



# Comparison of Observations and Predictions of Daytime Planetary-Boundary-Layer Heights and Surface Meteorological Variables in the Columbia River Gorge and Basin During the Second Wind Forecast Improvement Project

Laura Bianco<sup>1,2,6</sup>  · P. Muradyan<sup>3</sup> · I. Djalalova<sup>1,2</sup> · J. M. Wilczak<sup>2</sup> · J. B. Olson<sup>4</sup> · J. S. Kenyon<sup>1,4</sup> · R. Kotamarthi<sup>3</sup> · K. Lantz<sup>1,5</sup> · C. N. Long<sup>1,5</sup> · D. D. Turner<sup>4</sup>

Received: 11 August 2020 / Accepted: 23 June 2021 / Published online: 17 August 2021  
© The Author(s) 2021

## Abstract

The second Wind Forecast Improvement Project (WFIP2) is an 18-month field campaign in the Pacific Northwest U.S.A., whose goal is to improve the accuracy of numerical-weather-prediction forecasts in complex terrain. The WFIP2 campaign involved the deployment of a large suite of in situ and remote sensing instrumentation, including eight 915-MHz wind-profiling radars, and surface meteorological stations. The evolution and annual variability of the daytime convective planetary-boundary-layer (PBL) height is investigated using the wind-profiling radars. Three models with different horizontal grid spacing are evaluated: the Rapid Refresh, the High-Resolution Rapid Refresh, and its nested version. The results are used to assess errors in the prediction of PBL height within the experimental and control versions of the models, with the experimental versions including changes and additions to the model parametrizations developed during the field campaign, and the control version using the parametrizations present in the National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction operational version of the models at the start of the project. Results show that the high-resolution models outperform the low-resolution versions, the experimental versions perform better compared with the control versions, model PBL height estimations are more accurate on cloud-free days, and model estimates of the PBL height growth rate are more accurate than model estimates of the rate of decay. Finally, using surface sensors, we assess surface meteorological variables, finding improved surface irradiance and, to a lesser extent, improved 2-m temperature in the experimental version of the model.

**Keywords** Convective boundary layer · Model evaluation · Wind-profiling radars · Planetary-boundary-layer height

---

C. N. Long: Deceased, 19 November, 2019

---

✉ Laura Bianco  
Laura.Bianco@noaa.gov

Extended author information available on the last page of the article

## 1 Introduction

The second Wind Forecast Improvement Project (WFIP2) is a Department of Energy (DOE) and National Oceanic and Atmospheric Administration (NOAA) sponsored program (including partners from public and private institutions) whose goal is to improve the accuracy of numerical-weather-prediction (NWP) forecasts in complex terrain. WFIP2 includes an 18-month field campaign in the Pacific Northwest U.S.A. During the WFIP2, a large set of instruments was deployed in the states of Washington and Oregon, U.S.A., particularly in the Columbia River Gorge and basin area, to improve understanding of flows in complex terrain, and to assess and improve the accuracy of several of NOAA's weather forecast models. This area is well known for its high wind-energy generation, with approximately 6 GW of total installed capacity in the WFIP2 study region at the time of the field campaign.<sup>1</sup> This area is also unique for planetary-boundary-layer (PBL) processes driven or affected by the local topography, which are not always easy to forecast. Among these phenomena are cold pools (Whiteman et al. 2001; Zhong et al. 2001; McCaffrey et al. 2019), gap flows, both westerly (Sharp and Mass 2002, 2004; Banta et al. 2019), and easterly (Neiman et al. 2018, 2019), mountain waves (Durran 1990, 2003; Draxl et al. 2021), topographic wakes, and convective outflows. The 18-month duration of the campaign (October 2015–March 2017) allowed the sampling of the above-mentioned atmospheric phenomena that are important in this area throughout the annual cycle.

Details on the campaign, both from the observational and from the model perspectives, are presented in Shaw et al. (2019), Wilczak et al. (2019a, b), and Olson et al. (2019a). The objectives of the campaign were to strategically deploy suites of instruments to characterize the physical processes that drive key atmospheric phenomena in this area of complex terrain, to assess the shortcomings of the models at forecasting these particular processes, and finally to improve the model parametrizations, particularly with the scope of improving the forecast of the wind-turbine hub-height wind speed (hub-height wind-speed-model output is at 80 m; more specifically, Wiser and Bolinger 2018, mention that the average hub height of turbines installed in the U.S.A. in 2018 was 88.1 m).

The principal NWP models investigated during the WFIP2 campaign were NOAA's shorter-range models that are reinitialized and run on an hourly basis: the Rapid Refresh (RAP, 13-km grid spacing), the High-Resolution Rapid Refresh (HRRR, 3-km grid spacing) and its nested version (HRRRNEST, 750-m grid spacing). For the duration of the campaign, the models were continuously monitored and compared with the observations using the WFIP2 real-time model evaluation website (<http://wfip.esrl.noaa.gov/psd/programs/wfip2/>) which helped in the selection of case studies. Using these case studies, tests were iteratively performed on the model parametrizations, modifying them or adding completely new parametrizations in the process. The final parametrizations selected at the end of this process were then included in the experimental versions of the models, including revisions to the local and non-local turbulent mixing, subgrid-scale clouds and coupling to radiation, the addition of surface drag due to subgrid-scale (SGS) orography, and modification of the horizontal diffusion. Olson et al. (2019a, b) detail all these changes and additions to the model parameterizations, which are also summarized below.

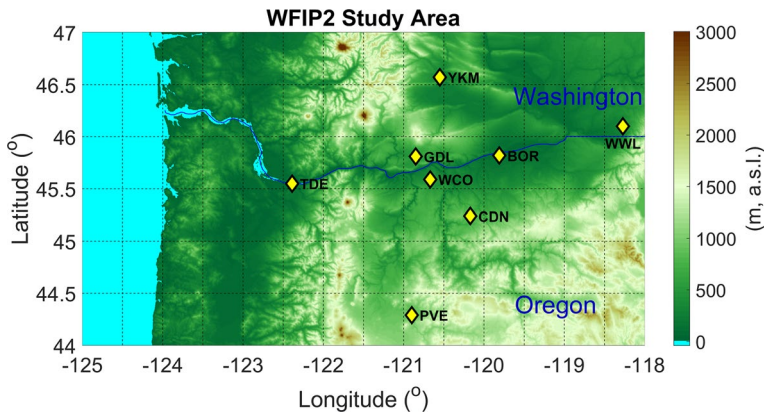
<sup>1</sup> A distribution of the wind turbines in the study area can be found here: <https://eerscmap.usgs.gov/uswtodb/viewer/#6.48/46/-120>.

To measure the final impact of these parametrization improvements and additions, toward the end of the campaign specific periods were selected to compare the experimental and control versions of the RAP, HRRR, and HRRRNEST models. These specific periods consisted of two 10-day periods in winter and summer 2016 (hereafter referred to as retrospective runs), and one 6-week period for each season (hereafter referred to as reforecast runs).

Although the main goal of the project was to improve the hub-height wind speed (with results presented in Bianco et al. 2019; Olson et al. 2019a; Pichugina et al. 2019,2020), robust model improvement occurs only if other forecast variables were not sacrificed. Several parallel studies were performed to assess how other key variables in the boundary layer were affected. For instance, Olson et al. (2019a) used surface observations to compare 2-m temperature and 10-m wind speed from the upgraded models with their previous versions over the entire contiguous U.S.A. domain, as well as the same set of wind-profiling radars (WPRs) (and their associated radio acoustic sounding systems, RASS) to assess model improvements on wind speed and temperature through the entire boundary layer. Mean absolute errors on wind speeds were found to be reduced over all four reforecast periods in the WFIP2 domain, especially at night and in winter (stable atmospheric conditions), and particularly in the rotor layer (layer of the atmosphere in which the rotor blades of the wind turbines rotate). Improved temperatures are mostly found in winter, particularly in cold-pool events, during which the performance of the models was originally poor, but largely improved due to WFIP2-led efforts. Improvements in the model performance during cold-pool events are particularly important for the wind energy, since during these events the stagnant cold layer of air, persisting over multiple days, often reduces wind-energy production to near zero. Olson et al. (2019a) also found a beneficial impact in the reduction of large model errors in wind-speed forecasting, with the largest improvements in the autumn and winter. In another study, Djalalova et al. (2020) analyzed the impact of the model modifications at forecasting wind-power ramp events, while Pichugina et al. (2019) compared a full year of wind profiles from Doppler lidars at three WFIP2 sites to the operational (at the time of their study) HRRR-NCEP (National Centers for Environmental Prediction) runs, showing how model errors varied from site to site, and highlighting several aspects of where the HRRR-NCEP model needed improvement. Finally, Pichugina et al. (2020) evaluated the performance of the HRRR and HRRRNEST models during the reforecast periods at forecasting the wind speed in the lowest 1 km of the atmosphere using scanning, pulsed Doppler lidars at three of the WFIP2 sites. Additionally, several studies using WFIP2 observations and model runs have been performed or are in preparation investigating other phenomena, such as surface fluxes (Grachev et al. 2020), cloud cover, and the surface energy balance.

Model forecast errors are generally interdependent, as explained in Benjamin et al. (2016). For example, in the boundary layer, insufficient cloud coverage in the model can result in excessive downward solar irradiance, causing warm and dry conditions near the surface, which in turn contributes to more turbulent mixing and to an excessive growth of the PBL, with entrainment of dry air from the free atmosphere into the PBL, which may further reduce the PBL-top cloudiness in a positive-feedback mechanism. For this reason, any improvement at any step of this positive feedback process taking place in the boundary layer should be beneficial to the other atmospheric variables as well.

Other studies have validated PBL height estimations from models using observations, such as Doppler lidars and radiosondes (Banks et al. 2015; Collaud Coen et al. 2014; Lee and De Wekker 2016 and reference within). The unique aspect of the present study is the



**Fig. 1** Topography of the WFIP2 study area with the locations of the observational sites (yellow diamonds). The abbreviation definitions can be found in Table 1

availability of a network of daytime PBL height observations collected over a long period of time and in complex terrain.

In this study, we assess improvements in the diurnal evolution and annual variability of the convective PBL height in the models, using eight 915-MHz WPRs deployed during the WFIP2 campaign, but we also take advantage of additional available meteorological measurements to evaluate the impact of model improvements on selected surface variables.

In Sect. 2, the observational and NWP model datasets as well as the differences between the retrospective and reforecast run configurations are described, Sect. 3 details the PBL height estimation methods for the observational and model datasets, Sect. 4 presents results for the retrospective and reforecast runs, and Sect. 5 presents the conclusions.

## 2 Dataset Description

The WFIP2 campaign was held in the states of Washington and Oregon, U.S.A., in an area of complex terrain dominated by the Cascade Mountains, the Columbia River basin to their east, and the Columbia River Gorge that transects the Cascades. The topographic features of the study area are presented in Fig. 1, where the locations of the eight sites considered in this study are represented with the yellow diamonds. Latitude, longitude, elevation, institution in charge of the instrument, and the operational period at each site for each instrument are provided in Table 1.

### 2.1 Observational Dataset

#### 2.1.1 Wind-Profiling Radars

The 915-MHz WPRs at the eight sites presented in Fig. 1 were deployed for nearly the entire duration of the WFIP2 campaign, with the exception of those in Goldendale, Walla Walla, and Yakima, which started operation a few months later. The primary information derived by these instruments is from the first moments of the Doppler spectra, which gives measurements of the wind components over the lowest few kilometres

**Table 1** List of the instruments used in this study with site name, site identification name, latitude (°N), longitude (°W), elevation (m above sea level, a.s.l.), institution in charge, and period of operation

Site name Instrument type	Latitude (°N)	Longitude (°W)	Elevation (m a.s.l.)	Institution	Time in operation
<i>Boardman, OR (BOR)</i>	45.82	119.81	112		
WPR				NOAA/ARL	10/2015–03/2017
Met Station and Micro- barograph				NOAA/ARL	10/2015–03/2017
<i>Condon, OR (CDN)</i>	45.24	120.17	891		
WPR				NOAA/PSL	10/2015–03/2017
Surface Radiation				NOAA/GML	02/2016–03/2017
Met Station and Micro- barograph				NOAA/PSL	10/2015–03/2017
<i>Goldendale, OR (GDL)</i>	45.81	120.85	501		
WPR				DOE/ANL	03/2016–03/2017
<i>Prineville, OR (PVE)</i>	44.29	120.90	991		
WPR				NOAA/PSL	10/2015–03/2017
Met Station				NOAA/PSL	10/2015–03/2017
<i>Troutdale, OR (TDE)</i>	45.55	122.39	12		
WPR				NOAA/PSL	10/2015–03/2017
Met Station and Micro- barograph				NOAA/PSL	10/2015–03/2017
<i>Wasco, OR (WCO)</i>	45.59	120.67	462		
WPR				NOAA/PSL	10/2015–03/2017
Surface Radiation				NOAA/GML	02/2016–03/2017
Met Station and Micro- barograph				NOAA/PSL	10/2015–03/2017
<i>Walla Walla, WA (WWL)</i>	46.10	118.27	383		
WPR				DOE/ANL	03/2016–03/2017
<i>Yakima, WA (YKM)</i>	46.57	120.55	330		
WPR				DOE/ANL	03/2016–03/2017

Oregon and Washington are abbreviated as OR and WA, and meteorological is abbreviated as Met

ARL, Air Resources Laboratory; PSL, Physical Science Laboratory; GML, Global Monitoring Laboratory; ANL, Argonne National Laboratory

of the atmosphere (Strauch et al. 1984; Ecklund et al. 1988). In addition, WPRs are also often used for derived products, such as to detect strong gap flows (Neiman et al. 2018) or, including the information contained in the other spectral moments, to compute the eddy dissipation rate (McCaffrey et al. 2017), melting-layer height (White et al. 2002), and the PBL height (White 1993; Angevine et al. 1994; Coulter and Holdridge 1998; Cohn and Angevine 2000; Bianco et al. 2008). The WPRs are often operated in two modes (high resolution and low resolution), which cover different height ranges with different vertical resolutions. For this study, the high-resolution mode, covering between about 100 m and 2500 m, was chosen because it provides the highest vertical resolution available (about 60-m resolution), while extending high enough to cover the entire boundary layer in the study region.

## 2.1.2 Meteorological Stations and Microbarographs

Surface meteorological stations were available at five of the eight WPR sites (Boardman, Condon, Prineville, Troutdale, and Wasco) using diverse instruments for measuring quantities related to meteorology such as pressure, temperature, humidity, precipitation, and solar and net irradiance near the Earth's surface. Four of these five sites (Boardman, Condon, Troutdale, and Wasco) also had available measurements of surface pressure from high-precision fast-response barometers combined with Nishiyama–Bedard quad-disk pressure probes (Nishiyama and Bedard 1991), hereafter referred to as “microbarographs”.

The measurements of 2-m temperature, solar irradiance, mixing ratio, and pressure from the meteorological stations were used to expand the analysis of these meteorological variables at these sites. Since these microbarographs measure pressure with higher accuracy than the meteorological stations, the data from the microbarographs were used for pressure observations at the four sites where they were available.

## 2.1.3 Surface Irradiance Measurements

Two of the eight sites with WPRs (Wasco and Condon) also had complete surface irradiance measurements from collocated stations that were operational at the sites within a few months of the official campaign start date. The station at the Wasco site recorded the four components for net irradiance (upwelling and downwelling shortwave and longwave irradiance) as well as diffuse and direct solar irradiance. The station at the Condon site recorded downwelling shortwave and longwave as well as diffuse and direct shortwave irradiance. Both these stations have the necessary inputs for the RadFlux analysis (Long and Ackerman 2000; Long et al. 2006) that produces a clear-sky identification and calculation of clear-sky irradiance components.

We used the data collected at these two sites to expand the analysis into model PBL height evaluation for clear-sky versus cloudy days.

## 2.2 Numerical Weather Prediction

The WFIP2 models targeted for improvement are the RAP (13-km horizontal grid spacing, Benjamin et al. 2016), the HRRR (3-km horizontal grid spacing), and its nested version, the HRRRNEST model (750-m horizontal grid spacing). Improvements in a number of model components include:

- A PBL mixing-length revision, which allows the mixing length to be independent of the height above ground and forces the turbulent eddies to be smaller than the depth of the model vertical grid interval in strong stratification;
- A mass flux addition to the Mellor–Yamada–Nakanishi–Niino (MYNN) PBL parametrization, which allows for direct coupling of the sub-cloud convective cores and the cloud layer above, improving coverage of shallow cumulus and profiles of temperature and humidity;
- A SGS coupling of clouds and radiation, which improves the downward shortwave forcing in shallow cumulus and stratocumulus conditions, improving the surface energy balance, which can in turn drive the turbulent mixing more accurately;

**Table 2** List of the periods chosen to test the experimental and control versions of the models

Retrospective periods		February 2016 10–20/2		August 2016 14–23/8
Reforecast periods	Spring 2016	Summer 2016	Autumn 2016	Winter 2017
	25/3–7/5	24/6–7/8	24/9–7/11	25/12/2016–7/2/2017

- A representation of drag associated with SGS orography;
- A representation of wind-farm drag, which reduces a high wind speed bias within and downwind of wind farms.

Moreover, because of the need for high-resolution simulations in complex terrain, special attention was paid to incorporate scale-adaptive physics into the MYNN eddy diffusivity-mass flux and gravity-wave-drag parametrizations. More details on the improved model configurations (which were then made operational after adding some minor adjustments) are provided in Olson et al. (2019a, b).

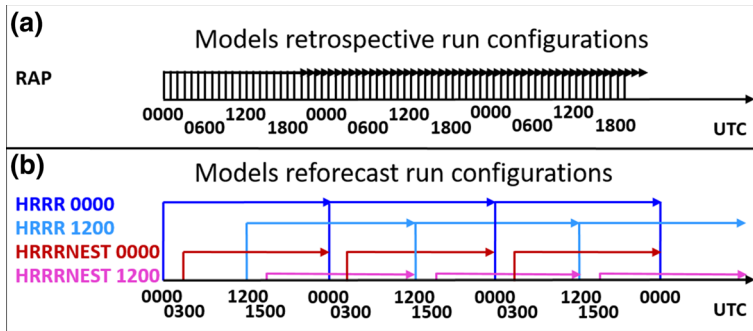
To measure the impact of these improvements, two 10-day retrospective periods (one in February 2016, and one in August 2016) were chosen to test the experimental and control versions of the models. Moreover, four (one for each season) 6-week reforecast periods were selected, over which the HRRR and HRRRNEST models were reassessed in comparison to the observations. Time frames for the retrospective and reforecast runs are presented in Table 2, and a description of these two sets of simulations is provided in the next section. The two sets of simulations are evaluated independently for their respective time periods.

### 2.2.1 Retrospective Runs

Two 10-day retrospective periods (Table 2) were chosen over which the RAP, HRRR, and HRRRNEST models were run in fully-cycled forecast-system mode, using the same data-assimilation and cycling procedures that exist in the operational RAP and HRRR models, for both the RAP (21 forecast horizons, i.e. 21 h into the future from the model initialization time), and HRRR (18 forecast horizons) models. The HRRRNEST model had no additional data assimilation, but was initialized off the 3-h HRRR forecast and ran concurrently with the HRRR run. The RAP model was run every hour, but the HRRR and HRRRNEST models were only run every third hour. All model forecast runs provided output every hour.

Due to the differences between the RAP, HRRR, and HRRRNEST model initialization and cycling designs, for the RAP retrospective runs we can create hourly model PBL height estimation time series for each forecast horizon (which can be used for a comprehensive model evaluation), while for the HRRR and HRRRNEST retrospective runs, the time series of the model PBL height for each forecast horizon contains values every third hour. Because of this, the number of PBL height estimations for the HRRR and HRRRNEST reforecast runs is significantly reduced compared with the RAP runs, producing noisier results. Also, the comparison to the RAP model would not be adequate unless we under-sampled the RAP output to align with the other models, which would degrade the quality of the results for this model as well. Therefore, we do not include these two models, but rather only look at the statistical performances of the RAP runs in the retrospective run analysis. Figure 2a presents a schematic representation of the RAP configuration for the retrospective runs.





**Fig. 2** **a** Schematic of the RAP model runs for the two 10-day retrospective runs. **b** Schematic of the HRRR and HRRRNEST model runs for the four 6-week reforecast runs. Vertical lines represent model initialization times, arrowheads represent forecast lengths

### 2.2.2 Reforecast Runs

Four 6-week reforecast periods (Table 2) were chosen for which the HRRR and HRRRNEST models were run in control and experimental configurations, for 24 forecast horizons. The models were initialized every 12 h (at 0000 UTC and at 1200 UTC, hereafter referred to as the 0000 and 1200 initialized runs, respectively). All model forecast runs provided predictions every hour, which were used for comparison with the observations. The HRRR model was run in a "cold-start" configuration, where initial conditions were supplied from the RAP model without additional data assimilation or antecedent cycling. Since there is a spin-up period associated with the model atmosphere adjusting to the higher resolution terrain, the HRRR model atmosphere is allowed to spin-up for 3 h before the HRRRNEST model is initialized from the HRRR 3-h forecast. This causes the first HRRRNEST model output to have a 3-h spin-up delay (see Fig. 2b for a schematic representation of the reforecast run configurations).

For the reforecast runs, we can compute statistics on each hour of the day for the 0000 and 1200 initialized runs, but it is not possible to calculate statistics as a function of the forecast horizon because there are only two initialization times per day. Nevertheless, we can still look at the statistical performances of the HRRR and HRRRNEST models, control and experimental runs.

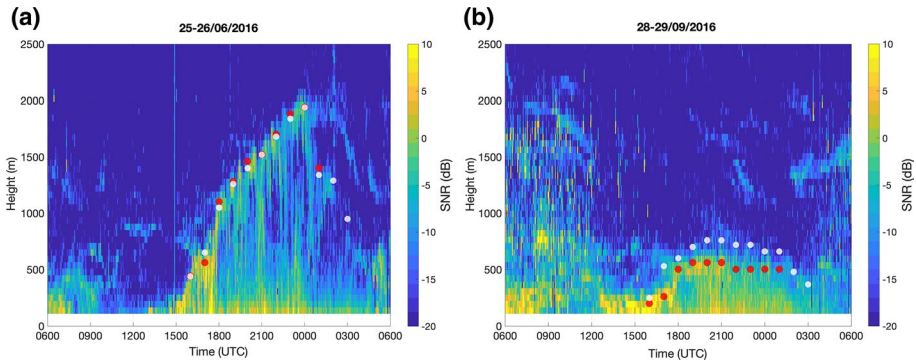
In both retrospective and reforecast runs, model output fields at the site locations are obtained by bilinearly interpolating the model grid-level output using the four closest grid points.

## 3 Estimated Planetary-Boundary-Layer Heights

### 3.1 Observed Planetary-Boundary-Layer Heights

Although WPRs are primarily used to measure the vertical profile of the horizontal wind vector (Strauch et al. 1984; Ecklund et al. 1988), it has been widely demonstrated that the information contained in the zeroth and second moments can be used for identifying the





**Fig. 3** Time–height cross-sections of SNR (in dB), with superimposed PBL height estimations determined with the fuzzy-logic (red dots) and SNR (white dots) approaches for two example days at the Condon site: 25/6/2016 (**a** sunrise at 1133 UTC and sunset at 0234 UTC) and 28/9/2016 (**b** sunrise at 1255 UTC and sunset at 0047 UTC)

top of the convective boundary layer (White 1993; Angevine et al. 1994; Coulter and Holdridge 1998; Cohn and Angevine 2000; Bianco and Wilczak 2002; Bianco et al. 2008). During the measurement campaign, two methods were used to retrieve the hourly PBL height estimates from the WPRs. The first method is the fuzzy-logic algorithm of Bianco et al. (2008) incorporating information on the returned signal-to-noise ratio (SNR) and the hourly variance and spectral width of the vertical velocity component. The PBL height is defined as the top of the lowest layer of the atmosphere that is directly affected by the surface and is continuously turbulent.

The second method, henceforth called the SNR approach, uses the automated algorithm of Coulter and Holdridge (1998). This method looks for the first and second peaks in the vertical profile of the SNR from the WPR hourly consensus measurements, and the first and second maxima in the ratio of the SNR differences above and below each range gate. The heights corresponding to these peaks are reported as four PBL height estimates. The algorithm then chooses the most frequently identified value, if any, as the PBL height (if one of the first two estimates coincides with one of the second two). If all four estimates are different, the closest value to the preceding PBL height estimate (larger from the previous estimate during the day, and smaller than the previous estimate from 1700 local standard time—at the site of interest—to sunset) is chosen.

The automated estimations of the PBL height with both the fuzzy-logic and the SNR approaches were also visually inspected to eliminate any outliers. Both methods are only effective for detecting the daytime convective PBL heights, because the WPRs have a range resolution that does not allow resolving the stable nocturnal PBL heights, which also frequently falls below the lowest height of the WPR device. Thus, the PBL height estimations are only available for times between sunrise and sunset, when the boundary layer is convective and deeper than the lowest WPR height (approximately 100 m). Moreover, PBL height estimations are not provided during precipitation times as in these cases the signal from precipitation overpowers that of clear air.

An intercomparison of the PBL heights with both methods conducted at the Condon site for the four 6-week reforecast periods (spring 2016, summer 2016, autumn 2016, winter 2017) generally shows reasonable agreement. An example of this can be seen in Fig. 3a, showing the time–height cross-section of the range-corrected SNR

with superimposed PBL height estimations using the fuzzy-logic (red dots) and the SNR (white dots) methods on 25 June 2016, with very good agreement from local sunrise to sunset, both realistically capturing the PBL height based on visual inspection of the SNR and spectral width profiles by subject-matter experts. However, occasionally the SNR approach places the PBL height consistently 100–200 m higher than the fuzzy-logic approach, as in the case of 28 September 2016 (Fig. 3b). In such cases, the composite fuzzy-logic approach, with the additional information from variance and spectral width of vertical velocity component, more realistically captures the top of the convective PBL, as confirmed by visual inspection of the SNR and spectral width profiles.

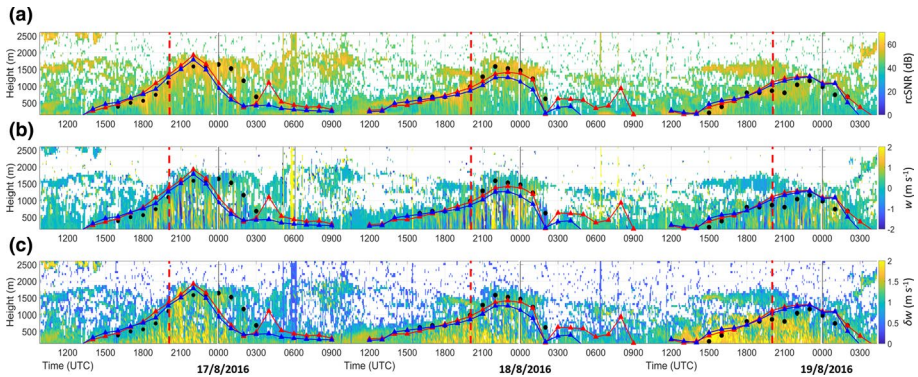
The correlation coefficients of the PBL heights determined by the SNR and the fuzzy-logic methods, and the mean difference between the two methods at the Condon site during the four reforecasting periods are  $R=0.86$  (5.8 m) for spring 2016,  $R=0.90$  (–6.25 m) for summer 2016, and  $R=0.79$  (45 m) for autumn 2016, with the number of simultaneous PBL height estimates involved in the statistics being 355, 250, and 254, respectively. While the differences between the two methods are similarly small in spring and summer, the SNR method places the PBL height lower in summer compared with the composite fuzzy-logic method (negative bias). Over winter 2017, when PBL height estimations are more challenging, both the fuzzy-logic and the SNR methods provide far fewer PBL height estimates compared with the other periods, only offering 18 simultaneous PBL heights, which prevents us from drawing conclusions on the general agreement of the two approaches during this season.

As the PBL height estimates obtained with both the fuzzy-logic and the SNR algorithms overall agree well, below only the convective PBL height estimates of the composite fuzzy-logic approach are considered.

### 3.2 Modelled Planetary-Boundary-Layer Heights

The models use the MYNN PBL parametrization, whose modifications require the PBL height as an internal variable, therefore its estimation needs particular attention. LeMone et al. (2013, 2014) showed that a potential temperature-based ( $\theta$ -based) definition of PBL height is generally accurate for convective boundary layers, and a turbulence kinetic energy-based (TKE-based) definition performs well for stable boundary layers. Olson et al. (2019b) expanded on this idea, using the  $\theta$ -based definition of PBL height of Nielsen-Gammon et al. (2008), and blending it with a TKE-based definition that uses criteria specified by Kosović and Curry (2000) and also used in Cuxart et al. (2006).

Specifically, the models search the lowest 250 m of the atmosphere for the minimum in  $\theta$  (min  $\theta$ ) and the maximum in TKE (max TKE), and determine the height where  $\theta > (\text{min } \theta + 1) ^\circ\text{C}$ , and the height where the TKE decreases below a threshold value (chosen to be 5%) of the maximum near-surface TKE  $< 5\%$  maximum TKE. Ultimately, these two definitions are blended by a hyperbolic-tangent function so that the  $\theta$ -based PBL height definition dominates for neutral and unstable conditions (when the  $\theta$ -based PBL height definition is  $> 250$  m), while the TKE-based definition dominates for stable conditions (when the  $\theta$ -based PBL height definition is  $\leq 250$  m), using a transition-layer depth of 300 m. This hybrid-PBL height algorithm has been shown to accurately diagnose the PBL height throughout the diurnal cycle (Fitch et al. 2013).



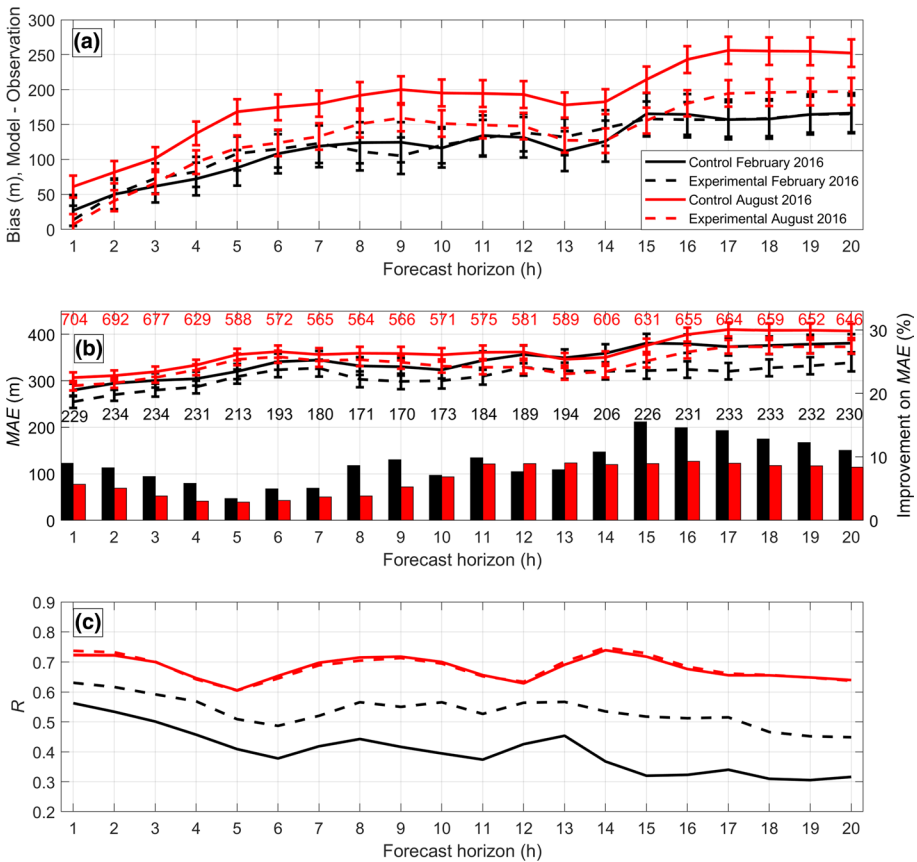
**Fig. 4** a Time–height cross-section of range-corrected SNR (rcSNR, in dB), with superimposed PBL height estimations from WPR (black circles), the RAP model control (red triangles and lines), and the RAP model experimental (blue triangles and lines) for the forecast horizon 1, at the Condon site for 16–18/8/2016. **b** As in (a) but for vertical velocity component ( $w$ , in  $m s^{-1}$ ). **c** as in (a) and (b) but for the spectral width of the vertical velocity component ( $\delta w$ , in  $m s^{-1}$ ). Vertical red-dashed bars indicate solar noon

## 4 Results

### 4.1 Retrospective Runs

An example is presented in Fig. 4 of how the RAP model (forecast horizon 1, of all hourly initialized runs) PBL heights compare to the fuzzy-logic heights for three days at the Condon site during the August 2016 retrospective run. Figure 4 represents the WPR observed range-corrected SNR, vertical velocity component, and spectral width of the vertical velocity component, the combination of which is used in the fuzzy-logic automated method to derive the PBL heights. Averaged over the three-day period, sunrise is at 1306 UTC, solar noon is at 2004 UTC, and sunset at 0301 UTC. Between sunrise and sunset, it is easy to detect the development of the convective boundary layer, which presents larger values of SNR at its top (identifiable in Fig. 4a), strong updrafts and downdrafts during the convective period (Fig. 4b), and larger values of the spectral width of the vertical velocity component (Fig. 4c) due to the turbulent nature of the boundary layer during convective periods. Figure 4 shows that the model PBL heights (RAP control and experimental shown in red and blue triangles and lines, respectively) are visually in good agreement with those obtained by the WPR (in black circles) over the daily cycle, with some instances where the model PBL heights are higher compared with those observed (for instance, during the hours 1600–2000 UTC of 16 August, and 2000–2200 of 18 August 2016 when the boundary layer was rapidly growing). During the late afternoon hours, the decay of the boundary layer sometimes occurs too early in the models (0000–0300 UTC of 17 August 2016) while on other days it is delayed (0100–0300 UTC of 19 August 2016). It can also be noted that although during night-time stable conditions the WPR PBL heights are not available for comparison, the RAP control estimations (red triangles) tend to result in much deeper and probably unrealistic values, which are lowered in the RAP experimental (blue triangles).

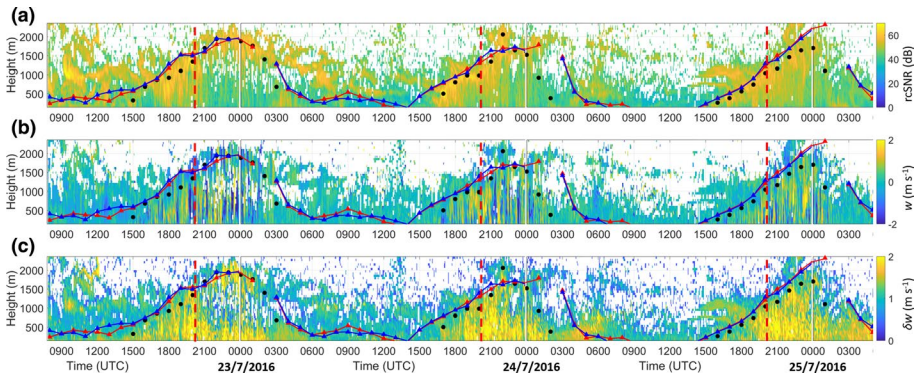
Statistical results for the February 2016 and August 2016 retrospective runs are presented in Fig. 5 for the RAP runs, all as a function of the forecast horizon. For both periods, as expected, we notice an increase in bias and mean absolute error (MAE), and a decrease in the value of  $R$  as a function of the forecast horizon (although the decrease in  $R$



**Fig. 5** **a** PBL height bias (model-observations) as a function of the forecast horizon, for the February 2016 (in black) and August 2016 (in red) RAP control (solid lines) and experimental (dashed) runs. Error bars denote  $\pm\sigma/\sqrt{n}$ , where  $\sigma$  is the standard deviation, and  $n$  the number of samples. **b** As in (a) but for MAE (lines, on the left axis) and for the improvements in the value of the MAE of the experimental versus control runs (bars on the right axis, computed as  $100 \times (\text{control} - \text{experimental}) / \text{control}$ , black for February 2016 and red for August 2016). **c** as in (a) and (b) but for  $R$  between the PBL height from the model runs and the observations

is less evident for the August 2016 retrospective period). The bias and the value of MAE are larger in August 2016 (red lines) compared with the February 2016 (black lines), as the PBL height estimates are also larger in summer compared with winter, due to stronger convective motions; in addition, the value of  $R$  is higher in August 2016 compared with February 2016, as during the wintertime the identification of the PBL height is more challenging. Also, the number of points involved in the statistics and presented in Fig. 5b for each forecast horizon (in red for the August 2016 and in black for the February 2016 retrospective periods) is much lower for February 2016 period compared with August 2016.

All model runs show an overestimation of the PBL height (positive bias). For the February 2016 retrospective run, the bias is similarly positive in both runs, but the magnitude of MAE improves in the experimental runs compared with the control runs (black bars, Fig. 5b), as well as  $R$  (black-dashed line above the solid-black line, Fig. 5c). For the August 2016 retrospective run, the bias and the MAE improve in the experimental run compared



**Fig. 6** As in Fig. 4, but for the HRRR control (red triangles and lines) and HRRRNEST control (blue triangles and lines) 0000 initialization time (forecast horizons 0–23 for the HRRR model and 3–23 for the HRRRNEST), at the Prineville site for 22–25/7/2016. Since we show the 0000 initialization runs the forecast horizon matches the clock time

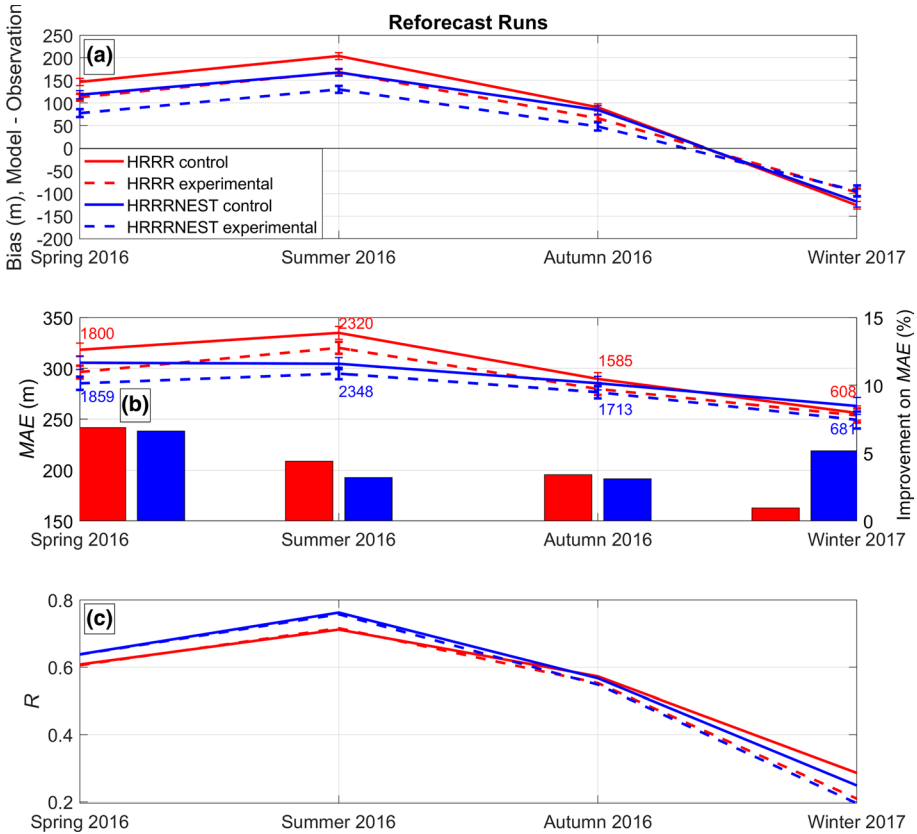
with the control run (red-dashed line below the solid-red line in the bias, Fig. 5a, and positive improvements denoted by the red bars, Fig. 5b), while the value of  $R$  is similar in the two runs. Overall the RAP experimental run provides better statistics compared with the RAP control run in both retrospective periods.

## 4.2 Reforecast Runs

An example of how the HRRR and HRRRNEST control PBL heights (in connected blue and red triangles respectively) compare to the WPR observations (in black circles) for three days during the summer 2016 reforecast run is presented in Fig. 6 for the Prineville site. We show only the forecast horizons of the 0000 initialization runs, as the other runs behave similarly. Visible in this figure is the HRRRNEST model output 3-h spin-up delay, while the HRRR model output gap at 0200 UTC is due to a missing output in the model. As in Fig. 4, the daily evolution of the convective boundary layer is clearly identifiable from the WPR observations (averaged over the three-day period, sunrise is at 1241 UTC, solar noon is at 2007 UTC, and sunset at 0333 UTC). It can be seen that while the models capture the daily cycle of the PBL heights well, there is some overestimation in comparison to the observations, particularly during periods of rapid boundary-layer growth on 1400–2000 UTC of 22/7/2016, and during the hours of boundary-layer decay on 0100–0200 UTC of 24/7/2016.

Statistical results of PBL height for the reforecast runs as a function of season are presented in Fig. 7 for the HRRR control and experimental runs (red solid and dashed lines, respectively) and for the HRRRNEST control and experimental runs (blue solid and dashed lines respectively). The 0000 and 1200 initialization time statistics are averaged together. For all periods, the experimental runs of both models result in a smaller bias and magnitude of MAE compared with the control runs, while the value of  $R$  is very similar between the control and experimental runs (except in winter). Also, the HRRRNEST model, with its smaller grid spacing, provides a smaller bias and value of MAE compared with the HRRR model. The winter 2017 reforecast period has significantly fewer points used in the statistics, as the identification of the PBL height from WPRs is in general more challenging in winter or stable conditions.

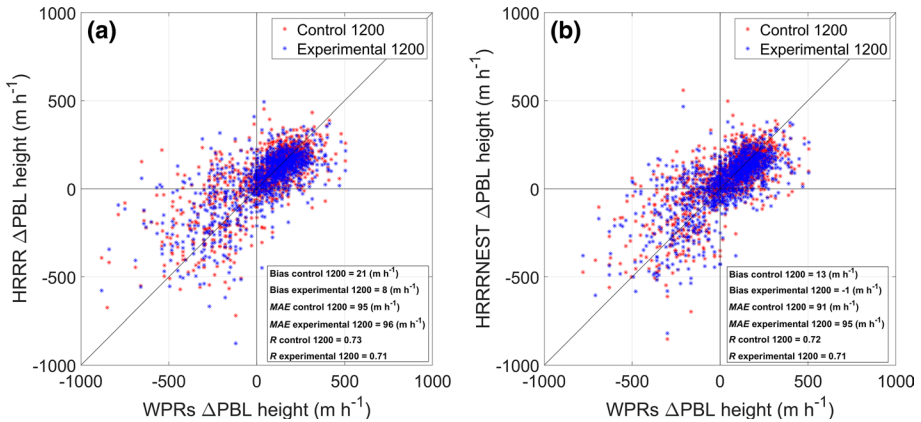




**Fig. 7** PBL height statistical results for the HRRR model (in red) and HRRRNEST model (in blue) for the four reforecast periods for the control (solid line) and experimental (dashed) runs. The panels in the figure show: **a** the bias (model-observations), **b** MAE (denoted with the lines, left axis; and the improvements of MAE of the experimental versus control runs in %, right axis), and **c** the correlation coefficient ( $R$ ) between the PBL height from the model runs and the observations for the four reforecast periods. The number of points involved in the statistics are presented in **(b)** with red numbers for the HRRR model and blue numbers for the HRRRNEST model (averaged for each model for the 0000 and 1200 initialization times)

Overall, for both the HRRR and HRRRNEST models, the experimental runs provide better statistics compared with their respective control runs, as visible in the bars presented in Fig. 7b, and the HRRRNEST model generally performs better than the HRRR model in both control and experimental runs.

Results showing the skill of the models at forecasting the time rate of change of PBL height over a period of an hour, in comparison to those observed, for all reforecast periods, are shown in the scatterplots of Fig. 8. In this case we use 1200 initialized runs, because the HRRRNEST model, with its 3-h spin-up delay, misses most of the afternoon decaying part of the PBL height development in its 0000 initialized runs. The correlation coefficients are very similar between the models (HRRR and HRRRNEST models versus observations) and between the control and experimental runs (versus observations). However, we note that positive values of the rate of change (a growing boundary layer) show better agreement with the observations compared with the negative values of the rate of change (a decaying boundary layer). Part of this larger discrepancy may be caused by the challenging



**Fig. 8** Rate of PBL height growth/decay ( $\text{m h}^{-1}$ ) of daytime HRRR (a) and HRRRNEST (b) modelled for the 1200 initialized runs versus WPR observed PBL heights. The control and experimental runs are shown in red and blue, respectively

task of determining the PBL height in its decaying boundary-layer phase, when typically, it is less well defined and difficult to capture accurately. Moreover, during the decaying phase, the modelled deep boundary layer is gradually replaced by a newly forming shallow stable boundary layer. However, because all observed PBL height estimations were visually inspected, instilling a high level of confidence in their values, the larger differences also reflect the sensitivity of the decaying boundary layer to small changes in surface heat flux and lapse rate that are difficult to model precisely. An additional model variable, independent of the PBL parametrization, is the subsidence rate, which can affect boundary-layer growth during the morning, but has a significant effect on boundary-layer decay during the afternoon. Despite the larger scatter for the decaying boundary layer, the models have little bias, accurately representing on average the late afternoon collapse of the boundary layer.

#### 4.2.1 Clear-Sky Days Versus Cloudy Days

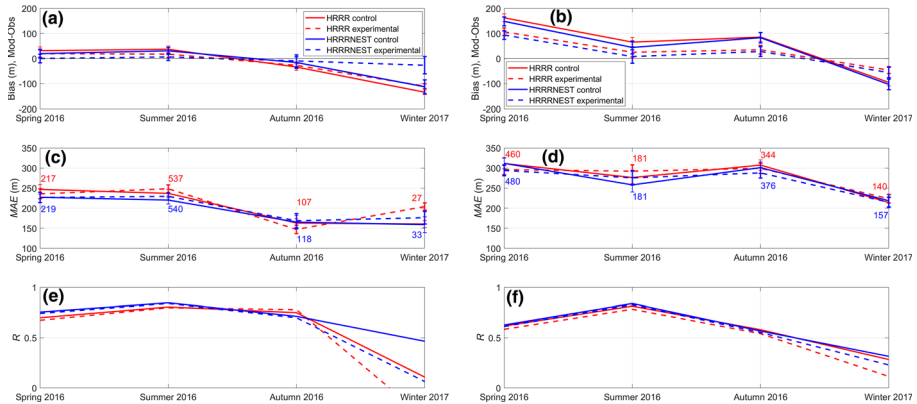
Cloud-topped convective boundary layers are more difficult to observe and to understand than cloud-free boundary layers. Grimsdell and Angevine (1998) analyzed the ability of WPRs to measure cloud-free and cloud-topped mixing heights compared with radiosonde observations. While the agreement was better in cloud-free conditions, even in cloud-topped conditions the correlation coefficient is 0.77. They also found that the peak in the radar return tends to be slightly higher (47 m, in their study) than the calculated cloud base determined from ceilometer data. They attribute this to increased turbulence within the cloud and to the effect of sharp moisture gradients at all edges of the cloud. For clouds higher than those analyzed by Grimsdell and Angevine (1998), above the PBL height, Wang et al. (2016) found that their presence does not influence the boundary-layer structure and therefore the WPR remains a good instrument for the estimation of PBL height.

In regard to the model PBL height estimations, the hybrid-PBL height diagnostic described in Olson et al. (2019a) uses the same diagnostic for both cloudy and clear conditions, but since the liquid and ice virtual potential temperature ( $\theta_{v|l}$ ) is less than  $\theta_v$  in the presence of liquid water or ice, the PBL height is often diagnosed higher than in clear conditions (all other things being equal).



**Table 3** Number of clear-sky days out of the total number of days for each of the four reforecast periods

Site name	Spring 2016	Summer 2016	Autumn 2016	Winter 2017
Condon	9/44	29/45	8/45	4/45
Wasco	12/44	32/45	6/45	5/45

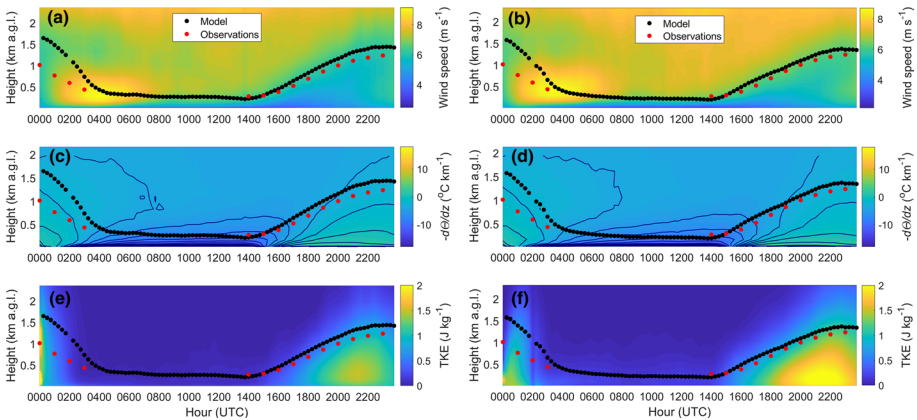


**Fig. 9** Statistics of the PBL height estimation (bias in panels **a** and **b**, MAE in panels **c** and **d**, and  $R$  in panels **e** and **f**) for the four reforecast periods for the clear-sky days (left) and for the cloudy days (right) at the Condon and Wasco sites having collocated surface radiation observations. The HRRR model is in red and the HRRRNEST model in blue (control runs in solid, and experimental runs in dashed lines)

Using the information from the surface radiation stations that were deployed at the Condon and Wasco sites, we repeated the analysis presented in Fig. 7, separating the clear-sky days from cloudy days. The fractional sky cover is a derived variable, calculated from broadband shortwave radiometer measurements using the algorithm developed by Long and Ackerman (2000) and Long et al. (2006). This algorithm uses surface measurements of downwelling total and diffuse shortwave irradiance for estimating fractional sky cover for an effective 160° field-of-view. The clear-sky days are here defined as those with a fractional sky cover over the entire daylight hours being less than 10%, regardless if these clouds are coupled to the boundary layer or not. All the other days are considered cloudy days. The number of clear-sky days identified at the two sites is summarized in Table 3 for the four reforecast periods. Due to the limited number of clear-sky days points going into the statistics of the winter 2017 reforecast period (27 for the HRRR model, and 33 for the HRRRNEST model), the statistics for this period during clear-sky days is less significant.

Results for the clear and cloudy-day analysis are presented in Fig. 9 (left panels are for the clear-sky days, right panels for the cloudy days). The number of points available in each season is presented in Fig. 9c, d for each of the four reforecast periods, with red numbers for the HRRR model and blue numbers for the HRRRNEST model (averaged for each model for the 0000 and 1200 initialization times).

With the exception of winter 2017 when fewer points are available for the analysis, the biases in the control version of the models at the two sites are close to zero in cloud-free conditions (Fig. 9a), and significantly positive in cloudy conditions (Fig. 9b). In comparison, the experimental versions of the models have near zero biases in spring, summer, and autumn 2016 for cloud-free conditions, and in summer and autumn 2016 in cloudy conditions. For the MAE (Fig. 9c, d) the experimental versions of the models perform similarly



**Fig. 10** Five-site average time–height cross-sections daily composite of model (HRRR control on the left, and HRRR experimental on the right, 0000 initialized runs, for the spring 2016 reforecast period) wind speeds (a, b in  $\text{m s}^{-1}$ ), lapse rate of potential temperature with contour lines (c, d in  $^{\circ}\text{C km}^{-1}$ ), TKE (e and f, in  $\text{J kg}^{-1}$ ). Superimposed are the PBL height estimations from the model and from the observations (in black and red dots respectively)

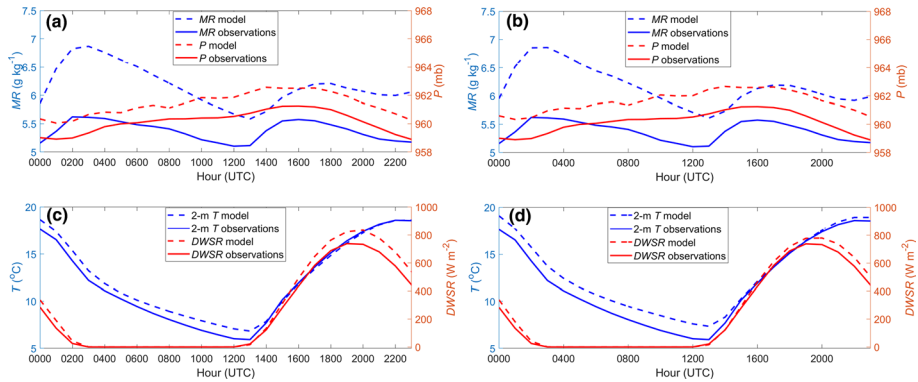
to the control versions (as expected from the analysis presented in the middle panel of Fig. 7), with generally smaller MAE values during clear-sky days (Fig. 9c) compared with cloudy days (Fig. 9d). Finally, values for the correlation coefficient are similar in Fig. 9e, f for both clear-sky days and cloudy days and for all model runs (as expected from the analysis presented in the lower panel of Fig. 7). When the rates of growth and decay were examined for clear-sky days and cloudy days separately (similar to the scatterplots presented in Fig. 8, but not shown) a slightly higher correlation was found for clear-sky days compared with cloudy days (for the HRRR runs  $R \approx 0.7$  for clear-sky days versus  $R \approx 0.6$  for cloudy days; while for the HRRRNEST runs  $R \approx 0.65$  for clear-sky days versus  $R \approx 0.57$  for cloudy days), with the differences between the correlation coefficients being statistically significant ( $p = 0.030$ ) due to the large sample sizes (approximately 500).

While the models clearly simulate the PBL height somewhat less accurately in cloudy conditions, the difference from the errors found on clear-sky days is not dramatic. Some additional improvements in the models are likely needed in cloudy conditions (coupled or not to the PBL), to improve the PBL-height forecast skill through the positive feedback process taking place in the boundary layer described previously.

#### 4.2.2 Impact of Surface Meteorological Variables

Other studies are in progress using observations and model runs from the WFIP2 dataset analyzing in detail cloud cover, the surface energy balance, and surface fluxes. Nevertheless, since pressure measurements from microbarographs were available at four of the WPR sites (Boardman, Condon, Troutdale, and Wasco), and measurements of 2-m temperature, surface irradiance, mixing ratio, and pressure from meteorological stations were available at the four sites with microbarographs plus at the Prineville site, we include a basic evaluation of them.

Figure 10 compares the composite wind speed, lapse rate of potential temperature, and TKE from the HRRR control (left) and HRRR experimental (right) 0000 runs, for all the



**Fig. 11** Five-site average time–height daily composite of model (dashed lines, HRRR control on the left and HRRR experimental on the right, 0000 initialized runs, for the spring 2016 reforecast period) and surface meteorological observation (solid lines) values of mixing ratio (**a, b**  $MR$  in blue, in  $\text{g kg}^{-1}$ ), 2-m pressure (**a, b**  $P$  in red, in mb), 2-m temperature (**c, d** 2-m  $T$  in blue, in  $^{\circ}\text{C}$ ), and downward solar irradiance (**c, d**  $DSWR$  in red, in  $\text{W m}^{-2}$ )

days of the spring 2016 reforecast period (as an example), averaged at all five sites. In these panels PBL height estimations from the model (in black) and observations (in red) are superimposed, showing better agreement for forecast horizons 20–23 (2000–2300 UTC) between the observations and the HRRR experimental 0000 initialized run, compared with the HRRR control 0000 initialized run.

The differences between the control and experimental model runs are easily observed by comparing the left and right panels. The first difference that we note is between Fig. 10a and b, where the low-level jet visible between 0100 and 0700 UTC, and centred around 500 m, is stronger in the experimental. This is found in summer as well (not shown). In Fig. 10c, d, we notice that the convective mixing of the boundary layer during daytime is more limited in height in the experimental version of the model (as denoted by the contour lines plotted for the same values of lapse rate in the control and experimental versions of the model), with the experimental version of the model PBL height estimations in closer agreement with the observations. The differences between the two versions of the model are particularly well pronounced in terms of TKE, with higher values for the HRRR experimental run in the convective boundary layer. The tendency of increased TKE in the convective boundary layer for the experimental (Fig. 10f) compared with the control runs (Fig. 10e) is found for all seasons and for both the HRRR and HRRRNEST runs (not shown). The increase in TKE is due to the addition of the TKE source term in the wind-farm parametrization and the mixing-length revision.

Figure 11 shows the diurnal variation of 2-m water vapour mixing ratio (a), 2-m pressure (b), 2-m temperature (c), and the downward solar irradiance (d) averaged over all five sites, for the observations and the model runs. Figure 11a, b shows that both the mixing ratio and pressure are overestimated in both the control and experimental versions of the model, with no visible improvement in the experimental version of the model, at least for the spring 2016 reforecast period presented here (a more comprehensive analysis on all the reforecast periods is presented below). The offset in pressure in Fig. 11a, b may be due to differences in terrain elevation between the models and observations, however, the models do accurately reproduce the diurnal variation in pressure. Figure 11c and d shows that both models tend to overestimate the observed downward solar irradiance, but maintain

very small biases in 2-m temperature at long forecast horizons (which in this case, since we show the 0000 initialization runs, match the clock time). Improvements in downward solar irradiance should result in a reducing of the daytime temperatures, but this is offset by warming induced by increases in entrainment, which can come from more resolved-scale mixing (larger vertical velocities, not shown) and/or parametrized mixing (increased TKE). The net effect appears to improve the PBL height estimations in the experimental runs compared with the control.

The statistics MAE and bias, and improvements in the MAE, for 2-m temperature, solar irradiance, mixing ratio and pressure over the four reforecast periods at the five above-mentioned sites, are presented in Tables 4 and 5. Results of the 0000 and 1200 model runs models are averaged together, and also site-averaged, over the entire daily cycle. From the values presented in Table 4 it is evident that the value of MAE for the 2-m temperature is slightly worse in spring 2016 and unchanged in autumn and summer 2016, but it is dramatically improved for winter 2017 (for the experimental versus control runs of both HRRR and HRRRNEST models). For solar irradiance, improvements in the value of MAE are found for all reforecast periods, with improvements of 13.7% for the HRRR experimental versus HRRR control runs, and 13.1% for the HRRRNEST experimental versus HRRRNEST control runs, when averaged over the four reforecast periods. Improvements in the value of MAE for the water vapour mixing ratio, while showing averaged positive values, are only found in autumn. Similarly, differences in model 2-m pressure MAE are zero for all seasons except spring for the HRRR model, and winter for the HRRRNEST model, where they are negative.

From Table 5, the 2-m temperature bias is small (just a few tenths of a °C) for all the seasons, and the experimental version of the models does not improve it, which can also be said regarding the biases of mixing ratio and 2-m pressure. However, improvements are large in the bias of downward solar irradiance for all reforecast periods, which is on average reduced by half in the experimental versus the control version of both the HRRR and HRRRNEST models.

This analysis of the surface measurements shows that the main improvements in PBL height are related to improvements in downward solar irradiance, but not so much in the mixing ratio, pressure, and temperature. This is also confirmed in the analysis of the RASS data presented in Olson et al. (2019a), who showed that improved temperatures were mostly found in winter, but not in the other seasons. As mentioned in Sect. 2.2, one of the modifications included in the experimental version of the models involves SGS coupling of clouds and radiation, which we believe is the main reason for the improvement found in the solar irradiance. Additional studies using WFIP2 observations and model runs are underway to analyze clouds and radiation in more detail. This analysis was directed to understand how changes in the model parametrizations impact the various interconnected meteorological variables, and to make sure that the improvements in one variable would not deteriorate another variable, but rather add a positive feedback in the forcing mechanisms of the boundary layer.

While performing the analysis presented in Tables 4 and 5, the surface variables were also analyzed at specific locations, and while the behaviours are mostly consistent at the various locations for the 2-m temperature, downward solar irradiance, and mixing ratio, it was found (not shown) that the pressure of the HRRR model yields a positive bias on the east side of the Cascades (Boardman, Condon, Prineville, and Wasco sites), and a negative bias on the site located to the west of the Cascades (Troutdale).

Differences in site elevations between models and observations can contribute to the pressure bias at the single sites, as for example, the HRRR model, with its 3-km horizontal

**Table 4** Mean absolute errors (MAE) for the different model runs (0000 and 1200 initialized runs averaged together) and their relative improvement (experimental versus control runs) of surface temperature, solar radiation, pressure, and mixing ratio for the four reforecast periods and averaged over them all

Model MAE	Variable	Spring 2016	Summer 2017	Autumn 2016	Winter 2017	All averaged	
HRRR	2-m T (°C)	Control	1.9	1.7	1.6	2.8	2.0
		Experimental	2.0	1.7	1.6	2.3	1.9
	Solar radiation (Wm <sup>-2</sup> )	Improvements	-5.2%	0.0%	0.0%	<b>17.9%</b>	<b>3.2%</b>
		Control	45	36	31	27	35
		Experimental	38	32	28	22	30
		Improvements	<b>15.6%</b>	<b>11.1%</b>	<b>9.7%</b>	<b>18.5%</b>	<b>13.7%</b>
	Pressure (mb)	Control	1.9	2.1	1.6	2.0	1.9
		Experimental	2.1	2.1	1.6	2.0	1.9
	Mixing ratio (g kg <sup>-1</sup> )	Improvements	-10.0%	0.0%	0.0%	0.0%	-2.5%
		Control	1	0.8	0.8	0.5	0.8
Experimental		1	0.8	0.7	0.5	0.7	
Improvements		0.0%	0.0%	<b>12.5%</b>	0.0%	<b>3.1%</b>	
HRRRNEST	2-m T (°C)	Control	2.0	1.7	1.6	2.6	2.0
		Experimental	2.1	1.7	1.6	2.2	1.9
	Solar radiation (Wm <sup>-2</sup> )	Improvements	-5.0%	0.0%	0.0%	<b>15.4%</b>	<b>2.6%</b>
		Control	46	36	34	29	36
		Experimental	39	32	31	24	31
		Improvements	<b>15.2%</b>	<b>11.1%</b>	<b>8.8%</b>	<b>17.2%</b>	<b>13.1%</b>
Pressure (mb)	Control	0.7	0.8	1.0	1.0	0.9	
	Experimental	0.7	0.8	1.0	1.1	0.9	
Mixing ratio (g kg <sup>-1</sup> )	Improvements	0.0%	0.0%	0.0%	-10.0%	-2.5%	
	Control	1	0.9	0.8	0.5	0.8	
	Experimental	1	0.9	0.7	0.5	0.8	
	Improvements	0.0%	0.0%	<b>12.5%</b>	0.0%	<b>3.1%</b>	

Improvements in the experimental versus control versions are in bold. Lower section: Bias of surface temperature, solar radiation, pressure, and mixing ratio for the different model runs (0000 and 1200 initialized runs averaged together) for the four reforecast periods and averaged over them all. Improvements in the experimental versus control versions are in bold

**Table 5** Bias of surface temperature, solar radiation, pressure, and mixing ratio for the different model runs (0000 and 1200 initialized runs averaged together) for the four reforecast periods and averaged over them all

(Model-Obs) Bias	Variable	Spring 2016	Summer 2017	Autumn 2016	Winter 2017	All averaged	
HRRR	2-m $T$ ( $^{\circ}\text{C}$ )	0.2	0.3	0.1	-0.5	0.0	
		Experimental	0.4	0.2	-0.1	0.2	
	solar radiation ( $\text{Wm}^{-2}$ )	34	29	15	18	24	
		Experimental	<b>13</b>	<b>21</b>	<b>5</b>	<b>11</b>	<b>12</b>
	pressure (mb)	1.3	1.5	1.2	1.5	1.4	
		Experimental	<b>1.5</b>	1.5	1.2	2.1	1.6
	mixing ratio ( $\text{g kg}^{-1}$ )	1.7	0.2	-0.3	0.0	0.4	
		Experimental	<b>0.8</b>	0.2	-0.2	0.0	0.4
	2-m $T$ ( $^{\circ}\text{C}$ )	0.2	0.4	0.0	-0.8	0.0	
		Experimental	0.3	0.6	0.2	-0.2	0.2
HRRRNEST	solar radiation ( $\text{Wm}^{-2}$ )	36	32	18	19	26	
		Experimental	<b>17</b>	<b>23</b>	<b>7</b>	<b>11</b>	<b>14</b>
	pressure (mb)	0.6	0.7	0.5	0.9	0.7	
		Experimental	0.6	<b>0.6</b>	<b>0.3</b>	1.1	<b>0.6</b>
	mixing ratio ( $\text{g kg}^{-1}$ )	0.7	0.2	-0.3	-0.1	0.1	
		Experimental	0.8	0.2	-0.2	<b>0.0</b>	0.2

Improvements in the experimental versus control versions are in bold

grid spacing, resolves the terrain less accurately than the HRRRNEST model with its 750-m horizontal grid spacing. On the other hand, the lack of accuracy in the HRRR model at forecasting the gradient in pressure between the west and east side of the Cascades can be important in other applications, as for example when validating the model at forecasting the intensity of gap flows usually present in this area of complex terrain through the gorge in the Cascade mountains (Banta et al. 2019), with consequences also in the forecast of wind up-ramp and down-ramp events, which are important in this area with large wind-energy production (Djalalova et al. 2020). Future studies are being conducted to address this topic.

## 5 Conclusions

Data collected during the second Wind Forecast Improvement Project (WFIP2, October 2015—March 2017, in the Pacific Northwest U.S.A.) are used to evaluate the impact of improvements and additions made to the RAP, HRRR, and HRRRNEST model parametrizations on simulating the PBL height, as well as surface meteorological forcing variables that influence the evolution of the PBL. Of interest were the RAP (13-km grid spacing), the HRRR (3-km grid spacing), and the HRRRNEST (750-m grid spacing) control and experimental models, with the experimental configuration including all the improvement to the parametrizations. We find that:

- Higher-resolution model versions outperform the lower-resolution versions in estimating the convective PBL height (with the HRRRNEST, with its smaller grid spacing, providing a smaller bias and value of the MAE compared to the HRRR);
- Experimental versions of the models with improved physical parametrizations perform better than control versions in estimating the convective PBL height (with improvements in the value of the MAE between 5–15% for the RAP model, depending on the forecast horizon-time after model initialization, and between a few percent to 8% in the HRRR and HRRRNEST models, depending on the season);
- All model runs estimate the rate of growth of PBL height more accurately than the rate of decay;
- During cloud-free days (days with fractional sky cover over the entire daylight hours being less than 10%) the NWP models have greater accuracy at simulating the PBL height than on cloudy days (days with fractional sky cover over the entire daylight hours being more than 10%);
- Downward solar irradiance MAE and bias are found to be improved in the experimental runs compared with the control for both the HRRR and HRRRNEST models. To a lesser extent, this is also true for 2-m temperature. The impact on the surface mixing ratio and pressure of the experimental version of the models is negligible.

Subsequent model development has been initiated to further alleviate the systematic biases noted above, including additional refinements to the SGS clouds and their interaction with the radiation to reduce the lingering solar radiation biases, and adding heating from compensational subsidence to the plume mixing in the mass-flux scheme to help reduce the high surface pressure biases in the Columbia Basin. The degree of success of these model development tasks will be investigated in the future.



While the main goal of this analysis was to evaluate model improvements at forecasting PBL heights, the rich dataset collected during the WFIP2 campaign allowed for a deeper investigation in the mechanisms controlling the PBL growth, providing insights into how the model can be further improved. Other studies are in preparation, going further into the details of cloud cover, surface radiation budget, and surface fluxes on the WFIP2 dataset.

**Acknowledgements** We thank all the people involved in WFIP2 for site selection, leases, instrument deployment and maintenance, data collection, and data quality control. Funding for this work was provided by the Department of Energy, Office of Energy Efficiency and Renewable Energy, Wind Energy Technologies Office, and by the NOAA Atmospheric Science for Renewable Energy (ASRE) program. We are very grateful for the constructive comments and suggestions provided by the three anonymous reviewers, which we believe have greatly improved the clarity of the manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Angevine WM, White AB, Avery SK (1994) Boundary layer depth and entrainment zone characterization with a boundary-layer profiler. *Boundary-Layer Meteorol* 68:375–385
- Banks RF, Tiana-Alsina J, Rocabenbosch F et al (2015) Performance evaluation of the boundary-layer height from lidar and the weather research and forecasting model at an urban coastal site in the North-East Iberian Peninsula. *Boundary-Layer Meteorol* 157:265–292. <https://doi.org/10.1007/s10546-015-0056-2>
- Banta RM, Pichugina YL, Brewer WA, Choukulkar A, Lantz KO, Olson JB, Kenyon J, Fernando HJS, Krishnamurthy R, Stoelinga MJ, Sharp J, Darby LS, Turner DD, Baidar S, Sandberg SP (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. <https://doi.org/10.1175/MWR-D-19-0188.1>
- Benjamin SG, Weygandt SS, Brown JM, Hu M, Alexander CR, Smirnova TG, Olson JB, James EP, Dorell DC, Grell GA, Lin H, Peckham SE, Smith TL, Moninger WR, Kenyon JS, Manikin GS (2016) A North American hourly assimilation and model forecast cycle: the Rapid Refresh. *Mon Wea Rev* 144:1669–1694. <https://doi.org/10.1175/MWR-D-15-0242.1>
- Bianco L, Wilczak JM (2002) Convective boundary layer depth: Improved measurement by Doppler radar wind profiler using fuzzy logic methods. *J Atmos Ocean Technol* 19:1745–1758. [https://doi.org/10.1175/1520-0426\(2002\)019%3c1745:CBLDIM%3e2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019%3c1745:CBLDIM%3e2.0.CO;2)
- Bianco L, Wilczak JM, White AB (2008) Convective boundary layer depth estimation from wind profilers: statistical comparison between an automated algorithm and expert estimations. *J Atmos Ocean Technol* 25:1397–1413. <https://doi.org/10.1175/2008JTECHA981.1>
- Bianco L, Djalalova IV, Wilczak JM, Olson JB, Kenyon JS, Choukulkar A, Berg L, Fernando HJS, Gmit EP, Krishnamurthy R, Lundquist JK, Muradyan P, Pekour M, Pichugina Y, Stoelinga MT, Turner DD (2019) Impact of model improvements on 80-m wind speeds during the second Wind Forecast Improvement Project (WFIP2). *Geophys Model Dev* 12:4803–4821. <https://doi.org/10.5194/gmd-12-4803-2019>
- Cohn SA, Angevine WM (2000) Boundary layer height and entrainment zone thickness measured by lidars and wind-profiling radars. *J Appl Meteor* 39:1233–1247. [https://doi.org/10.1175/1520-0450\(2000\)039%3c1233:BLHAEZ%3e2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039%3c1233:BLHAEZ%3e2.0.CO;2)
- Collaud Coen M, Praz C, Haeefe A, Ruffieux D, Kaufmann P, Calpini B (2014) Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing

- measurements as well as by the COSMO-2 model. *Atmos Chem Phys* 14:13205–13221. <https://doi.org/10.5194/acp-14-13205-2014>
- Coulter RL, Holdridge DH (1998) A procedure for the automatic estimation of mixed layer height. In: Proceedings of eighth atmospheric radiation measurement (ARM) program science team meeting, Tucson, AZ, Department of Energy Office of Energy Research, pp 177–180
- Cuxart J, Holtslag AAM, Beare RJ, Bazile E, Beljaars A, Cheng A, Conangla L, Ek MB, Freedman F, Hamdi R, Kerstein A, Kitagawa H, Lenderink G, Lewellen D, Mailhot J, Mauritsen T, Perov V, Schayes G, Steeneveld GJ, Svensson G, Taylor P, Weng W, Wunsch S, Xu KM (2006) Single-column model intercomparison for a stably stratified atmospheric boundary layer. *Boundary-Layer Meteorol* 118:273–303
- Djalalova IV, Bianco L, Akish E, Wilczak JM, Olson JB, Kenyon JS, Berg LK, Choukulkar A, Coulter R, Eckman R, Fernando HJS, Grit E, Krishnamurthy R, Lundquist JK, Muradyan P, Pekour M, Stoelinga M (2020) Ramp events validation during the second Wind Forecast Improvement Project (WFIP2) using the Ramp Tool and Metric (RT&M). *Weather Forecast* 35(6):2407–2421. <https://doi.org/10.1175/WAF-D-20-0072.1>
- Draxl C, Worsnop RP, Xia G, Pichugina Y, Chand D, Lundquist JK, Sharp J, Wedam G, Wilczak JM, Berg LK (2021) Mountain waves impact wind power generation. *Wind Energy Sci* 6:45–60. <https://doi.org/10.5194/wes-6-45-2021>
- Durrán DR (1990) Mountain waves and downslope winds. In: Blumen W (ed) *Atmospheric processes over complex terrain: Meteorological monographs*, vol 23. American Meteorological Society, Boston. [https://doi.org/10.1007/978-1-935704-25-6\\_4](https://doi.org/10.1007/978-1-935704-25-6_4)
- Durrán DR (2003) Lee waves and mountain waves, encyclopedia of atmospheric sciences, Holton JR, Pyle J, Curry JA. Elsevier, Amsterdam, pp 1161–1169. <https://doi.org/10.1016/B0-12-227090-8/00202-5>
- Ecklund WL, Carter DA, Balsley BBA (1988) UHF wind profiler for the boundary layer: brief description and initial results. *J Atmos Ocean Technol* 5:432–441
- Fitch AC, Lundquist JK, Olson JB (2013) Mesoscale influences of wind farms throughout a diurnal cycle. *Mon Wea Rev* 141:2173–2198. <https://doi.org/10.1175/MWR-D-12-00185.1>
- Grachev AA, Fairall CW, Blomquist BW, Fernando HJS, Leo LS, Otárola-Bustos SF, Wilczak JM, McCaffrey KL (2020) On the surface energy balance closure at different temporal scales. *Agric For Meteorol* 281:107823. <https://doi.org/10.1016/j.agrformet.2019.107823>
- Grimsdell AW, Angevine WM (1998) Convective boundary layer height measured with wind profilers and compared to cloud base. *J Atmos Ocean Technol* 15:1331–1338. [https://doi.org/10.1175/1520-0426\(1998\)015%3c1331:CBLHMW%3e2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015%3c1331:CBLHMW%3e2.0.CO;2)
- Kosović B, Curry JA (2000) A large eddy simulation study of a quasi-steady, stably stratified atmospheric boundary layer. *J Atmos Sci* 57:1052–1068. [https://doi.org/10.1175/1520-0469\(2000\)057<1052:ALESSO.2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<1052:ALESSO.2.0.CO;2)
- Lee TR, De Wekker SFJ (2016) Estimating daytime planetary boundary layer heights over a valley from rawinsonde observations at a nearby airport: an application to the Page Valley in Virginia, United States. *J Appl Meteorol Climatol* 55:791–809. <https://doi.org/10.1175/JAMC-D-15-0300.1>
- LeMone MA, Tewari M, Chen F, Dudhia J (2013) Objectively determined fair-weather CBL depths in the ARW-WRF model and their comparison to CASES-97 observations. *Mon Wea Rev* 141:30–54. <https://doi.org/10.1175/MWR-D-12-00106.1>
- LeMone MA, Tewari M, Chen F, Dudhia J (2014) Objectively determined fair weather NBL features in ARW-WRF and their comparison to CASES-97 observations. *Mon Wea Rev* 142:2709–2732. <https://doi.org/10.1175/MWR-D-13-00358.1>
- Long CN, Ackerman TP (2000) Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J Geophys Res* 105(D12):15609–15626. <https://doi.org/10.1029/2000JD900077>
- Long CN, Ackerman TP, Gaustad KL, Cole JNS (2006) Estimation of fractional sky cover from broadband short-wave radiometer measurements. *J Geophys Res* 111:D11204. <https://doi.org/10.1029/2005JD006475>
- McCaffrey K, Bianco L, Wilczak JM (2017) Improved observations of turbulence dissipation rates from wind profiling radars. *Atmos Meas Technol* 10:2595–2611. <https://doi.org/10.5194/amt-10-2595-2017>
- McCaffrey K, Wilczak JM, Bianco L, Grit E, Sharp J, Banta R, Friedrich K, Fernando HJS, Krishnamurthy R, Leo L, Muradyan P (2019) Identification and characterization of persistent cold pool events from temperature and wind profilers in the Columbia River Basin. *J Appl Meteorol Climatol* 58:2533–2551. <https://doi.org/10.1175/JAMC-D-19-0046.1>
- Neiman PJ, Gottas DJ, White AB, Schneider WR, Bright DR (2018) A real-time online data product that automatically detects easterly gap-flow events and precipitation type in the Columbia River Gorge. *J Atmos Ocean Technol* 35:2037–2052. <https://doi.org/10.1175/JTECH-D-18-0088.1>

- Neiman PJ, Gottas DJ, White AB (2019) A Two-Cool-Season Wind Profiler-based analysis of westward-directed gap flow through the Columbia River Gorge. *Mon Wea Rev* 147:4653–4680. <https://doi.org/10.1175/MWR-D-19-0026.1>
- Nielsen-Gammon JW, Coauthors, (2008) Multisensor estimation of mixing heights over a coastal city. *J Appl Meteorol Climatol* 47:27–43. <https://doi.org/10.1175/2007JAMC1503.1>
- Nishiyama RT, Bedard AJ Jr (1991) A “Quad-Disc” static pressure probe for measurement in adverse atmospheres: with a comparative review of static pressure probe designs. *Rev Sci Instrum* 62:2193–2204
- Olson JB, Kenyon JS, Djalalova I, Bianco L, Turner DD, Pichugina Y, Choukulkar A, Toy MD, Brown JM, Angevine W, Akish E, Bao J-W, Jimenez P, Kosović B, Lundquist KA, Draxl C, Lundquist JK, McCaa J, McCaffrey K, Lantz K, Long C, Wilczak J, Banta R, Marquis M, Redfern S, Berg LK, Shaw W, Cline J (2019a) Improving wind energy forecasting through numerical weather prediction model development. *Bull Am Meteorol Soc* 100:2201–2220. <https://doi.org/10.1175/BAMS-D-18-0040.1>
- Olson JB, Kenyon JS, Angevine WM, Brown JM, Pagowski M, Sušelj K (2019b) A description of the MYNN-EDMF scheme and coupling to other components in WRF-ARW. NOAA Tech Mem OAR GSD 61:37. <https://doi.org/10.25923/n9wm-be49>
- Pichugina YL, Banta RM, Bonin T, Brewer WA, Choukulkar A, McCarty BJ, Baidar S, Draxl C, Fernando HJS, Kenyon J, Krishnamurthy R, Marquis M, Olson J, Sharp J, Stoelinga M (2019) Spatial variability of winds and HRRR–NCEP model error statistics at three Doppler-Lidar sites in the wind-energy generation region of the Columbia River Basin. *J Appl Meteorol Climatol* 58:1633–1656. <https://doi.org/10.1175/JAMC-D-18-0244.1>
- Pichugina Y, Banta R, Brewer A, Baidar S, Berg L, Bianco L, Draxl C, Fernando HJS, Kenyon J, Lundquist JK, McCarty B, Olson J, Rai R, Roberts B, Sharp J, Shaw W, Stoelinga M, Turner DD, Wharton S, Wilczak J, Worsnop R (2020) Evaluating the WFIP2 updates to the HRRR model using scanning Doppler lidar measurements in the complex terrain of the Columbia River Basin. *J Renew Sust Energy* 12:043301. <https://doi.org/10.1063/5.0009138>
- Sharp J, Mass C (2002) Columbia Gorge gap flow: insights from observational analysis and ultra-high-resolution simulation. *Bull Am Meteorol Soc* 83:1757–1762. <https://doi.org/10.1175/1520-0477-83.12.1745>
- Sharp J, Mass C (2004) Columbia Gorge gap winds: their climatological influence and synoptic evolution. *Weather Forecast* 19:970–992. <https://doi.org/10.1175/826.1>
- Shaw W, Berg L, Cline J, Draxl C, Djalalova I, Grimit E, Lundquist JK, Marquis M, McCaa J, Olson J, Sivaraman C, Sharp J, Wilczak JM (2019) The second Wind Forecast Improvement Project (WFIP2): general overview. *Bull Am Meteorol Soc* 100(9):1687–1699. <https://doi.org/10.1175/BAMS-D-18-0036.1>
- Strauch RG, Merritt DA, Moran KP, Earnshaw KB, Van De Kamp D (1984) The Colorado wind-profiling network. *J Atmos Ocean Technol* 1:37–49
- Wang C, Shi H, Jin L, Chen H, Wen H (2016) Measuring boundary-layer height under clear and cloudy conditions using three instruments. *Particuology* 28:15–22. <https://doi.org/10.1016/j.partic.2015.04.004>
- White AB (1993) Mixing depth detection using 915-MHz radar reflectivity data. Preprints, Eighth Symposium on observations and instrumentation, Anaheim, CA. American Meteorological Society, pp 248–250
- White AB, Gottas DJ, Strem ET, Ralph FM, Neiman PJ (2002) An automated brightband height detection algorithm for use with Doppler radar spectral moment. *J Atmos Ocean Technol* 19:687–697
- Whiteman CD, Zhong S, Shaw WJ, Hubbe JM, Bian X, Mittelstadt J (2001) Cold pools in the Columbia Basin. *Weather Forecast* 16:432–447. [https://doi.org/10.1175/1520-0434\(2001\)016<0432:CPITCB.2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0432:CPITCB.2.0.CO;2)
- Wilczak JM, Stoelinga M, Berg L, Sharp J, Draxl C, McCaffrey K, Banta RM, Bianco L, Djalalova I, Lundquist JK, Muradyan P, Choukulkar A, Leo L, Bonin T, Pichugina Y, Eckman R, Long C, Lantz K, Worsnop R, Bickford J, Bodini N, Chand D, Clifton A, Cline J, Cook D, Fernando HJS, Friedrich K, Krishnamurthy R, Marquis M, McCaa J, Olson J, Otarola-Bustos S, Scott G, Shaw WJ, Wharton S, White AB (2019a) The second Wind Forecast Improvement Project (WFIP2): observational field campaign. *Bull Am Meteorol Soc* 100(9):1701–1723. <https://doi.org/10.1175/BAMS-D-18-0035.1>
- Wilczak JM, Olson J, Djalalova I, Bianco L, Berg L, Shaw W, Coulter R, Eckman RM, Freedman J, Finley C, Cline J (2019b) Data assimilation impact of tall towers, wind turbine nacelle anemometers, sodars and wind profiling radars on wind velocity and power forecasts during the first Wind Forecast Improvement Project (WFIP). *Wind Energy*. <https://doi.org/10.1002/we.2332>
- Wiser RH, Bolinger M (2018) wind technologies market report. Technical Report for the Wind Energy Technologies Office within the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy. <https://doi.org/10.2172/1559241>

Zhong S, Whiteman CD, Bian X, Shaw WJ, Hubbe JM (2001) Meteorological processes affecting the evolution of a wintertime cold air pool in the Columbia basin. *Mon Wea Rev* 129(10):2600–2613. [https://doi.org/10.1175/1520-0493\(2001\)129%3c2600:MPATEO%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129%3c2600:MPATEO%3e2.0.CO;2)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Laura Bianco<sup>1,2,6</sup>  · P. Muradyan<sup>3</sup> · I. Djalalova<sup>1,2</sup> · J. M. Wilczak<sup>2</sup> · J. B. Olson<sup>4</sup> · J. S. Kenyon<sup>1,4</sup> · R. Kotamarthi<sup>3</sup> · K. Lantz<sup>1,5</sup> · C. N. Long<sup>1,5</sup> · D. D. Turner<sup>4</sup>

<sup>1</sup> Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

<sup>2</sup> Physical Science Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA

<sup>3</sup> Argonne National Laboratory, Lemont, IL, USA

<sup>4</sup> Global Systems Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA

<sup>5</sup> Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA

<sup>6</sup> NOAA/Earth System Research Laboratory, 325 Broadway, mail stop: PSL3, Boulder, CO 80305, USA