[9], W.J. Shaw[5], M.J. Stoelinga[10], and R. Worsnop[7]

7 [1]CIRES, University of Colorado, Boulder, CO 80309, USA

8 [2]NOAAChemical Sciences Laboratory, Boulder, CO 80305, USA

9 [3]NOAA Global Systems Laboratory, Boulder, CO 80305 USA

10 [4]NOAA Physical Sciences Laboratory, Boulder, CO 80305, USA

11 [5]Pacific Northwest National Laboratory, Richland, WA 99352, USA

12 [6]National Renewable Energy Laboratory, Golden, CO 80401, USA

13 [7]University of Boulder, Boulder, CO 80309, USA

14 [8]University of Notre Dame, Notre Dame, IN 46556, USA

15 [9]Sharply Focused LLC, Portland, OR 97213, USA

16 [10]Vaisala, Seattle, WA 98121, USA.

17 [11]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

18 *Correspondence to*: Yelena Pichugina (Yelena.Pichugina@noaa.gov)

19   **Abstract.** The wind-energy (WE) industry relies on numerical weather prediction (NWP) forecast models as

20 foundational or base models for many purposes, including wind-resource assessment and wind-power forecasting.

21 During the Second Wind Forecast Improvement Project (WFIP2) in the Columbia River Basin of Oregon and

22 Washington, a significant effort was made to improve NWP forecasts through focused model development, to include

23 experimental refinements to the High Resolution Rapid Refresh (HRRR) model physics and horizontal grid spacing.

24 In this study, the performance of an experimental version of HRRR that includes these refinements is tested against a

25 control version, which corresponds to that of the operational HRRR run by NOAA/NCEP at the outset of WFIP2. The

26 effects of horizontal grid resolution were also tested by comparing wind forecasts from the HRRR (with 3-km grid

27 spacing) with those from a finer-resolution HRRR nest with 750-m grid spacing. Model forecasts are validated against

28 accurate wind-profile measurements by three scanning, pulsed Doppler lidars at sites separated by a total distance of

29 71 km. Model skill, and improvements in model skill, attributable to physics refinements and improved horizontal grid

30 resolution varied by season, by site, and during periods of atmospheric phenomena relevant to WE. In general, model

31 errors were the largest below 150 m AGL. Experimental HRRR refinements tended to reduce the mean absolute error

32 (MAE) and other error metrics for many conditions, but degradation in skill (increased MAE) was noted below 150 m

09 January 2024 21:36:56

33   AGL at the two lowest-elevation sites at night. Finer resolution was found to produce the most significant reductions

34   in the error metrics.

35

36

1 Introduction

38   Numerical weather prediction (NWP) forecast models are widely used in the wind energy (WE) industry as

39   foundational or base models for wind resource assessment and wind-power forecasting. Continual changes to the base

40   forecast models, by implementing updated model physics, parameterization schemes, and horizontal grid spacing, are

41   performed to improve model skill. Evaluating and quantifying the effects of these updates on model performance

42   require accurate measurements of atmospheric variables such as wind speed, wind direction, and turbulence (Bonin et

43   al. 2017) in various landscapes and atmospheric conditions.

44   The performance of mesoscale NWP models has been evaluated using measurements above the surface from

45   various instruments during short-term field experiments over complex terrain (Fast and Darby 2004, Yang et al. 2017,

46   Zhong and Fast, 2003), coastal areas (Draxl et al. 2014), and offshore (Drechsel et al. 2012, Krogsæter and J. Reuder

47   2015, Pichugina et al. 2017, Banta et al. 2018) documenting the relative strengths and weaknesses of model forecasts

48   at a horizontal grid spacing of 1 km or less.

49   In this study, the performance of the High Resolution Rapid Refresh (HRRR) and a HRRR nest (HRRRNEST),

50   developed at the Global Systems Laboratory (GSL) of the National Oceanic and Atmospheric Administration (NOAA)

51   /Earth Systems Research Laboratories (ESRL), is investigated by measurements from scanning, pulsed Doppler lidars

52   during the Second Wind Forecast Improvement Project (WFIP2) in the complex terrain of the Columbia River Valley

53   in Oregon and Washington in the United States. This 18-month long (September 2015 to April 2017), multi-institutional

54   collaborative project aimed to improve forecasting of wind flow complicated by mountainous terrain, coastal effects,

55   and the presence of numerous wind farms in this area. Overviews of the WFIP2 campaign describe the experiment and

56   scientific objectives (Shaw et al., 2019), instrumentation involved in the experiment and the logistics of their

57   deployments (Wilczak et al. 2019), and the model-physics updates made during WFIP2 (Olson et al. 2019).

58   WFIP2 measurements from three scanning lidars separated from each other by 30-40 km, sited at different terrain

59   elevations over a total distance of 71 km, were used previously (Pichugina et al., 2019) to analyze temporal and spatial

60   variability of the wind flow and to evaluate the performance of the operational HRRR-NCEP (National Centers for

61   Environmental Protection) model for the entire 18-month period. Pichugina et al. (2019) found significant differences

62   in the characteristics of model errors between seasons, over the diurnal cycle at each site and between sites. In winter,

63   the highest-elevation lidar site (Wasco, Oregon) showed anomalously high model wind speeds, which were attributed

64   to overly shallow predicted cold pools. In summer, the two lower-elevation sites (Arlington and Boardman, Oregon)

65   showed systematic low biases in the simulated wind-turbine rotor-layer wind speeds at night, due to a premature end

66   to a diurnal, sea-breeze related, marine-intrusion flow (Banta et al. 2020). Both of these error types influenced the

67   annual error statistics.

68   The HRRR-NCEP versions validated by Pichugina et al. (2019) were run operationally during WFIP2—version

69   1 of the HRRR-NCEP (HRRRv1) until 23 August 2016, then version 2 of the HRRR-NCEP (HRRRv2) afterward—as

70   described by Benjamin et al. (2016) and Olson et al. (2019). By the end of WFIP2, NOAA/ESRL had developed

71   experimental refinements for its nonoperational version of the HRRR.

72   For the present study, a provisional HRRR model was run in two domain configurations: a 'parent' HRRR with

73   standard 3-km horizontal grid spacing, and an experimental, finer-resolution, HRRR nest with 750-m horizontal grid

74   spacing. Additionally, both domain configurations were run with a control physics configuration and again with an

75   experimental physics configuration. The control configuration represents the state of HRRR physics at the beginning

76   of WFIP2 (September 2015), which also corresponds to the physics of HRRRv1, whereas the experimental

77   configuration represents physics developments made during WFIP2 (Olson et al. 2019, Bianco et al. 2019), to be

78   discussed in a later section.

79   We use the Doppler lidar measurements at the three sites to see how these updates affected model skill season

80   by season, site by site, and over periods with atmospheric phenomena relevant to WE, and to determine the effects of

81   finer resolution on model performance. Evaluating improvement in model skill involves comparing model errors from

82   one version of the model to the next. This evaluation requires accurate measurements with known levels of uncertainty,

83   against which to validate each model version. Here we exploit the accuracy of wind-speed profile measurements from

84   the Doppler lidars to perform these comparisons (Banta et al. 2013). The analyses quantify model error as related to

85   height, diurnal cycle, and season, and evaluate the version-to-version differences.

86   The paper is organized as follows: Section 2 describes the location of lidars and provides a short discussion of

87   the models used in the study and wind flow complexity in the WFIP2 research area. Section 2 also describes the periods

88   selected for model reforecasts and shows representativeness of the atmospheric phenomena observed during reforecast

89   periods relative to 2016 and to the entire length of the WFIP2 field campaign (Sep 2015-Mar 2017). This section also

90   illustrates signatures of atmospheric events (westerly Gap Flow and Cold Pool events) detected in lidar time-height

91   cross-sections over diurnal periods and in the composites over all days of the particular event observed during 2016.

92   Section 3 provides an overview of annual and seasonal wind flows in the lowest 1 km above ground level (AGL) from

93   lidar measurements and two runs of both models and presents annually/seasonally averaged profiles of validation

94   metrics for both models. Section 3 also focuses on the evaluation of model errors at 80 m AGL and presents analyses

95   for different wind directions and diurnal periods, and presents analyses of model performance during the most frequent

96   atmospheric events observed in the research area and presents model error dependency on both event type and

97   instrument location. Section 4 contains a discussion of the implications of these findings to model evaluation using

98   field data. Section 5 presents a summary and conclusions.

99   **2 Methods and forecast validation setup**

100   **2.1 Scanning Doppler lidars**

101   Two NOAA scanning, pulsed Doppler lidars (Leosphere WindCube 200S) were deployed 40 km apart near the

102   Wasco and Arlington, Oregon airports, providing real-time wind measurements from September 2015 through April

103   2017. The third scanning Doppler lidar (Halo Streamline XR) was deployed by the University of Notre Dame (UND)

09 January 2024 21:36:56

104 to a site near Boardman, 31 km east-northeast of Arlington (Fig. 1), and continuously operated from January through

105 December 2016. These instruments provided concurrent measurements of wind flow, each using similar scanning

106 sequences and data-processing techniques. They were located along a west-southwest/east-northeast line within a high-

107 density (~2,000 turbines) "wind-energy corridor" region of the Columbia River Basin (Banta et al. 2020).

108 Details of the lidars used during WFIP2 are described in Pichugina et al. (2019). These include lidar operational

109 parameters, details of scanning pattern, data quality control, and post-processing techniques used to obtain vertical

110 profiles of the mean wind variables (wind speed and direction) using line-of-sight (LOS) velocity measurements from

111 the near-surface up to 3.5 km. An assessment of the scanning Doppler lidar random measurement uncertainty of wind

112 speed in the complex terrain of WFIP2 (Pichugina et al. 2019) showed significant diurnal variability with a peak value

113 of ~0.02 m s^{-1} around local noon as well as dependence on terrain complexity for each site (with larger uncertainty at

114 the Wasco site compared to the lower-elevation Boardman site). Overall, the lidar measurement uncertainty was

115 determined to be small enough to quantify the accuracy of the operational HRRR-NCEP model to properly forecast

116 challenging wind-flow conditions in WFIP2. These accuracies are also sufficient for the comparison of different model

117 versions.

2.2 NWP models used in the study

119 The RAP and HRRR models are hourly updating, operational forecast models developed by researchers at

120 NOAA/GSL, and which employ the Advanced Research version of the WRF Model (WRF-ARW; Skamarock et al.

121 2008). Benjamin et al. (2016) describe the evolution of the RAP and HRRR models through 2016, and Olson et al.

122 (2019) describe the RAP and HRRR model-development efforts undertaken during WFIP2.

123 To support the goals of WFIP2 using the limited computational resources available to the project, a

124 nonoperational version of the HRRR was used for WFIP2 (Olson et al. 2019). This provisional WFIP2 HRRR

125 (hereafter simply "HRRR", unless otherwise noted) encompassed a smaller domain than its operational counterpart, as

126 shown by the large green box in Fig. 2a, but retained the standard 3-km horizontal grid spacing. Within the HRRR, a

127 nested domain (hereafter "HRRRNEST"), shown by the small green box in Fig. 2a, was run at 750-m horizontal grid

128 spacing. HRRR forecasts employed a so-called "cold start" initialization, where initial conditions were supplied from

129 the operational RAP without additional data assimilation or prior cycling, as described by Olson et al. (2019, pp. 2204–

130 2205). In turn, the HRRRNEST initializations were delayed 3 h after the HRRR initialization (see Bianco et al. 2019)

131 to minimize "spin up" problems at sub-kilometer scales. The HRRRR was initialized twice daily at 0000 and 1200

132 UTC (lagged 3 h by the HRRRNEST), and both the HRRR and HRRRNEST provided 24-h forecasts at 15-min output

133 intervals. The collection of HRRR and HRRRNEST forecasts used in this study are referred to as *reforecast* runs.

134 Table 7 (Appendix A) summarizes key aspects of the model design used for reforecasts.

135 During WFIP2, significant refinements were made to the HRRR and HRRRNEST, intended to improve how

136 these models represent complex terrain, the vertical mixing between the surface and the atmosphere, and the impact of

137 turbulence in the horizontal as well as vertical directions (Olson et al. 2019). To evaluate these improvements, HRRR

138 and HRRRNEST reforecasts for selected periods, to be discussed in the next subsection, were produced using a control

139 configuration (CNTR), which contained the parameterizations and other model routines present at the beginning of

140 WFIP2 (Olson et al 2019), and again using an experimental configuration (EXPR), which contained all physics

141 modifications developed during WFIP2. Table 8 summarizes key aspects of the CNTRL and EXPR configurations;
142 additional details are available in Olson et al. (2019). Reforecasts from both configurations were validated against
143 wind-profile measurements at three Doppler lidar sites to assess the impacts of these updates on model skill at locations
144 having different terrain complexity and wind-farm effects on wind flows (Fig1b). To assess the mean model skill over
145 the extended (~70 km) area, including the net impact of terrain and wind turbines there, the measured and modeled
146 profiles were averaged over the three lidar sites, referred in this paper as three-site composites.

147   An example of HRRRNEST-modeled, 80-m winds superimposed on the area topography (Fig. 2b) illustrates
148 the complexity of wind flow in the research area. The figure shows increased westerly winds through the narrow
149 Columbia River Gorge, (the only sea-level gap through this range [Sharp and Mass 2002]), driven by higher pressure
150 offshore to the west of the Cascade Mountains, as well as vortices of wind speed and direction caused by the mountain
151 ridge and local topography. Terrain complexity expressed as the standard deviation of the terrain elevations (SDE)
152 within a 3-km radius of each instrument location (Ascione et al. 2008), is largest for the Arlington site (71.1 m) followed
153 by 26.3 m for Wasco, and 18.6 m at Boardman (Pichugina et al. 2019). The lidar deployment locations and terrain
154 elevation of the surrounding area relative to a portion of the horizontal grid of each model (HRRR and HRRRNEST)
155 are shown in Fig. 2c.

156   Data from the three Doppler lidars were not assimilated into the model runs, so the lidar wind profiles provide
157 an independent validation dataset. For validation of models against lidar measurements, the gridded model wind values
158 from control and experimental runs of both models were extracted at the position of the lidar either by bi-linear
159 interpolation of winds from the surrounding four grid points or by using modeled winds from the nearest grid point
160 (Appendix B). Both methods show very similar results for each lidar location (Fig. 18; Table 4), with correlation
161 coefficients of 0.99 between the two extraction techniques, differences in mean wind speed of 0.01 to 0.22 m s^{-1}, and
162 standard deviations of 0.0 to 0.085 m s^{-1}. The offset differs from 0.02 to 0.08 m s^{-1} between the results and the slope
163 varies from 0.97 to 1.02 between the lowest and the highest value. Here, we use bilinear interpolation. For quantitative
164 comparisons of wind-speed profiles, the modeled values obtained at the location of each lidar were then linearly
165 interpolated to the heights of lidar measurement. The effects of the vertical-interpolation method and uncertainties of
166 two approaches—first, when measurements are interpolated to model output levels (lidar-to-model), and second, when
167 modeled variables are interpolated to the heights of lidar measurements (model-to-lidar)—are discussed in Pichugina
168 et al. (2017). This interpolation method "allows the fine-scale structure of the lidar data, with such features as thin shear
169 layers and LLJ noses, to be accounted for in the error computations." The LOS lidar measurements were then averaged
170 over vertical bins of 10 m in the lowest 200 m AGL, gradually increasing to 25 m up to 1 km AGL, 50 m up to 2 km
171 m AGL, and 100 m up to 5 km AGL.

2.3 Representativeness of the reforecast periods.

173   In addition to the assessment of model errors for different seasons and at different terrain complexity,
174 understanding of model performance during frequently occurring atmospheric phenomena is important for wind energy
175 operations. A variety of synoptic conditions were observed over the study area and were cataloged in an Event Log
176 (Draxl et al. 2019, Shaw et al. 2019, Wilczak et al. 2019) by a group of meteorologists after review of all available
177 WFIP2 observations, satellite imagery, real-time simulations from all WFIP2 models, and discussion of synopses of

09 January 2024 21:36:56

178   the daily weather. The determination of event type involved subjectivity associated with the identification process.

179   Besides, an event was assigned a category only if it was observed to occur over a major portion of the study area. An

180   overview of the mechanism and frequency of the events observed throughout the WFIP2 in the study area appears in

181   Wilczak et al. (2019).

182   Recent WFIP2 studies provide in-depth analyses of key phenomena to understand how well the HRRR

183   simulated individual flow types and how periods of these events may impact aggregate wind power generation within

184   the Bonneville Power Authority (BPA) balancing area

185   https://transmission.bpa.gov/Business/Operations/Wind/twndbspt.aspx).

186   These detailed studies include cold-pool periods (McCaffrey et al. 2019, Whiteman et al. 2001, Zhong et al.

187   2001), thermodynamically-driven westerly gap flows (Banta et al. 2020), and mountain waves and topographic wakes

188   (Draxl et al. 2020).

189   Two studies have used the WFIP2 dataset to evaluate the updates in the EXPR versions *vs.* the CNTR versions.

190   Olson et al. (2019) and Bianco et al. (2019) used 80-m AGL wind-speed data averaged over 22 sites, 19 equipped with

191   Doppler sodars, 2 with profiling lidars, and one with a scanning lidar, to calculate model-error and improvement

192   statistics including diurnal time series (for reference, the locations of these 22 sites are shown in Section 4). The Olson

193   et al. (2019) study also analyzed time-height cross-sections of wind speed averaged over eight 915-MHz wind-profiling

194   radars (WPRs). The studies found a strong diurnal behavior of the model errors, the largest errors occurring at nighttime.

195   Large positive biases were found during winter at higher-elevation sites (including Wasco) due to the modeled cold-

196   pool depths being too shallow. Significant low biases and large mean absolute error (MAE) occurred in summer at

197   night, due to the models not properly representing marine-intrusion type gap flows (Banta et al. 2020). These findings

198   all confirmed that the results and interpretations of Pichugina et al. (2019) using HRRR-NCEP model error statistics

199   averaged over the entire WFIP2 study period apply equally to the CNTR and EXPR reforecast runs. Olson et al. (2019)

200   and Bianco et al. (2019) showed that the updates to HRRR in EXPR produced significant improvements in winter

201   because of better predictions of cold-pool depth, but skill decreased in summer and spring at night. In Bianco et al.

202   (2019) the EXPR run updates were found to improve forecast skill by 0.1 to 0.2 m s^{-1} at 80 m, when averaged over all

203   four reforecast periods and all 22 locations, and error profile data from the WPRs indicated similar improvements

204   through a 200- to 600-m layer.

205   Reforecast periods for model runs were selected to provide a reasonable balance between computational expense

206   and coverage of consequential meteorological event types. These periods represent six weeks during each season, as in

207   Olson et al. (2019) and Bianco et al. (2019). An additional three weeks of simulation in August 2016 was added for a

208   detailed study of summertime thermally forced westerly gap-flow events (Banta et al. 2020, Olson et al. 2019). Analysis

209   of this additional model-output period is included in the present investigation.

210   Overall, the dataset for the reforecast period covers fewer days in the year than the operational HRRR-NCEP

211   dataset used by Pichugina et al. (2019), but the 15-min output of these models, along with the finer grid resolution of

212   the HRRRNEST, provide opportunities for more detailed analyses of wind-flow dynamics compared to the hourly

213   output that was saved from the operational model. The major error characteristics identified in the HRRR-NCEP by

214   Pichugina et al. (2019) are also found in the HRRR-CNTR reforecast runs, as just described.

215   Distributions of the observed events by event type for the whole experiment, for the annual period of 2016 (the

216   year of concurrent lidar measurements from all three lidars), and for the reforecast periods selected in 2016 are shown

217   in Fig. 3a. This figure establishes that each of the major events occurred during the selected reforecast periods.

218   The monthly distribution of events during reforecast periods (Fig. 3b) shows that westerly gap flows occurred

219   every month. Two types of westerly gap-flow forcing were identified during the event logging, one due to the synoptic

220   cross-barrier pressure gradient, and the other due to the diurnal heating-cooling cycle, such as the marine-intrusion

221   flows investigated by Banta et al. (2020). The westerly Gap flows occurred more often during warmer months of the

222   year (April-September) when a large fraction of these flows were of the thermally forced type (Fig. 3b). Cold-pool

223   events occurred mostly during winter and fall months, (McCaffrey et al. 2019). The frequent occurrence of westerly

224   gap flow (Fig. 3a, b) agrees with the regional prevalence of westerly winds in the annual and seasonal (except winter)

225   distributions of wind direction found by Pichugina et al. (2019) from both lidar observations and HRRR-NCEP output

226   of 80-m and rotor-layer winds.

227   The representativeness (%) of each event type during the reforecast periods is given in Table 1 by comparison

228   to the total number of events observed in 2016 and during 18 months of WFIP2. The relative frequency of each event

229   type over all reforecast periods was similar to the frequency observed for the year and for the WFIP2 period, except

230   for cold-pool events, which (at 9.2% of all reforecast events) were underrepresented in the reforecast sample.

231   **2.4 Signatures of atmospheric events in lidar observations**

232   Major event types captured by lidar measurements at a 15-min temporal resolution during the reforecast period were

233   cold pools, westerly gap flows, and easterly gap flows. The daily evolution of wind flow at each site (Fig. 4) for selected

234   days illustrates major signatures of these atmospheric event types as follow:

235   *Cold pool event on 26 February*. Weak easterly winds associated with the cold pool event occur throughout most

236   of the day. At all sites, average wind speed below 200 m slowed to <2 m s^{-1} by 0500 UTC and remained low for hours.

237   Note that this is below the average cut-in (~3-4 m/s) wind speed (i.e., minimum wind speed for power generation) for

238   most wind turbines.

239   *Westerly Gap Flow on 30 June*. Predominant westerly winds were associated with a marine intrusion event. Strong

240   nocturnal wind speeds (12-15 m s^{-1} in the rotor layer) decreased to 6-10 m s^{-1} for the rest of the day. A mesoscale

241   pressure gradient arising from E-W differential heating was superimposed upon a more steady-state synoptic-scale

242   pressure gradient.

243   *Easterly Gap Flow on 7 April*. An offshore surface low, moving northward during the day, provided the pressure

244   gradient for a weak easterly gap flow in the lowest 200-400 m, increasing with height.

245   *Mountain wave on 19 September*. A low-pressure system over the western half of the continental United States led

246   to northwesterly flow in the Columbia River Gorge, which caused stationary mountain waves east of the Cascade

247   Mountains. Doppler lidar cross-sections (Fig. 4d) reveal evidence for trapped lee waves through temporal fluctuations

248   of mean wind speed, but more accurate estimates of wave activity were made from high temporal-resolution lidar

249   measurements of vertical velocity as shown in Draxl et al. (2020). These measurements were obtained by the newly

250   developed micro-Doppler lidar (Schroeder et al. 2019) tested during a September-October 2016 deployment to the

251   Wasco site.

252 Differences in the diurnal flow structure and evolution of each event emerge among sites (Fig. 4). Further
253 averaging of wind profiles over all instances of each flow-event type at each site (Fig. 5a-c), and for the 3-site
254 composites (Fig. 5d), leads to smoothed but representative mean features of these flows. The Easterly flow composites
255 represent all days with easterly wind directions as detected by all three lidars, regardless of how they formed (i.e., they
256 could be associated with cold pools, easterly gap flows, or other forcings). Note that a mountain-wave phenomenon is
257 generally unique and may occur at different times of the day, so that diurnal composite time-height cross-sections are
258 unlikely to be representative of the structure or evolution of individual events.

259 Comparing wind-speed structure and behavior among the three lidar sites (Fig. 1b) reveals wind-flow variations
260 due to distance, elevation, and the presence of wind turbines. During prevalent westerly winds, the westernmost lidar
261 at Wasco documented an inflow profile upstream of the wind farms, whereas the lidar at the Arlington site shows how
262 the flow changed due to location and may give insight as to how the wind farms affect the shape and magnitude of the
263 wind profile (Banta et al. 2015, Fitch et al. 2012, Lundquist et al. 2019, Rajewski et al. 2013, Redfern et al. 2019,
264 Siedersleben et al. 2018, 2020). When the flow was from easterly directions, both lidars measured winds affected by
265 both terrain complexity and wind farms. The flow at the easternmost Boardman site, where winds were weakest of the
266 three sites, was least influenced by wind farms. Considering the above factors is important for selecting a proper site
267 for any measurement system.

268 Meteorological events often persisted for several days, accompanied by fluctuations in wind speed and/or wind
269 direction. Strong increases or decreases of wind speed over a few hours, known as wind *ramp* events, lead to a
270 corresponding ramping in wind power production. Errors in timing or intensity of these ramp events may cause
271 significant discrepancies between model-predicted and actual power output (Bianco et al. 2019).

272 The scanning lidars were located along the W-SW wind direction frequently observed over the research area.
273 During episodes of westerly or easterly winds, the variations in wind power computed from lidar measurements can
274 approximate the fluctuations of total power generated over the BPA area (see Wilczak et al. 2019 for methodology).
275 Examples of 80-m wind-speed ramps and wind power computed from lidar measurements at 3 sites as well as the BPA-
276 reported power generation are shown for a westerly gap-flow event on 13-19 August and a cold pool event on 07-13
277 February (Fig. 6). During the August case, marine-intrusion flows brought strong nocturnal winds to the area and a
278 strong diurnal cycle to both wind speed and total generated power.

279 The variations of wind speed and wind power computed from lidar measurements are similar to those of the
280 power generated over the BPA area. During the August period (Fig. 6a-c), large up or down power ramps reached 3
281 GW or more (17 August). Time offsets among the lidar-measured ramps in wind speed and between these ramps and
282 those in the normalized BPA power occurred as the wind-surge event propagated over the 30-40 km distance from one
283 site to the next (Fig. 1). These time offsets produced discrepancies between individual-site wind measurements and
284 total power produced but averaging the winds over the three sites smooths the temporal offsets and produces better
285 agreement and higher temporal correlations. The cold pool during the 07-13 February case (Fig. 6d-f) brought weak
286 winds to the area leading to low BPA power generation; the power ramped up to 2.5 GW by the end of this period as
287 the cold pool dissipated and wind speeds increased.

288   Distributions of wind speed and wind direction (Fig.6, bottom row) during the August reforecast period show

289 stronger wind speeds with predominant westerly winds at each site, whereas the February case had weaker winds and

290 a larger percentage of easterly flow. A better correlation between power computed from lidar measurements and BPA

291 generation is found for the August case when the winds were consistently from the west. The correlation coefficient

292 (R^2) values for this period are 0.83, 0.86, 0.78, and 0.95 for Wasco, Arlington, Boardman, and the 3-site composites,

293 respectively, illustrating how the three-site averaging correlates better with the total power generated in a month

294 (August) when propagating diurnal wind-surge events occurred often. For the February case of weak wind speeds and

295 a broader range of wind directions, correlation values are 0.52, 0.78, and 0.69 for Wasco and Arlington, and for the 3-

296 site composites, respectively. No linear relationship was found for the weakest winds at the lowest Boardman site

297 situated further to the east of wind farms in the area. Results from Fig. 6a and b illustrate that measurements from

298 Doppler lidars properly located along the prevalent winds may in some conditions be adequate for the assessment of

299 wind resources in the larger area of surrounding wind farms.

300   **3. Results: Model validation by lidar measurements**

301   HRRR model performance was quantified using several validation metrics: bias (BIAS), root mean square error

302 (RMSE) and mean absolute error (MAE) The validation metrics obtained for annual and seasonal periods and for cases

303 of interesting meteorological events observed in the study area, are analyzed as diurnal time-height cross-sections,

304 mean profiles, and time-series of hub-height (80-m AGL) winds. These analyses allow models to be assessed regarding

305 model resolution (HRRRNEST *vs.* HRRR) and model-physics updates (EXPR *vs.* CNTR runs). All types of analysis

306 performed also provide information about model errors between sites of different complexity as well as differences in

307 model performance for seasons and major events.

308   **3.1 Time-height analysis**

309   **3.1.1 Annual model performance**

310   Time-height sections from lidar measurements and HRRR-model output for the 00z run (Fig.7a-c), averaged

311 over all reforecast periods in 2016 (which we will refer to as *annual*, even though not all days were used), demonstrate

312 a similarity in diurnal variability of HRRR *vs.* lidar-measured wind speed. The influence of the frequent westerly gap

313 flows (46% of all events counted: Fig. 3 and Table 1) on the annual mean is evident at each site and for the 3-site

314 composites. Winds were fastest at night, and these nocturnal winds were stronger (≥ 10 m s^{-1}) at the highest, western-

315 most Wasco site, whereas the weakest winds (~7-8 m s^{-1}) were observed at the lowest, eastern-most Boardman site.

316 The diurnal variability of winds at the Arlington site most closely resembled the diurnal variability of the 3-site

317 composites.

318   The biases between HRRR-CNTR and lidar (Fig. 7d) and HRRR-EXPR and lidar (Fig. 7e) varied mostly

319 between -2 and +2 m s^{-1} at the Wasco and Arlington sites with larger (up to +/- 3 m s^{-1}) values above 400 m AGL at

320 Boardman. Both models tended to overestimate wind speeds above 300 m around and just after the evening transitions

321 (0000-0500 UTC) and between 1200 and 1800 UTC at Wasco. In the rotor layer, the runs (except for the CNTR at

322 Wasco) tended to underestimate wind speeds, by as much as 3 m s^{-1} for the EXPR run at Arlington.

09 January 2024 21:36:56

323   At Boardman, the winds in the lidar cross-section above ~500 m were considerably weaker than at the other two

324   sites. Both model versions, however, maintained the relatively stronger winds aloft, resulting in large positive biases

325   reaching 3 m s⁻¹. The large positive biases in the 3-site composite cross-sections thus mostly reflect these high biases

326   aloft at Boardman. Raw LOS lidar data were not available for the lidar at Boardman to use for the quality-assurance

327   procedures, as described by Pichugina et al. (2019), so these 500- to 1000-m-AGL values may be from vertical regions

328   of low lidar signal-to-noise ratio, and consequently may not be as accurate as at the two other sites. Therefore, as a

329   precaution, wind profiles in the next sections will be shown only up to 500 m.

330   MAE from both runs of both models (Fig. 8) amounted to ~2 m s⁻¹ within the rotor layer, the largest values

331   exceeding 3 m s⁻¹ between 02 and 14 UTC in the EXPR version at Arlington and 2.5 m s⁻¹ at Boardman. MAE above

332   200 m increased with height, which again revealed the largest values occurring aloft at Boardman.

333   The models are compared against each other in Figs. 9-10. The new physics and other updates are evaluated in

334   Fig. 9, which shows the MAE differences between EXPR and CNTR runs of the HRRR and HRRRNEST models,

335   respectively. Negative values denote MAE reduction (improvement) due to model updates in EXPR. The EXPR version

336   generally showed improvements of 0.5-1.0 m s⁻¹ (5-10%) aloft between 200 and 600 m, in agreement with the 'model-

337   difference' cross-sections in Olson et al. (2019) based on profile data averaged over the eight 915-MHz Radar Wind

338   Profilers (WPRs), as described above in Section 2.3. Below 150 m AGL, however, during local nighttime hours of

339   0300-1200 UTC, the lidar-based cross-sections show that EXPR updates produced decreases in the model skill of ~1

340   m s⁻¹ (up to 10%) at Arlington and Boardman compared to the CNTR, a rotor-layer degradation that was not well

341   characterized in the RWP cross-sections (Olson et al. 2019). Profiles presented later in Section 3.2.b show that these

342   low-level decreases in skill occurred during spring and summer. This result is consistent with summertime 80-m-level

343   reductions in skill during the late afternoon and evening hours seen in Olson et al. (2019), Bianco et al. (2019).

344   The magnitudes of MAE differences between EXPR and CNTR runs of both HRRR and HRRRNEST were

345   largest within the rotor layer for nighttime periods (03-13 UTC). Differences between EXPR and CNTR (Fig. 9a, c)

346   were smaller for Wasco than the other two sites and the 3-site composites, indicating that the model-change impacts

347   were site-dependent. The finer-resolution model, HRRRNEST (Fig. 9b, d) results generally show similar but smaller-

348   magnitude differences due to the model-physics updates.

349   Model grid-spacing effects are shown in Fig.10. MAE differences between CNTR runs of both models (row

350   10a) generally indicate that increasing resolution to a 750-m mesh produced improvements of ~0.5 m s⁻¹, except at the

351   surface at night. The EXPR runs (row 10b) mostly show improvements due to the finer resolution. However, the EXPR

352   model at Wasco showed slight (<0.4 m s⁻¹) tendencies to degrade skill for much of the time, and likewise for Arlington

353   at the surface at night (10-16 UTC).

354   The last row (Fig. 10c) shows the combined effect of improved physics and grid resolution of HRRRNEST over

355   HRRR. Overall improvement (~0.5 m s⁻¹) occurred for the nighttime/morning hours (03-18 UTC) in the 100-400-m

356   layer, similar to the pattern for the CNTR resolution impacts but larger. The combined effect of model physics and

357   finer grid resolution below 150 m at Arlington and below 50 m at Boardman during nighttime hours was to decrease

358   skill by >0.5 m s⁻¹, which was also reflected in the 3-site composites.

359   Overall, even in the annual average, wind speeds showed a strong diurnal modulation, the strongest winds

360   occurring at night (Fig.8a, Pichugina et al. 2019), and the model errors were also diurnal and largest at night. The

361   model-difference time-height patterns often displayed a complex and patchy nature, even for the sign of the

362   improvements. The effects of model updates and finer resolution tended to improve model skill, but the skill was seen

363   to decrease in certain instances: at levels near the surface at Arlington and Boardman, and comparisons showed

364   decreased skill most of the time for the finer resolution runs of the HRRR-EXPR through the lowest kilometer at Wasco.

365   The improvements in model skill were larger as a result of increasing the grid resolution than by implementing the set

366   of model-physics updates in EXPR. The largest improvements were for the combined updated-physics, EXPR/finer-

367   resolution run vs. the CNTR/coarser-resolution run (Fig. 10c).

368   **3.1.2 Model performance during observed events.**

369   Similar time-height cross-section analyses were performed for composites over particular events, including

370   westerly-gap-flow (Fig. 11) and cold-pool (Fig. 12) days, two major event types that often lasted for several days and

371   influenced wind-power production. These two event types have a strong influence on the annual wind and model-error

372   statistics (Pichugina et al. 2019, Banta et al. 2020).

373   The composites of winds during frequent westerly gap flow events (Fig. 11a-c) comprised 46% out of the total

374   number of events in the 2016 reforecast periods (Fig. 8a-c). Similar to annual averages, both runs of the HRRR captured

375   the mean diurnal trends of wind speed at each site, but show temporal and vertical discrepancies, which are smoothed

376   for the 3-site composite. Most notably, both HRRR versions produced the same early onset reduction in wind speeds

377   after 0500 UTC at Arlington and Boardman that was noted by Pichugina et al. 2019, Wilczak et al. (2019), and Banta

378   et al. (2020) for the HRRR-NCEP model, traced to poor representation of summertime, thermally forced flows related

379   to the penetration of a regional sea-breeze into this area. Biases (not shown) from CNTR and EXPR 0z runs of HRRR

380   ranged from -2.5 to 2.5 m s^{-1}, with a slightly larger value (~3 m s^{-1}) above 400 m at Boardman. The MAE differences

381   in row d of Fig. 11 show that the EXPR run reduced the magnitude of the wind-speed discrepancies (negative values,

382   green colors) above 200 m AGL at Arlington and Boardman somewhat, but had a larger MAE (>2 m s^{-1}) than the

383   CNTR within the rotor layer during nighttime hours at Arlington and (to a lesser degree) at Boardman, indicating that

384   the updates to the EXPR degraded model skill below 150 m—a pattern that was noted in the *annual* cross-sections of

385   the previous section. The magnitudes are larger here, though, indicating that the model improvements (+ or -) averaged

386   over all periods largely reflect how the model handled westerly gap flows. The similarity of these patterns to the annual

387   ones reinforces the dominance of westerly gap flows in HRRR error statistics as well as the basic wind-speed statistics

388   (Pichugina et al. 2019).

389   During cold-pool events (Fig. 11a-c), both runs of the model also captured many aspects of wind-speed behavior

390   at each site and for the composite, but the largest errors were due to the simulated weak-wind layer being too shallow

391   (Pichugina et al. 2019, also noted by McCaffrey et al. 2019, Olson et al. 2019, and Bianco et al. 2019 using different

392   data sources) and also to the mistiming of episodes of deepening or shallowing of this layer, which appeared even in

393   the composites. These effects led to ±3 m s^{-1} biases (not shown) with larger values at Boardman. MAE differences

394   between EXPR and CNTR runs of the model (Fig. 11d) varied between ±1 m s^{-1} (negative values again indicating

395 improvements for EXPR). Similar analyses of HRRRNEST (not shown) show smaller MAE for each run and smaller

396 MAE differences between runs.

397 Time-height cross-section analysis, using a high vertical and temporal resolution of lidar measurements,

398 provides simultaneous information on both temporal and vertical variability of winds and validation errors, but more

399 quantitative results can be seen in mean profile and time-series analysis.

3.2 Mean-profile analysis

3.2.1 Annual

402 Annually averaged profiles of measured and modeled wind speed from both runs of HRRR and HRRRNEST along

403 with the profiles of annual-mean errors are shown in Fig. 13 for each site and the 3-site composites. Both runs of the

404 HRRR and HRRRNEST under predicted wind speeds below 400 m (column 11a) (except for the HRRR-CNTR run at

405 Wasco), causing larger values of bias (column 13b) of up to ~2 m s^{-1} at Arlington and MAE (column 13c) of 2 m s^{-1} or

406 more at all sites. Model-skill impacts due to model-physics updates (column 13d) at Wasco show improvements except

407 very near the surface, whereas for the other two sites the improvements were seen only above 150 m, as noted in the

408 time-height cross-sections (Fig. 8). Below that level at these sites, the updates in EXPR generated larger MAE in the

409 rotor layer of 0.3 m s^{-1} annually, as also seen in Fig. 8. Finer model resolution (column 13e) mostly resulted in reduced

410 errors for the 750-m HRRRNEST compared to the 3-km HRRR model. Exceptions were for the HRRR-EXPR above

411 100 m at Wasco and the combined physics/resolution comparisons below 100 m at Arlington. The improvements were

412 largest (0.2-0.3 m s^{-1}) for EXPR runs below 200 m at the Arlington and Boardman sites, and were also significant for

413 the combined impacts.

3.2.2 Profiles averaged over seasons and major meteorological events

415 *Seasonal* (Fig.14a, c). Similar analyses of seasonally averaged wind speed errors (not shown here) for each site as well

416 as for the 3-site composites indicate again that the largest errors for both models occurred below 200 m, except for the

417 winter season.

418 Seasonal differences due to the EXPR updates (Fig. 14a) show that the improvement above 150 m seen in the

419 annual profiles (Figs. 8, 13) was also found for each season. In winter, improvement below 150 m resulted largely from

420 the improvements in the EXPR runs in better representing the cold-pool depth at Wasco, thus correcting the high bias

421 to the wind speeds there. The large annual degradations seen below 150 m in the 3-site annual average (occurring most

422 strongly at Arlington and Boardman) are shown here to be due to problems in summer and spring.

423 Going to the finer mesh (Fig. 14c) generally reduced errors above ~30 m AGL, except for the EXPR runs during

424 fall. The largest reductions are seen in the EXPR runs below 150 m in summer and spring, when the finer mesh produced

425 better simulations of the thermally forced, westerly gap flows.

426 *Major meteorological events* (Fig.14b, d). Results obtained for cold-pool days, frequent during winter months, are

427 similar to those obtained for the winter season but reduced in magnitude, whereas results for westerly gap flows most

428 resemble those for spring and summer. Easterly flows, which qualitatively exhibit behavior similar to the wintertime

429 results, observed over different seasons, show ~0.4 m s^{-1} improvement due to the combined effect of model physics

430 and model resolution.

12

3.3 Hub-height (80-m) wind

3.3.1 Time series analysis

Knowledge of the error in hub-height and rotor-disk wind-speed forecasts 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 hub-height of most wind turbines in the surrounding WFIP2 area. Figure 15 shows annually-averaged, diurnal time series of wind speed and model validation metrics for the HRRR CNTR and EXPR runs initialized at 00z, and the changes in model error due to updates in EXPR. Results were similar for the runs initialized at 12z (not shown here).

The bias plots (Fig. 15a, b, second row) show the underestimates of wind speed in the rotor layer noted previously for Fig. 7. Again the exception is the CNTR run at Wasco, where the speeds were overestimated by more than 1 m s^{-1} during nighttime hours—due to issues with the depth of the modeled wintertime cold pools, as discussed. Except for Wasco, the EXPR had larger negative biases and larger MAE than the CNTR (fourth row), including for the 3-site composites, consistent with the wind-speed cross-sections (Fig. 8) and profiles (Fig.13).

Grid-resolution effects can be seen by comparing the wind-speed and error statistics for (a) HRRR and (b) HRRNEST. The improved performance of the finer-meshed models is clear from the smaller magnitudes of the error values. The high biases at Wasco in the CNTR runs and low biases in EXPR at Arlington are still present, but much reduced. The smaller-magnitude errors led to smaller-magnitude differences due to EXPR updates (bottom rows).

3.3.2 Summary plots

Annual model-comparison results are summarized in Figure 16 where the percent change in forecast MAE (Δ-MAE) for both models is further broken down into nighttime and daytime hours.

Updates in model physics. Fig. 16a and b show Δ-MAE between CNTR and EXPR runs of the HRRR and HRRRNEST models due to the *model physics* updates in the EXPR runs. The plot shows how the differences were larger for the HRRR than the HRRRNEST and larger at night. Improvements are seen at Wasco, especially at night, but HRRR-EXPR performance was mostly worse than the CNTR's at the other sites and in the 3-site composite. The magnitudes of the Δ-MAEs were smaller for the nested model and these changes were of opposite sign at Wasco and Arlington, which canceled, resulting in small Δ-MAE in the composite.

Model horizontal grid resolution. Changes in model MAE due to finer-resolution grids are depicted in Fig. 14c, and d for each model version. All categories (except for Wasco at night) showed reduced error, with the relative decreases appearing largest for the EXPR version.

Overall Δ-MAE. Changes in model MAE due to updates in model physics and horizontal grid resolution are illustrated in Fig. 16e, f. The negative Δ-MAE values for both initial times imply 2-10% improvements at all three sites and 3 site composites except for nighttime hours at the Arlington.

Seasons and major events. The nature of the model errors also changed significantly from season to season and for different event types. Figure 17 shows the combined physics/resolution impacts of the updates implemented into HRRRNEST-EXPR runs as compared against HRRR-CNTR, shown for 3-site composites averaged over each season (Fig. 17a) and each major event (Fig. 17b). Improvement is indicated for all seasons, MAE reductions reaching ~10%

13

09 January 2024 21:36:56

during daytime in summer and spring, and during nighttime in winter. All events showed improvement for all hours, the largest appearing for easterly flows, especially at night. Small degradations are indicated for daytime cold pools and nighttime gap flows.

Power is generated by wind turbines if wind speeds are stronger the cut-in threshold of 3-4 m s^{-1} for typical power curves. Distributions of lidar-measured and modeled 80-m winds (see Appendix C, Fig. 21) and Table 6 shows that about 40% of wind speeds were less than 4 m s^{-1}. Although the analysis of the error statistics of wind power generation beyond the scope of this paper, we illustrate here an example of the power MAE change due to updates in the model physics, model horizontal grid resolution and overall changes that include all the above updates (Fig. 18). The power at each site was calculated using a typical power curve (Dupont et al. 2017). A trend of nighttime and daytime power MAE (Fig 18a-e) is similar to the 80-m wind speed MAE (Fig. 16a-e) but magnitudes are larger than those for wind speeds as the power is proportional to the cubic of the wind speed in regions of the power curve that are sensitive to wind speed (International Electrotechnical Commission, 2007). The overall impact from the updates in the model physics and the horizontal grid resolution on the speed and power Δ-MAE is illustrated in Table 2. The larger MAE improvement is observed at the Wasco sites where winds are much stronger compared to the Boardman site (see Appendix C, Fig. 21). Thus the power MAE improvement, and even more the power forecast error, depended on the location of the measurement in this complex-terrain setting.

4. Discussion

The overall findings here indicate that averaged over the four reforecast periods and the three lidar sites, the model updates to HRRR-EXP improved model skill above 200 m AGL, but degraded skill in the turbine rotor layer. Profiles showed that the decreases in skill were significant below 150 m at Arlington and Boardman. In the 3-site composites, time-height cross-sections and time series indicated that the degradation was strongest at night. These results contrast to those of Bianco et al. (2019), which found uniform improvement in area-averaged 80-m wind data, except during evening-transition hours in spring and summer. Measurements from the 19 sodars and three lidars deployed over the larger WFIP2 study area were averaged together in Bianco et al. (2019) and compared with similarly averaged output from the various versions of HRRR, as previously described. Calculated error values plotted as time series and bar graphs mostly showed improvement in HRRR skill as a result of EXPR updates.

The three Doppler-lidar sites in the present study were located within the wind-energy corridor of the WFIP2 study area along a line of the prevailing wind direction in the Columbia River Valley (Fig. 1). Previous analyses (Pichugina et al. 2019, Banta et al. 2020) have shown that model error characteristics and magnitudes can vary significantly among sites separated by a few tens of kilometers or more in the complex terrain of the WFIP2 region. In an attempt to reconcile the apparent discrepancies, the error data used in the Bianco et al. (2019) study were plotted for each measurement site to determine whether horizontal variations in error could be a factor in explaining the different results between the two studies.

Figure 19 shows site-by-site plots indicating which locations saw improvements due to the updates in the HRRRNEST EXPR and which locations saw decreases in skill compared to the HRRRNEST CNTR. For this plot, bias difference was broken into three categories: (blue) Improvement (EXPR |bias| < CNTR |bias|), (red): degradation (EXPR |bias| > CNTR |bias|), and (white) both values were about equal (EXPR |bias| ≈ CNTR |bias|). It is immediately

09 January 2024 21:36:56

504   obvious even in the sodar data that the sites that saw degradation during winter were low-elevation sites along the river.

505   In summer, and when averaged over all four reforecast periods, the decreased-skill sites were within the wind-energy

506   corridor along the river, and some of the locations where skill was improved were outside this area (i. e. sites

507   abbreviated as YKM, PVE). The Doppler lidars of this study were located along the strip where decreases in skill were

508   found. Thus, spatial differences in error characteristics do appear to be a factor in the differences in results between the

509   two studies. Evaluating model improvement by averaging over a wide area makes sense, for example, when measuring

510   the impact of improvements to a model on power production over a wind power generation area, as the total power

511   generated by all the turbines in that area generally matters most to the grid operator, and error characteristics at a single

512   site can be unrepresentative or misleading. However, when attempting to diagnose the sources of model errors, it is

513   more useful to do so at locations and over areas where the errors are occurring, rather than over large regions where

514   this error information has been averaged or smoothed out.

515   Forecast degradation at the sodar sites located along the river in Fig. 19 is thus in agreement with the forecast

516   degradations found by scanning lidars at the ARL, WCO, and BOR sites.

517

518   **5. Conclusions**

519   Doppler lidar measurements at three sites located along the prevalent wind directions in the Columbia River

520   Basin were used to evaluate the spatial and temporal variability of errors for two configurations of the model: the

521   'parent' HRRR (3-km horizontal grid spacing) and an experimental, finer-resolution, nested version HRRRNEST (750-

522   m horizontal grid spacing). Both models were run in control mode (CNTR)- the version available at the beginning of

523   WFIP2 in September 2015- and experimental mode (EXPR), with the model updates implemented by the end of WFIP2

524   (March 2017).

525   The evaluation of models was performed using three analysis types: time-height cross-sections of mean wind

526   speed and model-error statistics, vertical profiles of those quantities, and time series with accompanying summary bar

527   graphs of 80-m wind speed and error statistics. The results show how the updates in the EXPR versions affected model

528   skill season by season, site by site, and over periods of atmospheric phenomena relevant to wind energy.

529   The greatest improvements in the HRRR model skill came from reducing the grid interval from 3 km to 750

530   m. Combining the finer resolution with the EXPR updates produced the best overall improvement statistics, but only

531   somewhat better than the finer-resolution impacts. These conclusions are based on evaluations of both models using

532   Doppler-lidar measurements at the three sites located within the WFIP2 study area. The better performance of the

533   HRRRNEST runs compared to HRRR was found for all sites, seasons, and major atmospheric events.

534   Specifically, we find that:

535   ❑ The largest values of biases, MAE, RMSE and bias-corrected RMSE were found below 150 m for both runs of

536   HRRR and HRRRNEST.

537   ❑ Profiles of validation metrics computed for all reforecast periods (annual) below 200 m show bias error of 1.2 m

538   s^{-1} and 0.6 m s^{-1} for HRRR and HRRRNEST, respectively. MAE differences were small for both HRRRNEST

539   runs and varied between –0.2 and +0.2 m s^{-1} for HRRR runs (Fig. 17). For the 3-km HRRR, updates in the EXPR

540   version produced improvements of ~0.2 m s^{-1} during winter but produced decreases in skill (larger errors) in

541   summer and spring. The 750-m HRRRNEST versions were less affected by these updates, as the difference

542   between EXPR and CNTR was generally less than 0.05 m s^{-1} for the nested models.

543   ❑ Relative differences in annual 80-m wind speed MAE (Fig. 17) due to the experimental physics (EXPR *vs*. CNTR),

544   model horizontal grid resolution (HRRRNEST *vs*. HRRR) as well as the overall improvement (HRRRNEST EXPR

545   *vs*. HRRR CNTR) were largest for nighttime hours and for the westerly wind directions at each site and the 3-site

546   composites.

547   ❑ The biases between measured and modeled winds greatly depended on the location of the instrument (terrain

548   complexity), the season, and the mean wind speed as illustrated for CNTR and EXPR runs of HRRRNEST (Fig.

549   18).

550   ❑ Reducing the grid interval from 3 km to 750 m produced the greatest improvements in HRRR model skill, for the

551   model versions tested.

552

570   **Appendix A.** Description of the WFIP2 models evaluated in the paper

571   **Appendix B**. Correlation between model outputs of 80 m wind speed extracted at the lidar sites using bi-linear

572   interpolation or nearest grid point value (Fig. 20).

573   **Appendix C**. The annual distributions of 80-m wind speeds and wind directions at each site (Fig. 21).

574

575   **References**

09 January 2024 21:36:56

Ascione, A., Cinque, A., Miccadei, E., Villani, F. and Berti, C. 2008: The Plio-Quaternary uplift of the Apennine chain: New data from the analysis of topography and river valleys in Central Italy, *Geomorphology*, **102**(1), 105–118, doi:10.1016/j.geomorph.2007.07.022

Banta, R.M., Y.L. Pichugina, N.D. Kelley, W.A. Brewer, and R.M. Hardesty, 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. DOI:10.1175/BAMS-D-11-00057.1

Banta, R.M., Y.L. Pichugina, W.A. Brewer, E.P. James, J.B. Olson, S.G. Benjamin, J.R. Carley, L. Bianco, I.V. Djalalova, J.M. Wilczak, M.C. Marquis, J. Cline, and R.M. Hardesty, 2018: Evaluating and improving NWP forecasts for the future: How the needs of offshore wind energy can point the way. *Bull. Amer. Meteor. Soc.*, **99**, 1155-1176, doi: 10.1175/BAMS-D-16-0310.1 Banta, R.M., Y.L. Pichugina, W.A. Brewer, A. Choukulkar, K.O. Lantz, J.B. Olson, J. Kenyon, H.J.S. Fernando, R. Krishnamurthy, M.J. Stoelinga, J. Sharp, L.S. Darby, D.D. Turner, S. Baidar, and S.P. Sandberg, 2019: 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; doi.org/10.1175/MWR-D-19-0188.1

Banta, R.M., Pichugina Y.L., Brewer W.A., Lundquist J.K., Kelley N.D., Sandberg S.P., Alvarez R.J., Hardesty R.M., Weickmann A.M., 2015: 3-D volumetric analysis of wind-turbine wake properties in the atmosphere using high-resolution Doppler lidar. *J. Atmos. Oceanic Technol.,* **32**: 904–914. doi.org/10.1175/JTECH-D-14-00078.1

Benjamin, S.G., S.S. Weygandt, J.M. Brown, M. Hu, C. Alexander, T.G. Smirnova, J.B. Olson, E. James, D.C. Dowell, G.A. Grell, et al., 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669-1694; doi: http://dx.doi.org/10.1175/MWR-D-15-0242.1.

Bianco, L., I.V. Djalalova, J.M. Wilczak, J.B. Olson, et al., 2019: Impact of model improvements on 80 m wind speeds during the second Wind Forecast Improvement Project (WFIP2). *Geosci. Model. Dev.*, **12**, 4803-4821; doi.org: 10.5194/gmd-12-4803-2019

Bonin, T. A., Choukulkar, A., Brewer, W. A., Sandberg, S. P., Weickmann, A. M., Pichugina, Y. L., Banta, R. M., Oncley, S. P., and Wolfe, D. E, 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., 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. doi.org/10.1002/we.1555.

Draxl, Berg, Bianco, Bonin, Choukulkar, Clifton, Cline, Djalalova, Ghate, Grimit, Holub, Kenyon, Lantz, Long, Lundquist, McCaa, McCaffrey, Newman, Olson, Pichugina, Sharp, Shaw, Smith, and Toy, 2019: The Verification and Validation Strategy Within the Second Wind Forecast Improvement Project (WFIP 2). *National Renewable Energy Laboratory Report* www.nrel.gov/publications.

Draxl, C., R. Worsnop, G. Xia, D. Chand, J. K. Lundquist, Y. Pichugina, J. Sharp, G. Wedam, J. Wilczak, and L. Berg: 2020, Mountain waves impact wind power generation, submitted to Wind Energy Science.

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, doi.org/10.1175/JAMC-D-11-0247.1

09 January 2024 21:36:56

613 Dupont, Elise & Koppelaar, Rembrandt & Jeanmart, Hervé. (2017). Global available wind energy with physical and
614   energy return on investment constraints. Applied Energy. 10.1016/j.apenergy.2017.09.085.

615 International Electrotechnical Commission, 2007: Wind turbines - Part 12-1: Power performance measurements of
616   electricity producing wind turbines. IEC 61400-12-1, 90 pp.

617 Fast J.D., and L.S. Darby. 2004. "An Evaluation of Mesoscale Model Predictions of Down-Valley and Canyon Flows
618   and Their Consequences Using Doppler Lidar Measurements During VTMX 2000." *Journal of Applied*
619   *Meteorology* 43, no. 3:420-436.

620 Fitch, A. C., J. B. Olson, J. K. Lundquist, J. Dudhia, A. K. Gupta, J. Michalakes, and I. Barstad, 2012: Local and
621   Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model. *Mon. Weather Rev.*, **140**,
622   3017–3038, https://doi.org/10.1175/MWR-D-11-00352.1.

623 Lundquist, J.K., DuVivier, K.K., Kaffine, D. et al. Costs and consequences of wind turbine wake effects arising from
624   uncoordinated wind energy development. *Nat Energy* **4**, 26–34 (2019). https://doi.org/10.1038/s41560-018-
625   0281-2

626 Krogsæter, O., and J. Reuder, 2015: Validation of boundary layer parameterization schemes in the weather research
627   and forecasting model under the aspect of offshore wind energy applications—Part I: Average wind speed and
628   wind shear. *Wind Energy*, **18**, 769-782.

629 McCaffrey, K., J.M. Wilczak, L. Bianco, E. Grimit, J. Sharp, R. Banta, K. Friedrich, H.J.S. Fernando, R.
630   Krishnamurthy, L. Leo, and P. Muradyan, 2019: Identification and characterization of persistent cold pool
631   events from temperature and wind profilers in the Columbia River Basin. *J. Appl. Meteor. Climatol.*, **58**, 2533–
632   2551; doi.org/10.1175/JAMC-D-19-0046.1

633 Olson, J.B., J.S. Kenyon, I. Djalalova, L. Bianco, D.D. Turner, Y. Pichugina, A. Choukulkar, M.D. Toy, J.M. Brown,
634   W.M. Angevine, E. Akish, J.-W. Bao, P. Jimenez, B. Kosovic, K.A. Lundquist, C. Draxl, J.K. Lundquist, J.
635   McCaa, K. McCaffrey, K. Lantz, C. Long, J. Wilczak, R. Banta, M. Marquis, S. Redfern, L.K. Berg, W. Shaw,
636   and J. Cline, 2019: Improving wind energy forecasting through numerical weather prediction model
637   development. *Bull. Amer. Meteor. Soc.*, **100**, 2201-2220; doi.org/10.1175/BAMS-D-18-0040.1

638 Pichugina, Y. L., R.M. Banta, J.B. Olson, J.R. Carley, M.C. Marquis, W.A. Brewer, J.M. Wilczak, I. V. Djalalova, L.
639   Bianco, E.P. James, S.G. Benjamin, and J. Cline, 2017: Assessment of NWP forecast models in simulating
640   offshore winds through the lower boundary layer by measurements from a ship-based scanning Doppler lidar.
641   Monthly *Weather Review*, vol. **145**, no. 10, 4277-4301 doi: 10.1175/MWR-D-16-0442.1

642 Pichugina, Y.L., R.M. Banta, and Coauthors, 2019: Spatial variability of winds and HRRR-NCEP model error
643   statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia River Basin. *J.*
644   *Appl. Meteor. Climatol.*, **58**, 1633-1656. doi.org/10.1175/JAMC-D-11-040.1

645 Rajewski, D.A., E.S. Takle, J.K. Lundquist, S. Oncley, J.H. Prueger, T.W. Horst, M.E. Rhodes, R. Pfeiffer, J.L.
646   Hatfield, K.K. Spoth, and R.K. Doorenbos, 2013: Crop Wind Energy Experiment (CWEX): Observations of
647   Surface-Layer, Boundary Layer, and Mesoscale Interactions with a Wind Farm. *Bull. Amer. Meteor. Soc.*,
648   **94**, 655–672, https://doi.org/10.1175/BAMS-D-11-00240.

18

09 January 2024 21:36:56

Redfern, S., J. B. Olson, J. K. Lundquist, C. T. M. Clack, 2019, Incorporation of the Rotor-Equivalent Wind Speed into the Weather Research and Forecasting model's Wind Farm Parameterization. *Monthly Weather Review* **147**, 1029-1046.

Schroeder, Paul, W. Alan Brewer, Aditya Choukulkar, Ann Weickmann, Michael Zucker, Maxwell Holloway, Scott Sandberg. 2019: A compact, flexible, and robust micro pulsed Doppler Lidar. *Journal of Atmospheric and Oceanic Technology*, in peer review.

Sharp, J., & Mass, C. (2002). Columbia Gorge gap flow: Insights from observational analysis and ultra-high-resolution simulation. *Bulletin of the American Meteorological Society*, *83*(12), 1757-1762.

Sharp, J. 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., L.K Berg, J. Cline, C. Draxl, E. Grimit, J.K. Lundquist, M. Marquis, J. McCaa, J. Olson, C. Sivaraman, J. Sharp, J. Wilczak, 2019: The Second Wind Forecast Improvement Project (WFIP 2): General overview. *Bull. Amer. Meteor. Soc.*, 100, 1687-1699; 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.Siedersleben, S. K., J. K. Lundquist, A. Platis, K. Bärfuss, B. Cañadillas, B. Djath, J. Schulz-Stellenfleth, T. Neumann, J. Bange, S. Emeis, 2018, Micrometeorological Impacts of Offshore Wind Farms as seen in Observations and Simulations, *Environmental Research Letters* **13**, 124012.

Siedersleben, S. K., Platis, A., Lundquist, J. K., Djath, B., Lampert, A., Bärfuss, K., Cañadillas, B., Schulz-Stellenfleth, J., Bange, J., Neumann, T., and Emeis, S.: Turbulent kinetic energy over large offshore wind farms observed and simulated by the mesoscale model WRF (3.8.1), *Geosci. Model Dev*., **13**, 249–268, https://doi.org/10.5194/gmd-13-249-2020, 2020

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, doi:10.1175/1520-0434(2001)016,0432:CPITCB.2.0.CO;2.

Wilczak, J.M., M. Stoelinga, L.K. Berg, J. Sharp, C. Draxl, K McCaffrey, R.M. Banta, L. Bianco, I. Djalalova, J.K. Lundquist, P. Muradyan, A. Choukulkar, L. Leo, T. Bonin, Y. Pichugina, R. Eckman, C.N. Long, K. Lantz, R.P. Worsnop, J. Bickford, N. Bodini, D. Chand, A. Clifton, J. Cline, D.R. Cook, H.J.S. Fernando, K. Friedrich, R. Krishnamurthy, M. Marquis, J. McCaa, J.B. Olson, S. Otarola-Bustos, G. Scott, W.J. Shaw, S. Wharton, A.B. White, 2019: The Second Wind Forecast Improvement Project (WFIP2): Observational field campaign. *Bull. Amer. Meteor. Soc*., **100**, 1701-1723; doi.org/10.1175/BAMS-D-18-0035.1

Yang, B., Y. Qian, L. K. Berg, P. L. Ma, S. Wharton, V. Bulaevskaya, H. Yan, Z. Hou, and W. J. Shaw, 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. doi: 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.

19

685   Zhong, S., C. D. Whiteman, X. Bian, W. J. Shaw, and J. M. Hubbe, 2001: Meteorological processes affecting the
686   evolution of a wintertime cold-air pool in the Columbia Basin. *Mon. Wea. Rev.,* **129,** 2600–2613.

687   **Data Availability Statement**

AVAILABILITY OF DATA	DATA AVAILABILITY STATEMENT
Data generated at a central, large scale Facility: NOAA Meteorological Assimilation Data Ingest System (MADIS)	The information on instrument metadata such as location, dates of deployment, data-processing methods including time averaging, whether the data was 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
Data openly available in a public repository that does not issue DOIs	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 in available on DAP and on 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 through 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/
Data available on request due to privacy/ethical restrictions:	Power generation data. 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

688   **Table Captions**

689   Table 1. The number of days (#) having observed event types during reforecast periods in 2016, 12 months of concurrent

690   Doppler lidar measurements in 2016, and 18 months of the WFIP2 experiment. Occurrence (%) of a particular event

691   type during each period relative to the total occurrence of events.

692   Table. 2. The overall impact of the updates in model physics and horizontal grid resolution on wind speed and power

693   Δ-MAE.

694   Table 3. Observed 80-m mean wind speed and Bias of HRRR and HRRRNEST CNTR and EXPR runs during the

695   WFIP2 reforecast period (February 2016-March 2017) at 23 sites. The first three columns show site ID, site elevation

696   above sea level (ASL, m), and terrain complexity expressed as standard deviation (STD) of the elevation within 3 km

697   circle around each site.

698   Table 4. Statistics of the correlation between model outputs of 80 m wind speed extracted at the lidar sites using bi-

699   linear (Bi) interpolation or nearest grid point (N) value.

700   Table 5. Mean values for distributions of 80 m wind speed and direction from Fig. 21

701   Table 6. Percent of wind speeds in each region of a power curve of a typical wind turbine.

702   Table 7. Summary of the HRRR and HRRRNEST reforecast configurations.

703   Table 8. Description of the control and experimental physics configurations used in the reforecasts (adapted from

704   Table 2 in Olson et al. 2019), where model developments in support of WFIP2 are shown in bold. All components

705   not under development in WFIP2 (regular font) are described in Benjamin et al. (2016).

706   **Figure Captions**

707   Figure 1. (a) Map of the Columbia River Valley with the location of scanning Doppler lidars denoted by gold-filled

708   circles. Red circles indicate the highest peaks of the Cascade Mountains. (b) Google Earth detailed map of the study

709   area shows the location of two NOAA Doppler scanning lidars (200S) at Wasco (452 m ASL) and Arlington (262 m

710   ASL). The Notre Dame University Halo Streamline XR lidar is located at the Boardman site at 110 m ASL. The

711   white line indicates the east-west transect of the study region along the prevalent wind directions observed from

712   surface measurements and models during previous studies in this area and confirmed from the analysis of Doppler

713   lidar measurements in Pichugina et al. (2019)**.** The surrounding wind turbines (total rated capacity of ~3,800 MW)

714   are indicated by the clusters of dark yellow circles. An elevation gradient along the transect line is shown in (c).

715   Figure 2. (a) Domains of WFIP2 (green) HRRR and HRRRNEST models are shown along with the operational (white)

716   RAP (13-km grid spacing) and (red) HRRR-NCEP (3-km grid spacing) models. (Fig. 2a adapted from Olson et al.

717   2019). (b) An example of HRRRNEST-modeled wind flow at 80 m superimposed on area topography from WRF V3.7.

718   The example is taken from a 7-h forecast initialized at 000UTC on 15 May 2016. Locations of Doppler lidars are shown

719   by black circles. (c) Maps of the terrain elevations (in meters) at Wasco, Arlington, and Boardman research sites. The

720   location of each lidar (dark yellow triangles) is shown relative to the HRRR 3-km model corners of the grid cells (red

721   circles), and relative to the HRRRNEST 750-m model corners of the grid cells (white circles).

722   Figure 3. Atmospheric events observed over the WFIP2 research area from the analysis of models, satellites, and all

723   available ground observations. (a) The number of events observed during 18 months of the experiment, 12 months of

724   2016, and reforecast periods in 2016 shown by colors according to the legend. Multiple events can occur on the same

725   day. (b) Relative number (%) of major events in each month observed during the reforecast periods.

726   Figure 4. Time-height cross-sections of wind speed and wind direction at the (from top to bottom) Wasco, Arlington,

727   and Boardman site during atmospheric phenomena observed in the WFIP2 area: (a) Cold pool, (b) Westerly Gap Flow,

728   (c) Easterly Gap Flow, and (d) Mountain Waves. Wind speed is color-coded from 0 to 15 m s^{-1} according to the color

729   scale at the top of each panel. Black arrows show wind direction. Two white horizontal lines on each panel indicate

730   typical rotor-disk height, 50-150 m AGL.

731   Figure 5. Time-height cross-sections of wind speed averaged over all days in 2016 with associated atmospheric

732   phenomena: (from left to right) Cold pool; Westerly Gap flow, and Easterly Gap Flow. Event-averaged wind speed is

733   shown for each site: (a) Wasco, (b) Arlington, (c) Boardman, and (d) for 3-site composites. Wind speeds are color-

734   coded according to the color scale shown for each event on the top of this figure.

735   Figure 6. Time series of (a, d) 80-m wind speed observed at the 3 lidar sites and (b, e) wind power computed by

736   normalized power curve during a westerly gap flow/marine intrusion event on 13-18 August 2016, and cold pool event

737   on 07-12 February 2016. (c, f) Total power generation over the Bonneville Power Authority (BPA) area during these

738   periods. The bottom row shows distributions of wind speed and wind directions for the (g, h) August and (i, j) February

739   cases. Different colors indicate each site according to the legends.

740   Figure 7. Time-height cross-sections of annual-mean wind speed from (a) lidar, (b) HRRR CNTR run, and (c) HRRR

741   EXPR run for each site (from left to right columns: Wasco, Arlington, and Boardman), and for the 3-site composites.

742   Two bottom rows show the wind speed difference (bias) between (d) HRRR CNTR run and lidar, and (e) HRRR EXPR

743   run and lidar, expressed as HRRR minus lidar-measured values. White lines indicate 50 m and 150 m AGL

744   Figure 8. Annual statistics: same as Fig. 7d-e but for MAE from (a, b) HRRR CNTR and EXPR runs, (c, d)

745   HRRRNEST CNTR and EXPR runs. All panels are shown for 00z runs of both models; 12z runs were similar.

746   Figure 9. Annual MAE differences due to changes in model physics and other updates (EXPR-CNTR) made in (a)

747   HRRR and (b) HRRRNEST. Bottom two rows (c, d) show these differences expressed as percentages, normalized by

748   mean, lidar-measured wind speeds.

749   Figure 10. Annual MAE differences due to changes in model horizontal spacing (HRRRNEST-HRRR) for (a) CNTR

750   and (b) EXPR runs of models. The bottom row (c) shows MAE difference due to combined effects of changes in model

751   physics and horizontal spacing.

752   Figure 11. Time-height cross-sections of wind speed from lidar, HRRR CNTR, and HRRR EXPR 00z runs averaged

753   over (a-c) westerly gap flow. The bottom row (d) shows the MAE difference between EXPR and CNTR runs. Columns

754   are shown for (from left to right) Wasco, Arlington, Boardman sites, and 3-site composites.

755   Figure 12. Same as Fig. 11 but composited over cold-pool events.

756   Figure 13. Annual profiles of wind speed and validation metrics are shown for the (from top to bottom), Wasco,

757   Arlington, and Boardman sites, and the 3-site composites. (a) Measured wind speed is shown by a black solid line, and

758   the (a-c) CNTR and EXPR runs are shown by red and blue colors for (solid) HRRR and (dashed) HRRRNEST

759   according to the legend at the top left panel. (d) MAE improvement due to model physics: error differences (m s^{-1})

760   between EXPR and CNTR runs of (black) HRRRNEST and (gray) HRRR models for (solid) MAE, (dashed) RMSE,

761   and (dotted) bias-corrected RMSE (RMSEu). (e) MAE improvement due to finer resolution in the (red) CNTR runs

762   (HRRRNEST CNTR – HRRR CNTR) of both models, and (blue) finer resolution in the EXPR runs (HRRRNEST

22

763   EXPR – HRRR EXPR), and the (black) total effect of physics and resolution (HRRRNEST EXPR – HRRR CNTR).

764   Negative values mean improvement.

765   Figure 14. (a-b) 3-site composite profiles of error differences (m s^{-1}) between EXPR and CNTR runs of (black)

766   HRRRNEST and (gray) HRRR models: (solid) MAE, (dashed) RMSE, and (dotted) bias-corrected RMSE (RMSEu).

767   Error differences are shown for (a) calendar seasons and (b) four major events. (c-d) MAE differences due to (red) finer

768   resolution in the CNTR runs (HRRRNEST CNTR – HRRR CNTR), and (blue) finer resolution in the EXPR runs

769   (HRRRNEST EXPR – HRRR EXPR). The total effect of improved physics and model resolution (HRRRNEST EXPR

770   –HRRR CNTR) is shown by black lines. Improvements are shown for (c) calendar seasons and (d) four major events.

771   Negative values mean improvement.

772   Figure 15. Time-series of 80-m annual (15a from top to bottom) winds speed, BIAS, MAE, and the difference in

773   validation metrics between EXPR and CNTR runs of HRRR initialized at 00z.

774   Mean values in the top three panels are from (black) lidar measurements, and the (red) CNTR and (blue) EXPR runs.

775   Horizontal dashed lines of the corresponding color represent mean values of Bias and MAE from both runs further

776   averaged over the diurnal cycle. Black lines on the bottom panels show the difference between two runs (EXPR-CNTR)

777   for (solid) RMSE, (dashed) unbiased RMSE, and (dotted) MAE according to the legend at the top right corner. Fig.

778   15b is the same as Fig. 15a, except for HRRRNEST.

779   Figure 16. Relative differences in annual 80-m wind speed MAE (Δ-MAE, in %) due to the (a, b) experimental physics

780   and (c, d) model horizontal grid resolution at each site and 3-site composites, (e) the overall (physics + resolution) Δ-

781   MAE for initial time 0z, (f) same as (e) but for the initial time 12z. Each panel shows Δ-MAE for (gray) data averaged

782   for all hours, (dark blue) nighttime, and (dark yellow) daytime hours. Nighttime and daytime hours for the initial time

783   0z (Fig. 16 a-e) are 0300-1200 UTC. Nighttime and daytime hours for the initial time 12z (Fig. 16 f) are 00-1200 UTC

784   and 1500-2400 UTC. The Δ-MAEs were computed as follow: (a) [HRRR EXPR – HRRR CNTR]/HRRR CNTR; (b)

785   [HRRRNEST EXPR – HRRRNEST CNTR]/HRRRNEST CNTR; (c) [HRRRNEST CNTR – HRRR CNTR]/HRRR

786   CNTR; (d) [HRRRNEST EXPR – HRRR EXPR]/ RRR EXPR; (e-f) [HRRRNEST EXPR – HRRR CNTR]/HRRR

787   CNTR. The negative values at each panel imply improvement.

788   Figure 17. The combined physics/horizontal grid resolution relative Δ-MAE for 3-site composites of 80-m winds

789   averaged over (a) each season and (b) major event types. Combined relative MAE defined as [HRRRNEST EXPR –

790   HRRR CNTR] / HRRR CNTR.

791   Figure 18. Same as Fig. 16 a-e but for the power MAE computed by using 80 m wind speed measured by lidar and

792   predicted by CNTR and EXPR runs of HRRR and HRRRNEST for 0z initial time.

793   Figure 19. Maps of the terrain overlapped with the bias difference (colored circles) between HRRRNEST EXPR and

794   CNTR runs (Bianco et al. 2019) at the locations of 19 sodars and 3 different types of lidars as described in the text.

795   The bias difference is shown in three categories according to the legend on each map. Sites, where the absolute value

796   of the bias from EXPR runs, is smaller (larger) than from the CNTR runs indicated some improvement (degradation)

797   in model physics are shown correspondingly by blue (red) circles and indicated in the legend as "Yes" ("No"). Sites,

798 where biases from CNTR and EXPR runs are almost equal, are indicated by yellow and "0" in the legend. Sites where

799 observed or modeled data were not available (no data) are indicated by white circles. Maps are shown for mean values

800 over reforecast periods in the (a) winter, (b) summer seasons, and (c) all four-seasons of WFIP2.

801 Figure 20. Scatter plots between modeled of 80-m wind speed extracted at the lidar sites using bi-linear interpolation

802 or nearest grid point value: (a) for HRRRNEST (750 m horizontal grid) and (b) HRRR (3 km horizontal grid) models.

803 Red and blue points on all panels represent CNTR and EXPR runs respectively.

804 Figure 21. The annual distributions of 80-m wind speeds and wind directions at each site are shown for (red) control

805 and (blue) experimental runs of (a) HRRR and (b) HRRRNEST models. Lidar-measured variables are shown by black

806 curves at each panel. Gray shaded areas indicate wind speeds (4-12 m s$^{-1)}$ in a non-rated region of a typical wind turbine

807 power curve (Dupont et al. 2017).

808 **Tables with Captions**

809 Table 1. The number of days (#) having observed event types during reforecast periods in 2016, 12 months of concurrent

810 Doppler lidar measurements in 2016, and 18 months of the WFIP2 experiment. Occurrence (%) of a particular event

811 type during each period relative to the total occurrence of events.

	Cold Pool		W-Gap Flow		E-Flow		Mount. Wakes		Topo. Wakes		CO		Other		Total
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#
WFIP2	136	17	339	43	92	12	90	11	80	10	12	2	39	5	788
2016	64	13	237	47	57	11	63	13	55	11	9	2	21	4	506
Refcst	22	9	110	46	26	11	33	14	27	11	5	2	16	7	239

812

813   Table. 2. The overall impact of the updates in model physics and horizontal grid resolution on wind speed and power

814   Δ-MAE.

	Wasco		Arlington		Boardman		3 sites	
	Speed	*Power*	*Speed*	*Power*	*Speed*	*Power*	*Speed*	*Power*
All	-7.13	-13.71	-1.35	-9.11	-5.61	-7.01	-4.76	-10.59
Nighttime	-10.65	-15.44	2.06	-4.43	-6.85	-6.17	-5.31	-9.02
Daytime	-3.81	-12.14	-4.58	-13.35	-4.43	-7.78	-4.23	-12.02

815

816 Table 3. Observed 80-m mean wind speed and Bias of HRRR and HRRRNEST CNTR and EXPR runs during the
817 WFIP2 reforecast period (February 2016-March 2017) at 23 sites. The first three columns show site ID, site elevation
818 above sea level (ASL, m), and terrain complexity expressed as standard deviation (STD) of the elevation within 3 km
819 circle around each site.

Site ID	Elev. (m)	Terrain std (m)	Instrument	Obs Speed (m s⁻¹)	Bias (m s⁻¹)			
					CNT 3 km	EXP 3 km	CNT 750 m	EXP 750 m
AO1	706	64	sodar	5.67	0.069	-0.111	0.05	0.015
AO2	356	13	sodar	6.04	-0.009	-0.195	-0.04	-0.087
AO3	116	12	sodar	4.78	-0.164	-0.324	-0.183	-0.209
AO4	432	34	sodar	6.04	0.145	-0.063	0.048	-0.039
AO5	456	13	sodar	6.3	0.125	-0.048	0.015	-0.043
AO6	731	81	sodar	6.85	-0.025	-0.151	-0.053	-0.075
AO7	166	55	sodar	4.78	0.078	-0.133	0.157	0.067
AO8	703	98	sodar	4.43	-0.148	-0.183	0.106	0.084
AO9	836	57	sodar	2.47	0.84	0.59	0.8	0.681
ARL	266	56	200S	4.78	-0.133	-0.276	-0.144	-0.214
BOR	112	6	sodar	4.78	-0.092	-0.248	-0.091	-0.14
CDN	891	25	sodar	6.17	0.013	-0.102	-0.008	-0.032
DCR	795	26	sodar	5.35	0.185	0.02	0.108	0.084
GDL	501	16	sodar	4.32	0.234	0.112	0.039	0.042
GDR	725	81	WindCube	7.19	-0.067	-0.18	-0.078	-0.1
PVE	991	42	sodar	4.16	0.152	-0.026	0.144	0.135
RFS	62	80	sodar	4.78	0.095	-0.083	-0.026	-0.062
RTK	708	19	sodar	5.02	0.29	0.156	0.188	0.181
VCR	542	69	ZephIR	7.13	-0.098	-0.255	-0.067	-0.112
WCO	462	25	sodar	6.67	0.07	-0.108	-0.027	-0.082
WWL	382	34	sodar	4.18	0.114	-0.109	0.082	0.051
YKM	330	19	sodar	2.82	0.213	0.095	0.109	0.102
AO1	706	64	sodar	5.67	0.069	-0.111	0.05	0.015

820

821   Table 4. Statistics of the correlation between model outputs of 80 m wind speed extracted at the lidar sites using bi-

822   linear (Bi) interpolation or nearest grid point (N) value.

Site	Model	R^2	intercept	slope	Mean Bi	STD Bi	Mean N	STD N	Speed diff
				HRRRNEST					
	CNTR								
Wasco		0.99	0.06	0.99	6.88	3.82	6.93	3.82	-0.05
Arlington		0.99	0.08	1.02	5.73	3.57	5.91	3.65	-0.19
Boardman		0.99	0.05	1.00	5.48	3.86	5.53	3.86	-0.05
	EXPR								
Wasco		0.99	0.05	1.00	6.53	3.71	6.58	3.73	-0.06
Arlington		0.99	0.06	1.03	5.24	3.32	5.46	3.43	-0.22
Boardman		0.99	0.04	1.00	5.16	3.57	5.20	3.57	-0.04
				HRRR					
	CNTR								
Wasco		0.99	0.03	0.99	7.03	3.84	7.00	3.82	0.03
Arlington		0.99	0.02	0.97	5.71	3.55	5.55	3.46	0.16
Boardman		0.99	-0.04	1.02	5.16	3.69	5.22	3.76	-0.06
	EXPR								
Wasco		0.99	0.02	1.00	5.98	3.38	6.00	3.39	-0.02
Arlington		0.99	-0.05	1.00	4.56	2.76	4.49	2.76	0.06
Boardman		0.99	-0.04	1.01	4.32	2.98	4.31	3.01	0.01

823

27

824   Table 5. Mean values for distributions of 80 m wind speed and direction from Fig. 21

| Site | | Westerly flows | | | | Easterly flows | | | |
| | | HRRRNEST | | HRRR | | HRRRNEST | | HRRR | |
		Speed	Dir	Speed	Dir	Speed	Dir	Speed	Dir
Wasco	*Lidar*	7.7	270.5	7.9	271.0	4.1	92.4	4.2	91.5
	CNTR	7.2	265.8	8.0	267.5	4.3	95.1	4.8	87.7
	EXPR	6.9	266.0	7.0	267.7	3.9	94.9	3.9	88.9
Arlington	*Lidar*	7.2	263.6	7.4	263.7	3.1	73.4	3.2	73.2
	CNTR	6.1	259.7	6.8	261.5	3.3	64.5	3.6	71.5
	EXPR	5.6	258.2	5.5	258.0	3.1	64.9	3.0	71.2
Boardman	*Lidar*	7.0	251.4	7.1	252.0	2.8	85.5	2.8	84.8
	CNTR	6.3	252.1	6.3	249.4	2.6	54.8	2.9	57.5
	EXPR	5.8	250.4	5.4	247.0	2.5	57.9	2.5	58.6

825

826   Table 6. Percent of wind speeds in each region of a power curve of a typical wind turbine.

		HRRR			HRRRNEST		
Wasco		*Lidar*	*CNTR*	*EXPR*	*Lidar*	*CNTR*	*EXPR*
	0-4 ms⁻¹	40	35	43	42	40	43
	5-12 ms⁻¹	52	57	54	51	54	52
	13-25 ms⁻¹	8	8	3	7	6	5
Arlington							
	0-4 ms⁻¹	41	48	58	42	50	55
	5-12 ms⁻¹	50	49	42	50	47	43
	13-25 ms⁻¹	9	3	0	8	3	2
Boardman							
	0-4 ms⁻¹	54	55	61	54	55	57
	5-12 ms⁻¹	41	43	39	41	42	42
	13-25 ms⁻¹	5	2	1	5	3	2

827

828   Table 7. Summary of the HRRR and HRRRNEST reforecast configurations.

Acronym	Description	Initial- and Lateral-Boundary Conditions	Physics	Forecast Cadence and Forecast Length
HRRR	WFIP2 provisional HRRR (see Fig. 2a, large green box) 3-km horizontal grid spacing, 50 vertical layers	Initial: cold start from operational RAP; no data assimilation or prior cycling Lateral boundary: operational RAP	Control (CNTR); see Table 8, second column	Twice-daily forecast initializations (i.e., 0000 and 1200 UTC) Forecasts to 24 h, output at 15-min intervals
			Experimental (EXP); see Table 8, third column	
HRRRNEST	One-way concurrent nest within WFIP2 provisional HRRR (see Fig. 2a, small green box) 750-m horizontal grid spacing, 50 vertical layers	Initial: parent HRRR, after 3-h "spin-up" period Lateral boundary: parent HRRR	Control (CNTR); see Table 8, second column	As in HRRR, except nest initialization delayed 3 h
			Experimental (EXP); see Table 8, third column	

829

830 Table 8. Description of the control and experimental physics configurations used in the reforecasts (adapted from
831 Table 2 in Olson et al. 2019), where model developments in support of WFIP2 are shown in bold. All components
832 not under development in WFIP2 (regular font) are described in Benjamin et al. (2016)

Model Component	Control (CNTR)	Experimental (EXPR)
Land Surface	RUC 9-level	RUC 9-level
Surface Layer	MYNN	**MYNN**
PBL	MYNN level 2.5	**MYNN-EDMF**
SW Radiation	RRTMG	RRTMG
LW Radiation	RRTMG	RRTMG
Microphysics	Thompson Aero	Thompson Aero
Shallow Convection	---	**MYNN-EDMF**
Horizontal Diffusion	Smagorinsky on sigma levels	**Smagorinsky on cartesian coordinates**
Orographic Drag	---	**Small-scale gravity wave drag (inactive when $\Delta x \leq 1$ km)**
Wind-Farm Drag	---	**Fitch et al. 2012**
Vertical Levels	51 levels	51 levels
Vertical Coordinate	sigma	**Hybrid sigma-pressure**
Albedo	AVHRR	AVHRR

833
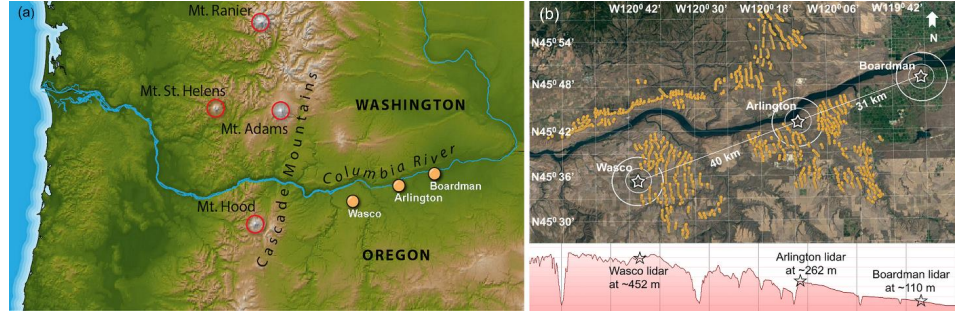
834

835 Figure 1. (a) Map of the Columbia River Valley with the location of scanning Doppler lidars denoted by gold-filled

836 circles. Red circles indicate the highest peaks of the Cascade Mountains. (b) Google Earth detailed map of the study

837 area shows the location of two NOAA Doppler scanning lidars (200S) at Wasco (452 m ASL) and Arlington (262 m

838 ASL). The Notre Dame University Halo Streamline XR lidar is located at the Boardman site at 110 m ASL. The

839 white line indicates the east-west transect of the study region along the prevalent wind directions observed from

840 surface measurements and models during previous studies in this area and confirmed from the analysis of Doppler

841 lidar measurements in Pichugina et al. (2019). The surrounding wind turbines (total rated capacity of ~3,800 MW)

842 are indicated by the clusters of dark yellow circles. An elevation gradient along the transect line is shown in (c).

843

Figure 2. (a) Domains of WFIP2 (green) HRRR and HRRRNEST models are shown along with the operational (white)
RAP (13-km grid spacing) and (red) HRRR-NCEP (3-km grid spacing) models. (Fig. 2a adapted from Olson et al.
2019). (b) An example of HRRRNEST-modeled wind flow at 80 m superimposed on area topography from WRF V3.7.
The example is taken from a 7-h forecast initialized at 000UTC on 15 May 2016. Locations of Doppler lidars are shown
by black circles. (c) Maps of the terrain elevations (in meters) at Wasco, Arlington, and Boardman research sites. The
location of each lidar (dark yellow triangles) is shown relative to the HRRR 3-km model corners of the grid cells (red
circles), and relative to the HRRRNEST 750-m model corners of the grid cells (white circles).

851
852 Figure 3. Atmospheric events observed over the WFIP2 research area from the analysis of models, satellites, and all
853 available ground observations. (a) The number of events observed during 18 months of the experiment, 12 months of
854 2016, and reforecast periods in 2016 shown by colors according to the legend. Multiple events can occur on the same
855 day. (b) Relative number (%) of major events in each month observed during the reforecast periods.
856

857

858 Figure 4. Time-height cross-sections of wind speed and wind direction at the (from top to bottom) Wasco, Arlington,

859 and Boardman site during atmospheric phenomena observed in the WFIP2 area: (a) Cold pool, (b) Westerly Gap Flow,

860 (c) Easterly Gap Flow, and (d) Mountain Waves. Wind speed is color-coded from 0 to 15 m s^{-1} according to the color

861 scale at the top of each panel. Black arrows show wind direction. Two white horizontal lines on each panel indicate

862 typical rotor-disk height, 50-150 m AGL.

863

864

865 Figure 5. Time-height cross-sections of wind speed averaged over all days in 2016 with associated atmospheric

866 phenomena: (from left to right) Cold pool; Westerly Gap flow, and Easterly Gap Flow. Event-averaged wind speed is

867 shown for each site: (a) Wasco, (b) Arlington, (c) Boardman, and (d) for 3-site composites. Wind speeds are color-

868 coded according to the color scale shown for each event on the top of this figure.

869

09 January 2024 21:36:56

870

Figure 6. Time series of (a, d) 80-m wind speed observed at the 3 lidar sites and (b, e) wind power computed by
normalized power curve during a westerly gap flow/marine intrusion event on 13-18 August 2016, and cold pool event
on 07-12 February 2016. (c, f) Total power generation over the Bonneville Power Authority (BPA) area during these
periods. The bottom row shows distributions of wind speed and wind directions for the (g, h) August and (i, j) February
cases. Different colors indicate each site according to the legends.

Figure 7. Time-height cross-sections of annual-mean wind speed from (a) lidar, (b) HRRR CNTR run, and (c) HRRR EXPR run for each site (from left to right columns: Wasco, Arlington, and Boardman), and for the 3-site composites. Two bottom rows show the wind speed difference (bias) between (d) HRRR CNTR run and lidar, and (e) HRRR EXPR run and lidar, expressed as HRRR minus lidar-measured values. White lines indicate 50 m and 150 m AGL.

883
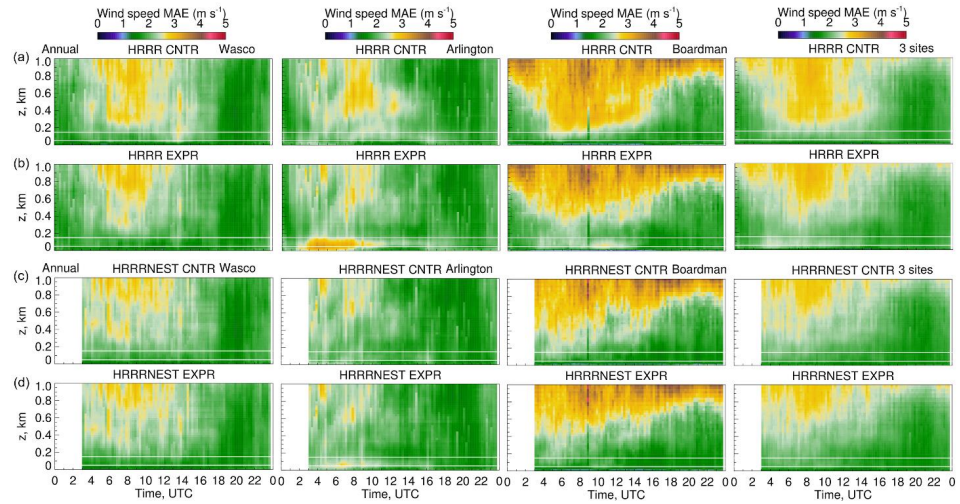
884

885 Figure 8. Annual statistics: same as Fig. 7d-e but for MAE from (a, b) HRRR CNTR and EXPR runs, (c, d)
886 HRRRNEST CNTR and EXPR runs. All panels are shown for 00z runs of both models; 12z runs were similar.

887

888

889
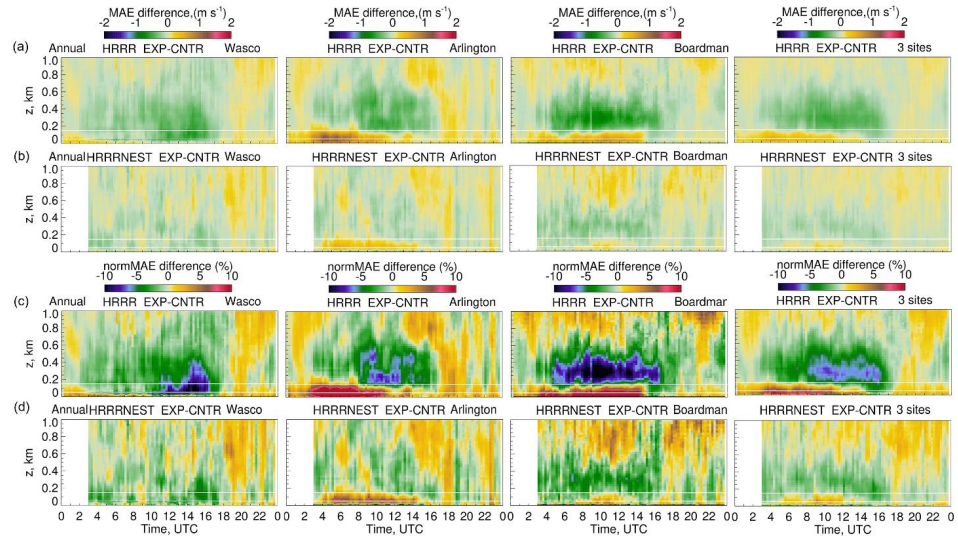
Figure 9. Annual MAE differences due to changes in model physics and other updates (EXPR-CNTR) made in (a) HRRR and (b) HRRRNEST. Bottom two rows (c, d) show these differences expressed as percentages, normalized by mean, lidar-measured wind speeds.

Figure 10. Annual MAE differences due to changes in model horizontal spacing (HRRRNEST-HRRR) for (a) CNTR and (b) EXPR runs of models. The bottom row (c) shows MAE difference due to combined effects of changes in model physics and horizontal spacing.

Figure 11. Time-height cross-sections of wind speed from lidar, HRRR CNTR, and HRRR EXPR 00z runs averaged over (a-c) westerly gap flow. The bottom row (d) shows the MAE difference between EXPR and CNTR runs. Columns are shown for (from left to right) Wasco, Arlington, Boardman sites, and 3-site composites.

905
906   Figure 12. Same as Fig. 11 but composited over cold-pool events.

09 January 2024 21:36:56

09 January 2024 21:36:56
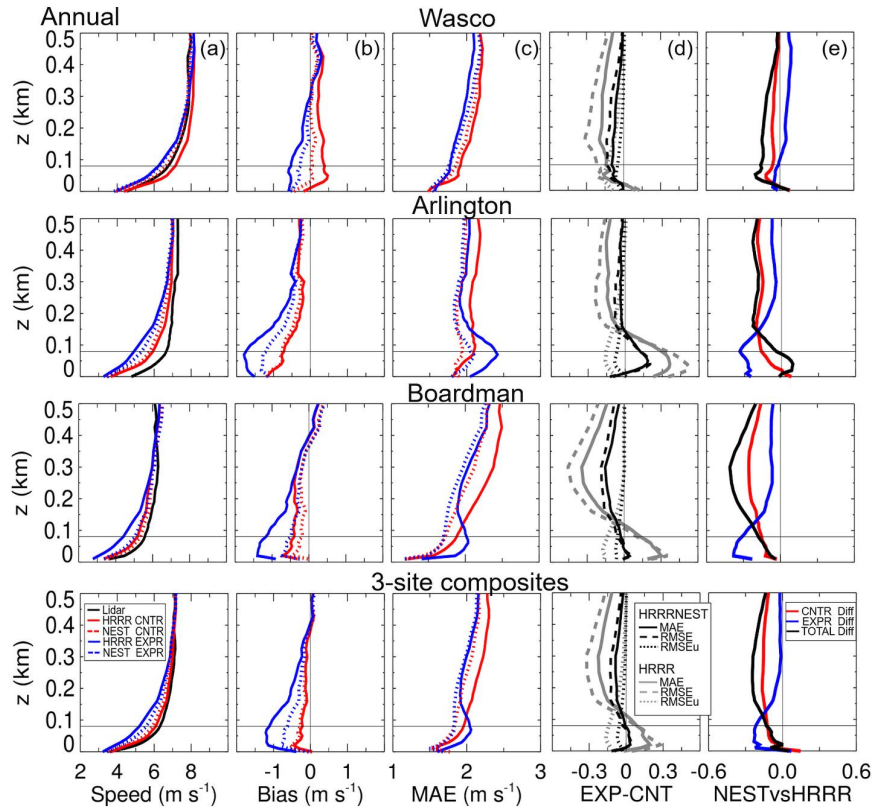
Figure 13. Annual profiles of wind speed and validation metrics are shown for the (from top to bottom), Wasco, Arlington, and Boardman sites, and the 3-site composites. (a) Measured wind speed is shown by a black solid line, and the (a-c) CNTR and EXPR runs are shown by red and blue colors for (solid) HRRR and (dashed) HRRRNEST according to the legend at the top left panel. (d) MAE improvement due to model physics: error differences (m s^{-1}) between EXPR and CNTR runs of (black) HRRRNEST and (gray) HRRR models for (solid) MAE, (dashed) RMSE, and (dotted) bias-corrected RMSE (RMSEu). (e) MAE improvement due to finer resolution in the (red) CNTR runs (HRRRNEST CNTR – HRRR CNTR) of both models, and (blue) finer resolution in the EXPR runs (HRRRNEST EXPR – HRRR EXPR), and the (black) total effect of physics and resolution (HRRRNEST EXPR – HRRR CNTR). Negative values mean improvement.

09 January 2024 21:36:56

917

Figure 14. (a-b) 3-site composite profiles of error differences (m s^{-1}) between EXPR and CNTR runs of (black)
HRRRNEST and (gray) HRRR models: (solid) MAE, (dashed) RMSE, and (dotted) bias-corrected RMSE (RMSEu).
Error differences are shown for (a) calendar seasons and (b) four major events. (c-d) MAE differences due to (red) finer
resolution in the CNTR runs (HRRRNEST CNTR – HRRR CNTR), and (blue) finer resolution in the EXPR runs
(HRRRNEST EXPR – HRRR EXPR). The total effect of improved physics and model resolution (HRRRNEST EXPR
–HRRR CNTR) is shown by black lines. Improvements are shown for (c) calendar seasons and (d) four major events.
Negative values mean improvement.

45

(a) Annual HRRR

(b) Annual HRRRNEST

925

926   Figure 15. Time-series of 80-m annual (15a from top to bottom) winds speed, BIAS, MAE, and the difference in
927   validation metrics between EXPR and CNTR runs of HRRR initialized at 00z.
928   Mean values in the top three panels are from (black) lidar measurements, and the (red) CNTR and (blue) EXPR runs.
929   Horizontal dashed lines of the corresponding color represent mean values of Bias and MAE from both runs further
930   averaged over the diurnal cycle. Black lines on the bottom panels show the difference between two runs (EXPR-CNTR)
931   for (solid) RMSE, (dashed) unbiased RMSE, and (dotted) MAE according to the legend at the top right corner. Fig.
932   15b is the same as Fig. 15a, except for HRRRNEST.

933

Figure 16. Relative differences in annual 80-m wind speed MAE (Δ-MAE, in %) due to the (a, b) experimental physics

and (c, d) model horizontal grid resolution at each site and 3-site composites, (e) the overall (physics + resolution) Δ-

MAE for initial time 0z, (f) same as (e) but for the initial time 12z. Each panel shows Δ-MAE for (gray) data averaged

for all hours, (dark blue) nighttime, and (dark yellow) daytime hours. Nighttime and daytime hours for the initial time

0z (Fig. 16 a-e) are 0300-1200 UTC. Nighttime and daytime hours for the initial time 12z (Fig. 16 f) are 00-1200 UTC

and 1500-2400 UTC. The Δ-MAEs were computed as follow: (a) [HRRR EXPR – HRRR CNTR]/HRRR CNTR; (b)

[HRRRNEST EXPR – HRRRNEST CNTR]/HRRRNEST CNTR; (c) [HRRRNEST CNTR – HRRR CNTR]/HRRR

CNTR; (d) [HRRRNEST EXPR – HRRR EXPR]/ RRR EXPR; (e-f) [HRRRNEST EXPR – HRRR CNTR]/HRRR

CNTR. The negative values at each panel imply improvement.

943
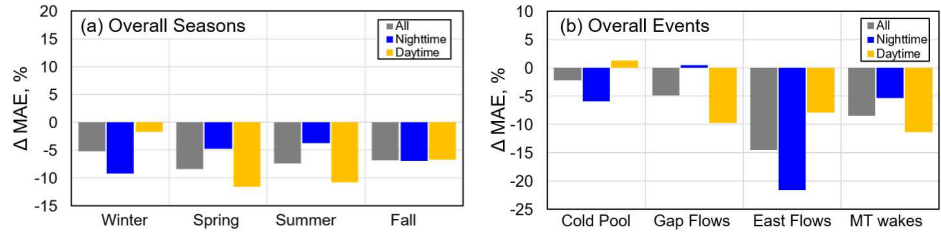
Figure 17. The combined physics/horizontal grid resolution relative Δ-MAE for 3-site composites of 80-m winds averaged over (a) each season and (b) major event types. Combined relative MAE defined as [HRRRNEST EXPR – HRRR CNTR] / HRRR CNTR.

09 January 2024 21:36:56
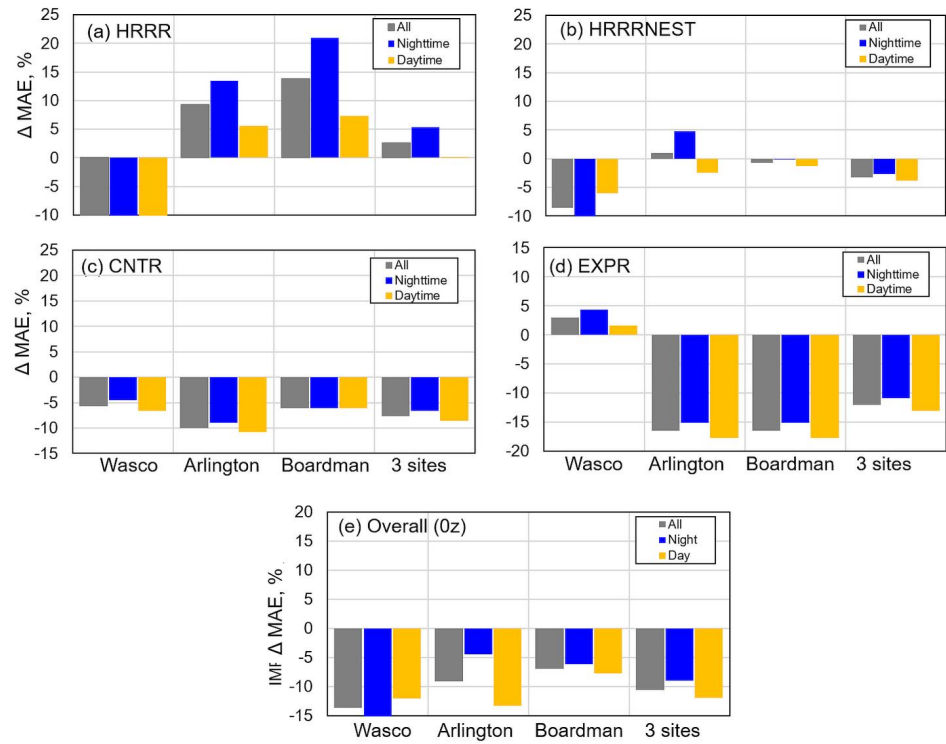
949

950 Figure 18. Same as Fig. 16 a-e but for the power MAE computed by using 80 m wind speed measured by lidar and
951 predicted by CNTR and EXPR runs of HRRR and HRRRNEST for 0z initial time.

952

953   Figure 19. Maps of the terrain overlapped with the bias difference (colored circles) between HRRRNEST EXPR and

954   CNTR runs (Bianco et al. 2019) at the locations of 19 sodars and 3 different types of lidars as described in the text.

955   The bias difference is shown in three categories according to the legend on each map. Sites, where the absolute value

956   of the bias from EXPR runs, is smaller (larger) than from the CNTR runs indicated some improvement (degradation)

957   in model physics are shown correspondingly by blue (red) circles and indicated in the legend as "Yes" ("No"). Sites,

958   where biases from CNTR and EXPR runs are almost equal, are indicated by yellow and "0" in the legend. Sites where

959   observed or modeled data were not available (no data) are indicated by white circles. Maps are shown for mean values

960   over reforecast periods in the (a) winter, (b) summer seasons, and (c) all four-seasons of WFIP2.

961

962   Fig. 20. Scatter plots between modeled of 80-m wind speed extracted at the lidar sites using bi-linear interpolation or

963   nearest grid point value: (a) for HRRRNEST (750 m horizontal grid) and (b) HRRR (3 km horizontal grid) models.

964   Red and blue points on all panels represent CNTR and EXPR runs respectively.

965

966

Figure 21. The annual distributions of 80-m wind speeds and wind directions at each site are shown for (red) control
and (blue) experimental runs of (a) HRRR and (b) HRRRNEST models. Lidar-measured variables are shown by black
curves at each panel. Gray shaded areas indicate wind speeds (4-12 m s^{-1}) in a non-rated region of a typical wind turbine
power curve (Dupont et al. 2017).

967
968
969
970

09 January 2024 21:36:56

09 January 2024 21:36:56

09 January 2024 21:36:56

09 January 2024 21:36:56

(a) Annual — MAE difference, (m s⁻¹) — HRRR EXP–CNTR — Wasco / Arlington / Boardman / 3 sites

(b) Annual — HRRRNEST EXP–CNTR — Wasco / Arlington / Boardman / 3 sites

(c) Annual — normMAE difference (%) — HRRR EXP–CNTR — Wasco / Arlington / Boardman / 3 sites

(d) Annual — HRRRNEST EXP–CNTR — Wasco / Arlington / Boardman / 3 sites

09 January 2024 21:36:56

09 January 2024 21:36:56

09 January 2024 21:36:56

09 January 2024 21:36:56

(a) Annual HRRR

(b) Annual HRRRNEST

09 January 2024 21:36:56

09 January 2024 21:36:56