

IMPROVING WIND ENERGY FORECASTING THROUGH NUMERICAL WEATHER PREDICTION MODEL DEVELOPMENT

JOSEPH B. OLSON, JAYMES S. KENYON, IRINA DJALALOVA, LAURA BIANCO, DAVID D. TURNER, YELENA PICHUGINA, ADITYA CHOUKULKAR, MICHAEL D. TOY, JOHN M. BROWN, WAYNE M. ANGEVINE, ELENA AKISH, JIAN-WEN BAO, PEDRO JIMENEZ, BRANKO KOSOVIC, KATHERINE A. LUNDQUIST, CAROLINE DRAXL, JULIE K. LUNDQUIST, JIM MCCAA, KATHERINE MCCAFFREY, KATHY LANTZ, CHUCK LONG, JIM WILCZAK, ROBERT BANTA, MELINDA MARQUIS, STEPHANIE REDFERN, LARRY K. BERG, WILL SHAW, AND JOEL CLINE

Operational numerical weather prediction models are being developed to improve wind energy forecasts by leveraging a multiscale dataset from the Second Wind Forecast Improvement Project field campaign in the U.S. Northwest.

Numerical weather prediction (NWP) models provide the foundation for forecasting a wide range of meteorological phenomena, from tropical cyclones to gentle breezes. The development of many operational NWP models has traditionally been motivated, in large part, by imperatives to improve forecasts of high-impact weather events and routine, near-surface “sensible” weather, while comparatively little effort has been devoted to improving wind forecasts at heights of 50–200 m AGL, where wind turbines harvest wind energy. Currently wind energy constitutes 6% and 4% of the electricity production of the United States and the world, respectively, and the rate of growth since 2001 is 17% and 21%, respectively. Wind energy is expected to become a large component of the electrical-generation portfolio of United States and the world as a whole (AWEA Data Services 2017; Global Wind Energy Council 2018). In particular, the 2015 Wind Vision of the Department of Energy (DOE) study has

mapped out a target scenario for wind energy to provide 35% of the United States’ electricity demands by 2050 (Department of Energy 2015). However, winds are an inherently variable source of electric generation, and for commonly used wind turbines, a 1 m s^{-1} change in rotor-layer wind speeds from 7 to 8 m s^{-1} can result in energy output changes up to 50%, owing to the cubic relationship between wind speed and power (International Electrotechnical Commission 2007). Furthermore, these changes in wind speeds over short time intervals ($\Delta t < 4 \text{ h}$), known as *wind ramps*, make forecasting of available wind energy resources very challenging. Due to these sensitivities, the efficiency of wind energy operations and the integration of wind energy into electric grids and electricity markets are greatly affected by the accuracy of wind forecasts. To this end, the strategic aims of NWP model development must broaden, to include the goal of improved forecasts of rotor-layer winds.

A substantial amount of wind-generation capacity exists in regions of complex terrain. Skillful wind forecasts in complex terrain are challenging, owing largely to the prevalence of terrain-modulated flows, such as mountain wakes, mountain waves, gap flows, valley cold pools, and mountain–valley circulations, all of which can be difficult to simulate due to limitations in NWP models or they may have inherent limits of predictability.

In the U.S. Northwest, the Columbia basin (CB) is an example of a high-wind-resource region within complex terrain. The CB is situated east of the Cascade Range, which not only features towering volcanic summits—some 4,000 m above mean sea level—but also a transecting river cut, known as the Columbia River Gorge (CRG). The CRG provides a near-sea level airflow conduit across the Cascade Range, and is a favored location for intense horizontal pressure gradients to develop, generating strong gap flows. During the warm season, a pronounced, diurnally varying horizontal pressure gradient within the CRG often emerges in response to land–sea temperature contrasts, producing strong westerly gap flows through the CRG that are directed into the CB (Fig. 1a). During the cold season, the horizontal pressure gradient in the CRG evolves largely in response to mobile synoptic-scale features, and both westerly and easterly gap flows through the CRG can develop (Baker et al. 1978; Sharp and Mass 2004). Moreover, in the winter, cold pools can deepen within the CB, shielding the wind farms within the basin from overlying winds. Throughout the year, electrical generation from wind in the CB can reach near 5 gigawatts, but rotor-layer wind forecasts in this region pose a difficult forecasting challenge, with errors often exceeding 2.8 m s^{-1} , which is approximately one standard deviation of the hub-height wind speed errors in modern NWP models (shown later, Fig. 11).

The overarching objective of the Second Wind Forecast Improvement Project (WFIP2) is to develop NWP models in a manner that leads to improved low-level wind forecasts in regions of complex terrain for short-range (i.e., 1–24 h) applications (Shaw et al. 2019). These model improvements are expected to proceed from a better understanding of the physical processes associated with the wind flow in and around wind farms. Due to the complexity of these processes and the feedbacks they may exert, forecast errors in rotor-layer winds may originate from numerous model components or model initial conditions. While model initial-condition improvements were investigated in the first WFIP (Wilczak et al. 2015), the focus of WFIP2 is primarily on developing improved model physical parameterizations and the application of improved model numerics. Forecast improvements are sought for a range of model horizontal grid spacing (Δx), from very high resolution ($\Delta x \leq 1 \text{ km}$) models, coarse resolution ($\Delta x > 10 \text{ km}$) models, and in between, to provide improved operational numerical guidance as well as higher-resolution modeling within the academia and the private sector. Therefore, the WFIP2 model-development effort attempts to develop scale-adaptive physical parameterizations that can represent subgrid-scale processes across all scales.

In support of the WFIP2 model-development effort, the U.S. DOE—in collaboration with NOAA, Vaisala Inc., other private firms, universities, and DOE national laboratories—deployed numerous wind profiling and scanning instruments within or near the CRG and CB (Fig. 1b) as part of an 18-month WFIP2 field study that spanned October 2015–March 2017 (Wilczak et al. 2019). The scientific challenges were outlined at the beginning of the field project and aggressive model-development goals were set within the limited timeline of the project. The following

AFFILIATIONS: OLSON, KENYON, DJALALOVA, BIANCO, PICHUGINA, CHOUKULKAR, TOY, ANGEVINE, McCAFFREY, LANTZ, AND LONG—Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and National Oceanic and Atmospheric Administration/Earth System Research Laboratory, Boulder, Colorado; TURNER, BROWN, BAO, WILCZAK, BANTA, AND MARQUIS—National Oceanic and Atmospheric Administration/Earth System Research Laboratory, Boulder, Colorado; AKISH—Science and Technology Corporation, Boulder, Colorado; JIMENEZ AND KOSOVIC—National Center for Atmospheric Research, Boulder, Colorado; K. A. LUNDQUIST—Lawrence Livermore National Laboratory, Livermore, California; DRAXL—National Renewable Energy Laboratory, Golden, Colorado; J. K. LUNDQUIST AND REDFERN—Department of Atmospheric and Oceanic Sciences, University of Colorado

Boulder, Boulder, and National Renewable Energy Laboratory, Golden, Colorado; McCAA—Vaisala, Inc., Seattle, Washington; BERG AND SHAW—Pacific Northwest National Laboratory, Richland, Washington; CLINE—National Oceanic and Atmospheric Administration/National Weather Service, Washington, D.C.

CORRESPONDING AUTHOR: Joseph B. Olson, joseph.b.olson@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-18-0040.1

In final form 22 May 2019

©2019 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

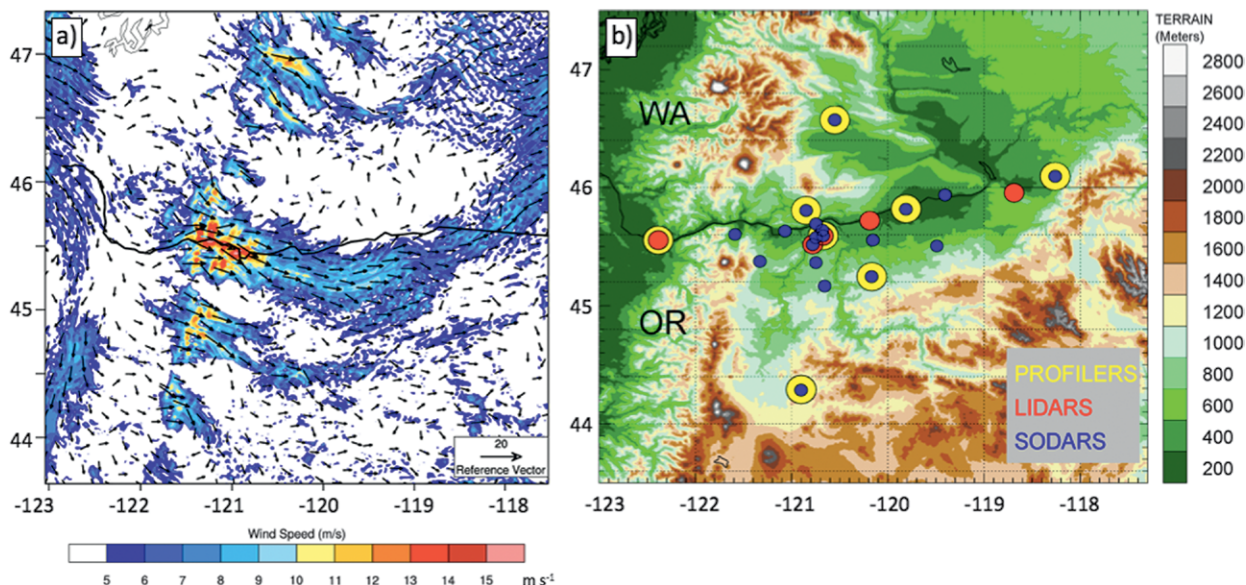


FIG. 1. (a) An example of a gap flow event taken from an 18-h forecast of the 80-m wind speed and vectors from the HRRRNEST for 5 Jul 2016. The 3-km HRRR (parent domain) was initialized at 1200 UTC 5 Jul 2016 and the HRRRNEST was initialized off the 1-h HRRR forecast and run concurrently. All wind vectors $<4 \text{ m s}^{-1}$ were omitted. **(b)** A station map of the WFIP 2 region included all stations with 915 MHz and RASS (yellow), lidars (red), and sodars (blue) over roughly the same regions as in (a).

sections overview these efforts and present results of this intensive model development effort.

SCIENTIFIC CHALLENGES. There are several scientific challenges with NWP model development that are general to any forecast application, but there are additional challenges specific to applications in complex terrain. The primary challenges include

- linking the evolution of terrain-modulated flows to specific physical processes;
- improving the physical parameterizations that represent subgrid-scale physical quantities/processes;
- improving numerical methods for solving continuous equations on discrete grids;
- improving initial conditions;
- improving the accuracy and representativeness of lower-boundary conditions (terrain elevation, soil state, vegetation coverage, albedo, etc.);
- improving the scale-adaptive flexibility of physical parameterizations, thereby allowing a parameterization to run satisfactorily at any grid spacing;
- optimizing a set of physical parameterizations to perform well together as a full physics suite; and
- unambiguously attributing forecast errors to specific model components or initial conditions.

Any of these challenges can hinder the model-development process, so model developers must remain

cognizant of all of them. In a project such as WFIP2, not all of these challenges can be addressed within the scope of the project; instead, the primary emphasis is placed on the improvement of model physical parameterizations by improving the representation of specific processes and the accuracy of numerical methods. In the context of complex terrain, this focus implies a requirement for scale-adaptive physical parameterizations.

Improving model physical parameterization is a broad scientific challenge. One approach is to ensure that all important physical processes are represented in the model. An example of such a process is the representation of subgrid-scale clouds and their interaction with radiation, which is essential for simulating a realistic surface energy balance and properly driving the turbulence parameterization. Other examples include the impacts of horizontal heterogeneity on surface fluxes and PBL mixing, the representation of the effects of subgrid-scale orography (e.g., Beljaars et al. 2004; Steeneveld et al. 2008), and the representation of wind farms (e.g., Fitch et al. 2012, 2013a,b). Some of these processes are often neglected or only crudely included. Without the representation of all relevant processes, the attribution of wind speed forecast errors to specific model components is practically impossible, and existing parameterizations may be inappropriately designed to compensate for nonrepresented processes.

Another approach for improving physical parameterizations is to adapt the parameterizations to perform under a less restrictive set of assumptions. One commonly used simplification is the flat-terrain approximation, which can cause errors in surface momentum and scalar fluxes when slopes become sufficiently steep ($>30^\circ$; Epifanio 2007). In these conditions, it may be justifiable to represent the horizontal momentum stress and heat fluxes within a full 3D surface flux and turbulence parameterization schemes.

Traditionally, most physical parameterization schemes have been designed for forecasting applications at specific grid spacing, causing poor performance when applied in other configurations. With a growing reliance on community modeling codes and parameterization interoperability, parameterization schemes must be designed with scale-adaptive flexibility, allowing the scheme to self-adjust for any choice of model grid spacing. For example, parameterizing all of the turbulent vertical mixing may be an appropriate approximation when Δx is relatively large ($>>1$ km). However, as Δx decreases below the traditional lower limit for mesoscale simulations (~ 1 km), this approximation may not be justified. Moreover, horizontal turbulent mixing is typically calculated in separate horizontal diffusion schemes, with no direct communication with the parameterized vertical mixing, but vertical and horizontal mixing in the convective PBL is often accomplished by the same turbulent eddies and PBL rolls, suggesting that fully 3D turbulence schemes are appropriate for $\Delta x < 1$ km (Boutle et al. 2014).

In addition to improved physical parameterizations, numerical methods within the dynamic core of NWP models can be improved, especially for application to complex terrain. Terrain-following coordinate used in many NWP models work well over smoothly varying terrain, but numerical errors can arise in complex terrain. As the horizontal resolution increases, finescale terrain features are captured, leading to larger terrain slopes and resulting in a skewed computational grid. Large grid skewness can introduce numerical errors and can even lead to numerical noise and computational instabilities. These deficiencies can potentially impact each spatially discretized term of the Navier–Stokes equations, including the horizontal pressure gradient, diffusion, and advection terms. Methods for reducing these errors include alternative formulations of terrain-following coordinates which smooth quickly with altitude (Leuenberger et al. 2010; Schär et al. 2002), more accurate finite difference stencils (Mahrer 1984;

Klemp 2011; Zängl 2012), use of the immersed boundary method (Lundquist et al. 2010, 2012; Ma and Liu 2017), and improvements to grid quality by controlling grid aspect ratio (Daniels et al. 2016). Improved numerical methods can improve the ability to accurately represent the physics of atmospheric processes that drive winds and turbulence at hub height.

In our experience of developing operational models, the approximate time scale of physical parameterization development is two years or greater before a scheme reaches maturity and demonstrates improved skill within a defined physics suite, primarily because of the time-consuming iterative process of diagnosing untoward behavior, modifying the model to reduce errors and performing further evaluation. Because of this, most of the science challenges overviewed above may not be surmountable within the span of a single model development project, but it is our intention to make progress. The following section overviews the model development tasks initiated in WFIP2 toward this end.

MODEL AND COMPONENTS TARGETED FOR DEVELOPMENT. NOAA’s Rapid Refresh (RAP; $\Delta x = 13$ km; Benjamin et al. 2016) and High-Resolution Rapid Refresh (HRRR; $\Delta x = 3$ km) models were selected as the basis for development during WFIP2. The reason for choosing these models was threefold. First, both models utilize a common set of well-tested physical parameterizations, known as the RAP/HRRR physics suite. Second, both models are run operationally by NWS/NCEP, such that improvements made during WFIP2 would be readily transferable into upgraded versions of these models. And third, both models utilize the underlying Advanced Research version of the WRF Model (WRF-ARW; Skamarock et al. 2008), such that WFIP2 model improvements would be transferable to the open-source WRF-ARW repository. The RAP/HRRR physics suite and with the WRF-ARW dynamical core comprise the model framework for the WFIP2 model-development effort. The domains of the RAP and HRRR are shown in Fig. 2 with white and gold boxes, respectively.

To support the goals of WFIP2 using the limited computational resources afforded to the project, a nonoperational version of the HRRR is utilized for WFIP2. This provisional WFIP2 HRRR configuration encompasses a smaller domain than its operational counterpart, as shown by the large green box in Fig. 2. The provisional WFIP2 HRRR is also a “cold start” configuration, where initial conditions are supplied from the RAP without additional data

assimilation or antecedent cycling. Aside from these differences, the provisional WFIP2 HRRR utilizes the same Δx and physics suite as its operational counterpart. Hereafter, the term “HRRR” will refer to the provisional WFIP2 configuration of the HRRR, unless otherwise noted.

Recognizing that very high resolution horizontal grid spacing is required to resolve terrain features that largely govern the formation of wind flows in complex terrain, a nested domain with $\Delta x = 750$ m, hereafter called the HRRRNEST, is run inside the provisional WFIP2 HRRR. This grid spacing is chosen because a fundamental assumption often made in mesoscale PBL parameterizations—namely, that the magnitude of the vertical gradients of basic-state variables far exceeds that of the horizontal gradients, known as the PBL approximation—is regarded as questionable for $\Delta x < 1,000$ m in simple terrain and may even be questionable for $\Delta x < 4,000$ m in complex terrain. Moreover, the 3D local diffusion schemes applied in large-eddy simulations (LES) are generally inappropriate for $\Delta x > 500$ m. This intermediate scale of Δx , where $500 \text{ m} < \Delta x < 1,000$ m, has been termed the “terra incognita” (Wyngaard 2004), where the traditional approaches to PBL parameterizations begin to lose their applicability. Thus, the choice of $\Delta x = 750$ m makes the HRRRNEST a useful platform for model development within terra incognita. The domain of the HRRRNEST is shown by the small green box in Fig. 2.

The complete set of model physical parameterizations and relevant numerical methods targeted for development in WFIP2 is summarized in Table 1. This set of components represents a combination of new parameterizations, improvements to existing parameterizations, and improvements to numerical methods. Together, this set of updated model components is hypothesized to address some of the science challenges discussed above, while also being suitable for future implementation in the operational RAP and

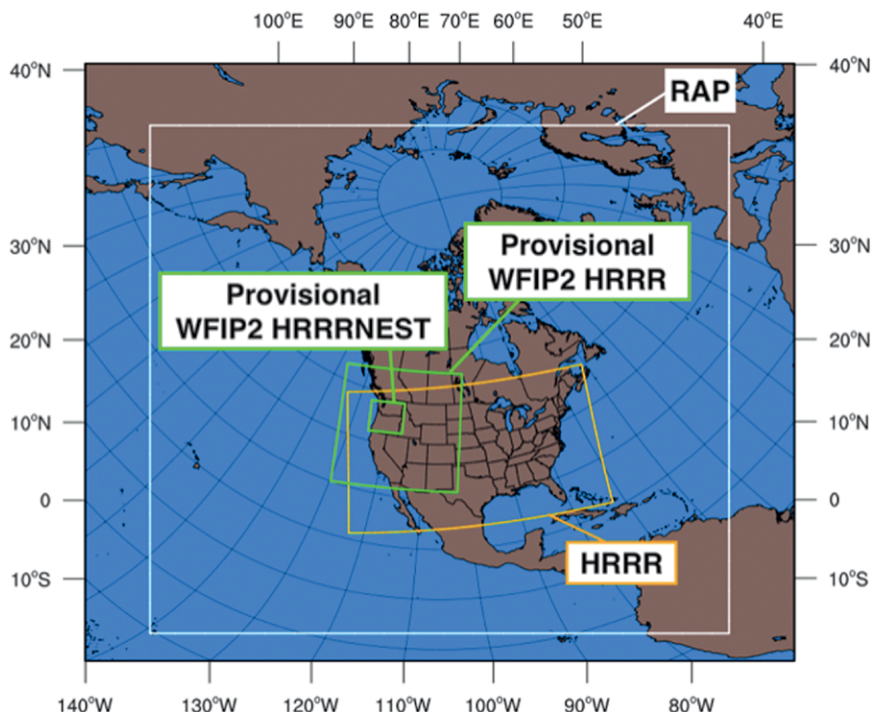


FIG. 2. Map showing the domains of the RAP ($\Delta x = 13$ km; white box), HRRR ($\Delta x = 3$ km; gold box), provisional WFIP2 HRRR ($\Delta x = 3$ km; large green box), and provisional WFIP2 HRRRNEST ($\Delta x = 750$ m; small green box), as annotated.

HRRR. A brief discussion of each component listed in Table 1 follows:

PBL local mixing: Mixing length revision. Mixing lengths describe the distance parcels can be displaced by turbulence processes within a known meteorological environment. They have been singled-out as important factors for regulating the behavior of some turbulence parameterization schemes as far back as Mellor and Yamada (1982). This revision to the mixing length formulation in the Mellor–Yamada–Nakanishi–Niino (MYNN) PBL scheme (Nakanishi and Niino 2009), described in detail in Olson et al. (2019), is focused on improving the forecast performance in stable PBLs. This is accomplished by reformulating the mixing length to be independent of height above ground (i.e., using a “z-less” formulation) whenever strong static stability limits the depth of turbulent eddies to be smaller than the depth of the model layer. This formulation helps to better maintain stable boundary layers.

PBL nonlocal mixing: Mass-flux scheme. The original MYNN PBL scheme only mixed scalars and momentum locally, that is, down the gradient produced by differencing adjacent model levels. This neglects the

representation of the nonlocal turbulent transport by thermal plumes in convective PBLs, which produce countergradient mixing of heat at the top of the PBL and can mix higher momentum aloft down into the PBL. This nonlocal mixing is best represented by mass-flux schemes, which are basically models of thermal plumes that advect parcel properties in the vertical. A mass-flux scheme, following Neggers (2015) and Sušelj et al. (2014), was added to the MYNN PBL scheme, making it an eddy-diffusivity mass-flux (EDMF) scheme. The details of this scheme are described in Olson et al. (2019) and examples of performance in single column testing are shown in Angevine et al. (2018).

SGS clouds and coupling to radiation. The subgrid-scale (SGS) cloud representation of Chaboureaud and Bechtold (2002, 2005) was implemented with minor modifications (Olson et al. 2019). This addition provides both a convective (from the mass-flux scheme) and a nonconvective (from the eddy diffusivity scheme) component of the SGS cloud mixing ratio, cloud fraction, and the SGS buoyancy flux produced by SGS clouds. The primary impact is to improve the surface energy balance, which can then more accurately drive the turbulent mixing.

Drag due to SGS topography. The representation of drag due to SGS orography was added to the HRRR

TABLE 1. Summary of all WFIP2-led model development tasks associated with particular model components, their general impact on forecasts within the CB, and their progress status. In this table, RAP and HRRR versions refer to their respective operational versions at NWS/NCEP. “EXP” implies the new/modified component was used in the experimental simulations. “OPER” implies the new/modified component has been integrated into the operational RAP and HRRR. “UD” implies still under development. See text for discussion.

Physical parameterization	Aspect	General impact on forecasts	Status
ID turbulence (MYNN-EDMF)	Local mixing: z-less mixing length	Improves maintenance of cold pools and stable layers in general by reducing mixing lengths in strong stratification (see Results section, part c).	EXP OPER
	Subgrid-scale (SGS) clouds	Improves representation of SGS stratus, shallow cumulus and the simulated downward shortwave radiation. Small direct impact on low-level winds.	EXP OPER
	Nonlocal mixing: mass-flux scheme	Improves coverage of shallow cumulus and improves profiles of temperature and humidity compared to LES. Small impact on low-level winds during the day.	EXP OPER
3D turbulence	Scale-aware 3D-TKE scheme	Improved representation of finescale turbulence. Expected benefits at subkilometric scales.	UD
Subgrid-scale orographic drag	Small-scale gravity wave drag (SSGWD) and form drag	SSGWD improves the maintenance of cold pools by reducing the near-surface wind speeds and therefore reduces the near-surface vertical wind shear in stable conditions (see Results section, part c). Form drag has smaller impact than the SSGWD at high resolutions ($\Delta x < 5$ km).	EXP OPER
Surface layer	3D surface momentum fluxes	Includes horizontal fluxes. Expected benefits at subkilometer scales only.	UD
Wind farm parameterization	Elevated momentum drag and TKE source	Including wind directional awareness and rotor-equivalent wind speed (small impact). Improves high wind speed bias within wind farms but can contribute to a slight low wind speed bias near wind farms.	EXP
Numerics	Aspect	General impact on forecasts	Status
Horizontal finite differencing	1) Diffusion gradient in physical x – y – z coordinates (not on σ levels), 2) Modified numerical stencils for horizontal diffusion	Horizontal diffusion calculations in physical space improve the maintenance of cold pools by no longer mixing vertically when model vertical coordinates follow steep terrain (see “Cold pool improvements” section). Modified stencils for horizontal diffusion show error reductions in idealized cases.	EXP OPER

physics suite in two forms: a small-scale gravity wave drag (Steenneveld et al. 2008; Tsiringakis et al. 2017) and form drag (Beljaars et al. 2004). The small-scale gravity wave drag acts in stable PBLs and the form drag acts in all conditions. Both are tapered off by $\Delta x = 1$ km, so neither are active in the HRRRNEST. Improved representation of drag will be shown to improve near-surface wind speeds.

Surface layer scheme. Fundamental to Monin–Obukhov theory is the flat-terrain approximation, implying that all momentum, heat, and moisture fluxes are in the vertical. However, horizontal stresses can become as large as the vertical stresses when the slopes become sufficiently large. The 3D surface flux algorithm of Epifanio (2007) has been targeted for implementation, with the intent to interface with 1D PBL schemes as well as 3D schemes.

3D turbulence scheme. A new 3D TKE scheme was developed to improve very high resolution ($\Delta x < 1$ km) simulations, where the impact of horizontal fluxes can be of similar magnitude as the vertical fluxes. This new scheme includes a diagnostic parameterization of all six turbulent 3D stress components and computational stress divergence. A separate manuscript is in preparation which will detail the features of this new scheme and highlight improvements in case studies.

Horizontal finite differencing. The WRF Model has had the capability to perform horizontal diffusion in Cartesian space instead of along terrain-following sigma coordinates for years, but this option was not sufficiently computationally stable for operational use. During WFIP2, several bug fixes were found in the horizontal diffusion code and modifications were introduced to improve the conservation of scalars. Improvements to the maintenance of mountain valley cold pools are demonstrated in the results section and are shown to reduce errors for a commonly used atmosphere-at-rest test case (Lundquist 2018). This option is a replacement to mixing along sigma coordinates, which can produce artificial vertical mixing when the vertical coordinated become sloped along steep topography.

Wind farm parameterization. A representation of wind farm drag was introduced by adopting the WRF wind farm parameterization (Fitch et al. 2012, 2013a,b). Additional work to add the effects of wind directional changes across the rotor layer as well as the use of the rotor-equivalent wind speed as an alternative to the hub-height wind speed was also investigated (Redfern et al. 2019).

TESTING FRAMEWORK AND STRATEGY.

Addressing the model development goals in WFIP2 required a two-stage approach. First, starting from known limitations in the operational RAP/HRRR physics suite and an a priori knowledge of systematic wind speed forecast errors in the operational RAP/HRRR, specific components of the model were targeted for development at the outset of WFIP2. Later, during the field campaign, real-time model validation against measurements permitted a more complete characterization of wind speed forecast errors within the rotor layer. This real-time comparison of model forecasts with field measurements was highlighted in weekly weather discussions, which included scientists from the public and private sectors. These activities were essential for defining the industry's primary forecast problems and were crucial in shaping the model-development priorities during the field project.

In the first stage of model development, candidate model components (i.e., new or modified) were developed to alleviate systematic model forecast errors. Development of these candidate components utilized a hierarchy of approaches (Fig. 3, left side), which include both single-column model (SCM) tests and 3D test cases. SCM testing was used primarily as an early step of development, to determine code functionality in an idealized framework. Following satisfactory SCM tests, candidate components were tested in 3D case studies and validated against WFIP2 observational data. Cases were often run numerous times to incorporate new changes until simulated low-level flow features better matched the observational data. During this stage, other forecast fields in the CB (e.g., cloud cover, 2-m temperature, precipitation) were also compared against the control simulation and conventional data to ensure that the model changes were not detrimental to these fields. Next, candidate components were tested in non-WFIP2 cases outside the CB to demonstrate that these components did not adversely impact forecasts of other phenomenon not pertinent to WFIP2 (e.g., severe convection, low cloud ceilings, lake-effect snow events).

The set of candidate model components that had successfully emerged from developmental testing were aggregated and deemed the *experimental* model configuration (Table 2, third column). Although this configuration encapsulates a large portion of WFIP2-related development, some candidate components, such as the 3D surface-layer and 3D turbulence schemes, were not included in the experimental configuration. Development of these schemes is regarded as a longer-term (beyond 3 years) research topic. A *control* model configuration, representing

Model Testing Framework

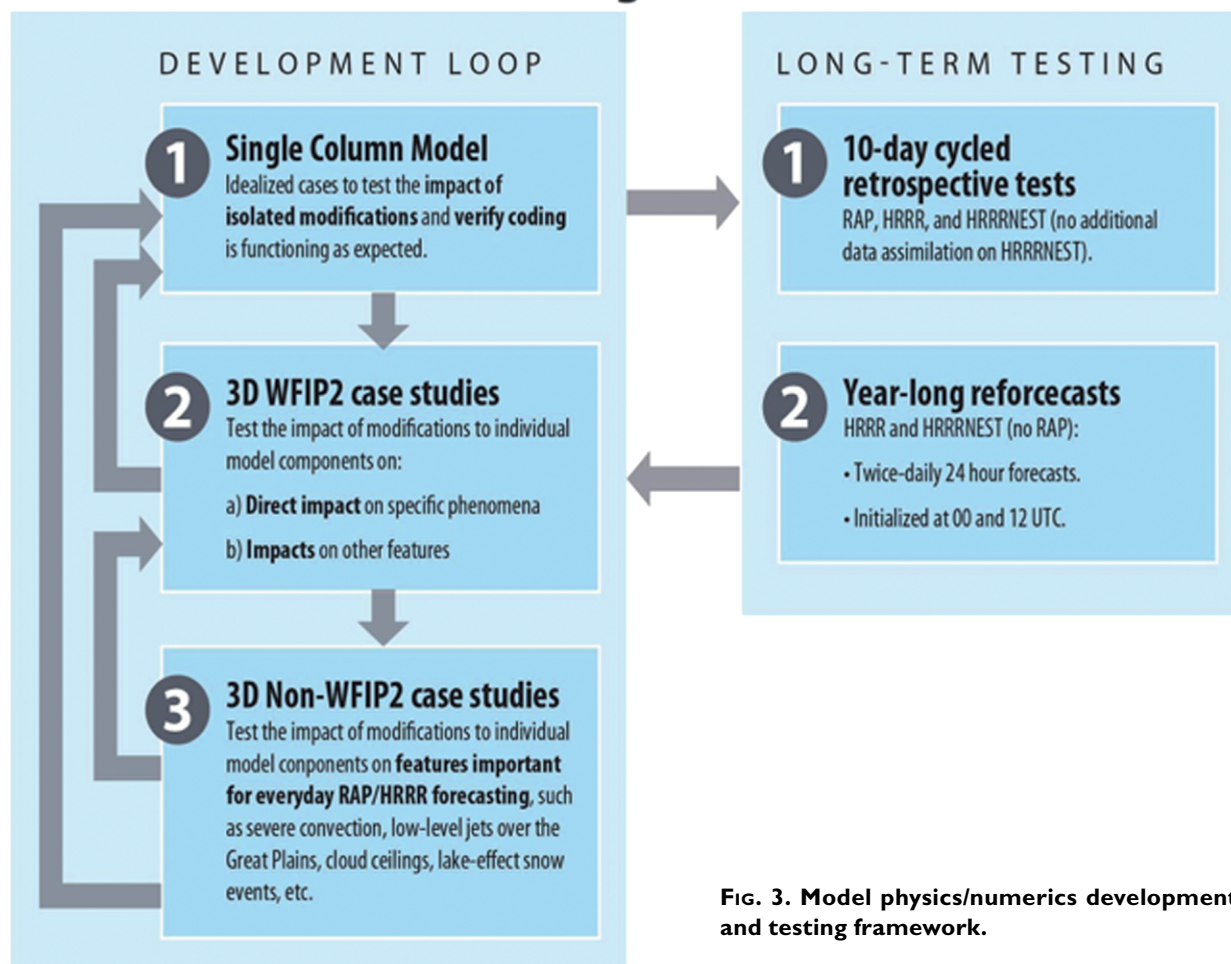


FIG. 3. Model physics/numerics development, and testing framework.

the state of the operational RAP/HRRR physics suite and numerics at the outset of WFIP2 circa summer of 2015, was also defined (Table 2, second column), which approximately corresponds to the configuration used in the operational RAPv2 and HRRRv1 (these versions were run operationally at the beginning of WFIP2). Both the control and experimental simulations utilized the same underlying WRF-ARW code version, such that forecast-performance differences are solely attributable to the configuration differences in Table 2.

The experimental model configuration is hypothesized to produce the forecast improvements sought in WFIP2 relative to the control configuration. To test this hypothesis, two long-term model-production frameworks—retrospectives and reforecasts—were utilized for both the control and experimental configurations (Fig. 3, right side). Retrospective tests, consisting of two 10-day forecast periods, utilize the same data-assimilation and cycling procedures

as exist in the operational RAP and HRRR. Due to their limited duration, results of retrospectives will not be shown. Reforecasts, on the other hand, consist of four forecast periods of ~6 weeks each, centered at the middle of each season (Table 3), thereby sampling a considerable portion of the annual cycle.

RESULTS. This section will highlight the test results from the reforecasts, presented in a variety of ways to determine the degree of forecast improvement and to assess which weather regimes were most improved.

80-m wind speed evaluation. Each set of simulations (HRRR and HRRRNEST, control and experiment) is compared against the 19 sodars deployed throughout the WFIP2 study area (Fig. 1b). A pronounced diurnal cycle in 80-m wind speed mean absolute error (MAE) is evident in the four-season composites for the HRRR (Fig. 4a) and HRRRNEST (Fig. 4d), with the largest MAEs found at night for both the control and

TABLE 2. Control, experimental, and possible future model configurations of the HRRR with WFIP2-led model developments in bold. All components not under development in WFIP2 (regular font) are described in Benjamin et al. (2016) and references therein.			
Model component	Control	Experimental	Future
Land surface	RUC 9-level	RUC 9-level	RUC 9-level
Surface layer	MYNN	MYNN	3D surface stresses
PBL	MYNN level 2.5	MYNN-EDMF	3D TKE scheme
SW radiation	RRTMG	RRTMG	RRTMG
LW radiation	RRTMG	RRTMG	RRTMG
Microphysics	Thompson Aero	Thompson Aero	Thompson Aero
Shallow convection	—	MYNN-EDMF	Mass flux from MYNN-EDMF
Horizontal diffusion	Smagorinsky on sigma levels	Smagorinsky on cartesian coordinates	3D TKE scheme on cartesian coordinates
Orographic drag	—	Small-scale gravity wave drag (inactive when $\Delta x \leq 1$ km)	Small-scale gravity wave drag form drag (inactive when $\Delta x \leq 1$ km)
Wind farm drag	—	Fitch et al. (2012)	Wind direction-aware, rotor-equivalent wind speed
Vertical levels	51 levels	51 levels	100 levels
Vertical coordinate	Sigma	Hybrid sigma–pressure	Hybrid sigma–pressure with immersed boundary method at $\Delta x \leq 1$ km
Albedo	AVHRR	AVHRR	MODIS + solar zenith angle dependence

experimental versions. The HRRRNEST has slightly smaller errors than the HRRR at night, but the errors are comparable during the day. The improvement in 80-m wind speed forecasts is expressed as the difference in MAE (control minus experiment; Figs. 4b,e). For most hours of the day, the experimental versions have a reduced MAE by about 0.05–0.2 m s^{−1} for the HRRR and 0.02–0.15 m s^{−1} for the HRRRNEST, which results in a 5%–10% improvement for the HRRR and 2%–7% improvement for the HRRRNEST (Figs. 4c,f). The confidence intervals (Figs. 4b,e), show that these results are statistically significant with

95% confidence for only a few hours of the night and early morning.

Considering the individual 6-week periods reveals that the degree of improvement varied substantially for each season (Fig. 5). Spring and summer showed the least improvements, where the HRRRNEST (red lines) generally showed small positive or neutral results throughout the diurnal cycle, while the HRRR (blue lines) struggled during the evening transition (0000–0500 UTC) but was mostly neutral otherwise. The forecast improvements were most robust in the fall and winter (Figs. 5c,d), where both the

TABLE 3. Description of the reforecast model runs used in WFIP2 and their associated time periods.				
Reforecast period name	Inclusive dates (~6 weeks)	Domains	Run cadence and forecast length	Initial and lateral boundary conditions
Spring	25 Mar to 7 May 2016	HRRR and HRRRNEST (provisional domains)	12-hourly forecast initializations (i.e., 0000 and 1200 UTC); forecasts to 24 h	Operational RAP analyses provide cold-start initial conditions and lateral boundary conditions; no data assimilation or cycling
Summer	24 Jun to 7 Aug 2016			
Fall	24 Sep to 7 Nov 2017			
Winter	25 Dec 2016 to 7 Feb 2017			

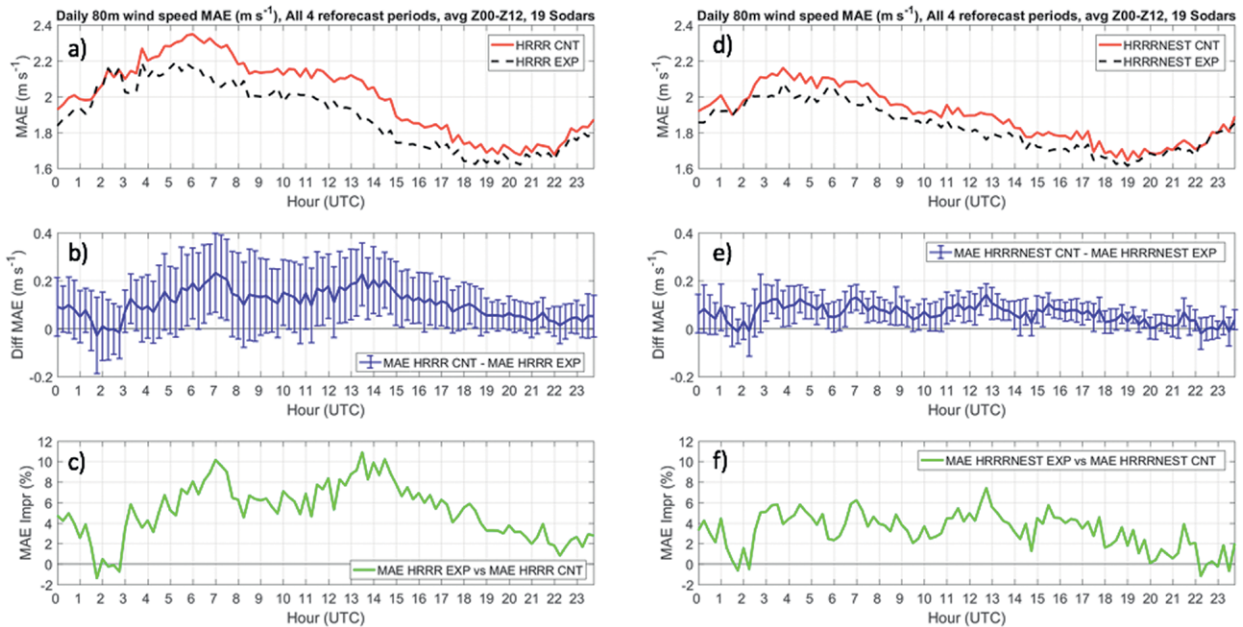


FIG. 4. The 80-m wind speed forecast verification results, as obtained from sodar measurements at 19 locations during all four reforecast periods. Shown are diurnal composites of (a) MAE (m s^{-1}) of the HRRR control (red solid line) and HRRR experimental (black dashed line) configurations; (b) MAE difference (m s^{-1}) between the HRRR control and HRRR experimental configurations (blue solid line), overlaid with 95% confidence intervals; and (c) percentage improvement of the HRRR experimental configuration MAE over the HRRR control configuration MAE (green solid line). (d)–(f) As in (a)–(c), but for HRRRNEST control and HRRRNEST experimental configurations.

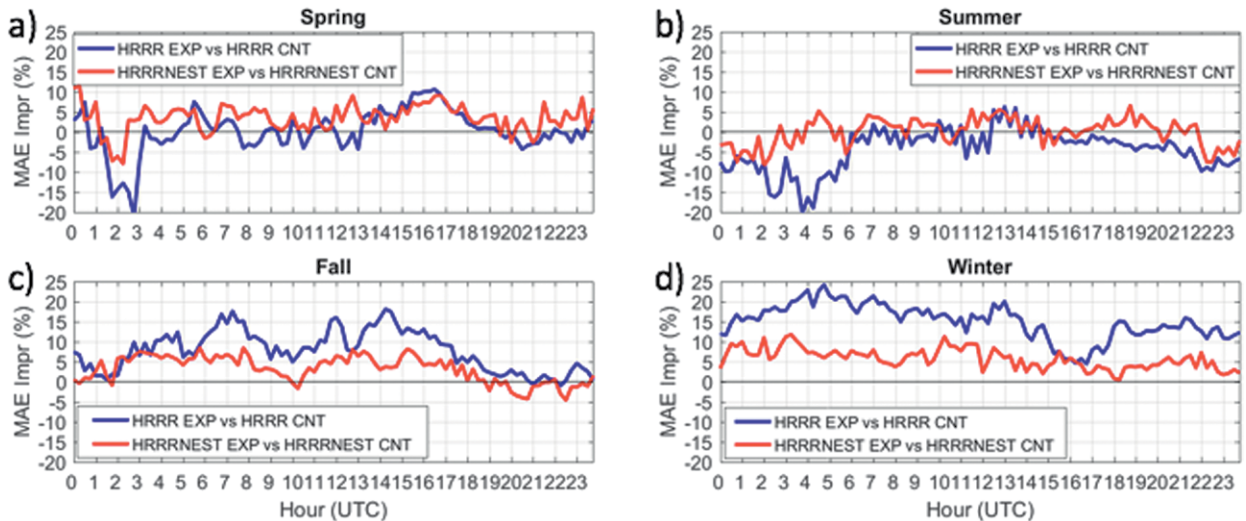


FIG. 5. Diurnal composites of the percentage improvement of 80-m wind speed forecast MAE in the HRRR experimental configuration over the HRRR control configuration (blue) and the HRRRNEST experimental configuration over the HRRRNEST control configuration (red) during the (a) spring, (b) summer, (c) fall, and (d) winter reforecast periods.

improvement to the HRRR and HRRRNEST stayed positive or near-neutral for most of the diurnal cycle. The HRRR showed larger improvements than the HRRRNEST, with MAE reductions often exceeding 10% in the fall and 15% in the winter. These largest improvements were associated with better

maintenance of cold pools (demonstrated in the “Cold pool improvements” section).

Vertical profile evaluation. Vertical profile evaluations of simulated wind speeds help assess how representative the 80-m wind speed improvements are in the rest of

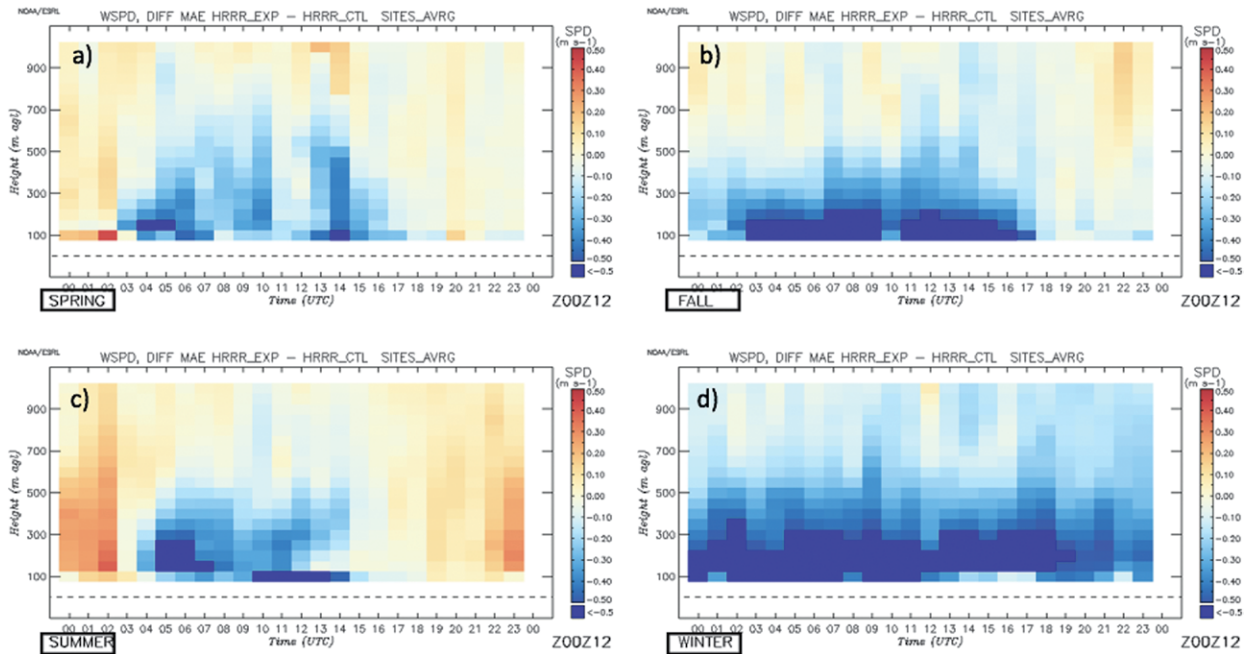


FIG. 6. Diurnal composite time–height cross sections depicting the change in MAE of wind speed forecasts (color fills, m s^{-1}) between the experimental and control HRRR ($\Delta x = 3 \text{ km}$) configurations for the (a) spring, (b) fall, (c) summer, and (d) winter reforecast periods, as verified against eight 915-MHz radar wind profilers. Shades of blue (red) correspond to MAE improvements (degradations) in the experimental configuration over the control configuration.

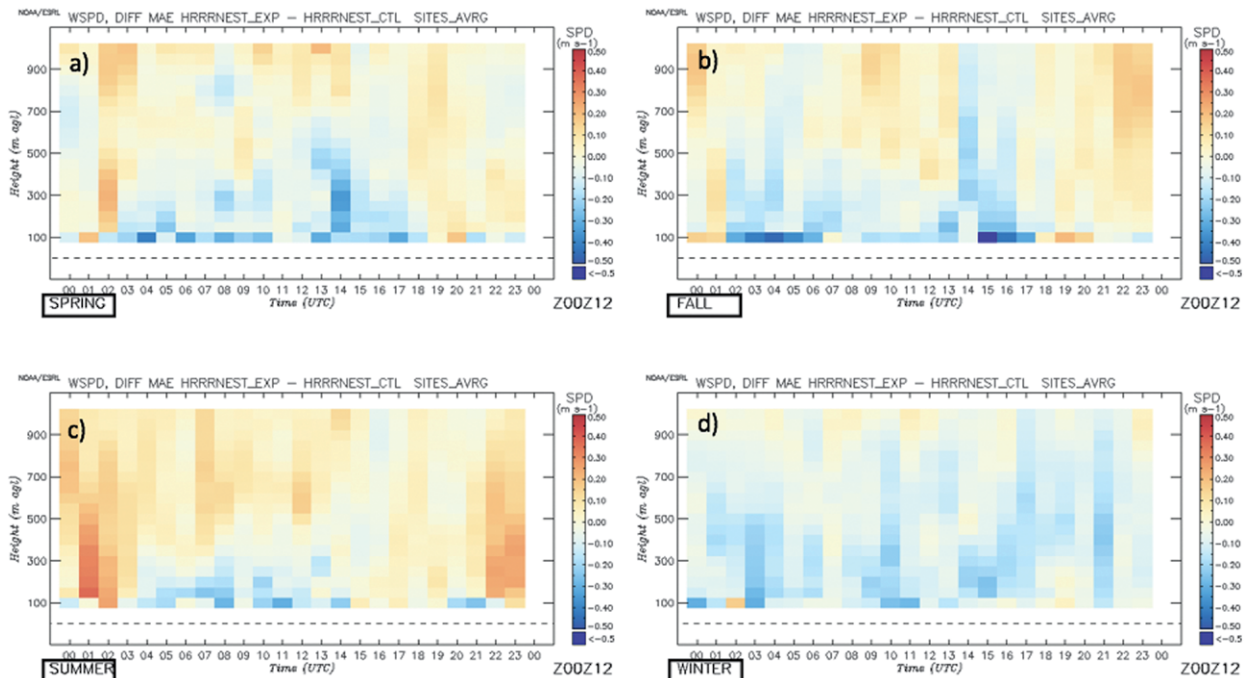


FIG. 7. As in Fig. 6, but for HRRRNEST ($\Delta x = 750 \text{ m}$) forecasts.

the planetary boundary layer. Diurnal composites of the wind speed MAE differences between the HRRR control and experiment are computed by comparison to the eight 915 MHz radar wind profilers in the WFIP2 region (locations in Fig. 1b) for each of the

four seasons (Fig. 6). Reduced wind speed MAEs (blue shades) dominate all seasons, especially at night and during the winter, where reductions in MAE exceed 0.5 m s^{-1} up through 300 m AGL for most of the diurnal cycle. The only significant degradations near the rotor

layer are found in the daytime during the summer, where the control MAE were typically smallest to begin with. Part of this degradation has since been removed by adding momentum transport in the mass-flux scheme, which acts to mix higher wind speed down into the rotor-layer to improve the negative wind speed bias during the daytime (not shown). The depth and magnitude of the improvements in the HRRRNEST are much smaller (Fig. 7). The largest improvements in the HRRRNEST are mostly isolated near the rotor layer for most seasons, and the same degradation seen in the HRRR is found in the daytime summer.

Improvements to wind forecasts are less valuable if accompanied by degradations in the predictions

of other important variables, such as temperature. To investigate this, the HRRR and HRRRNEST reforecasts were compared to Radio Acoustic Sounding System (RASS) virtual temperature measurements for the winter, when the largest wind forecasts improvements were found (Fig. 8). Improvements in virtual temperature MAE were found to exceed 0.5°C in the rotor layer for approximately half the diurnal cycle in the HRRR but less than half that improvement was found in the HRRRNEST.

Cold pool improvements. Improvements to low-level temperature were only robust in the winter, where cold pool mix-outs were found to be the primary forecast challenge. An example of wind speed improvements tied to cold pool mix-out events is shown for a 10-day period in January 2017 (Fig. 9). The mean rotor-layer wind speeds from three sodars located in the middle of the CB (Fig. 9a) show weak wind speeds ($<4\text{ m s}^{-1}$) when cold pools are established and wind ramps occur as the cold pools are sufficiently eroded away, resulting in stronger mean winds ($>4\text{ m s}^{-1}$). The simulated biases (Fig. 9b) and standard deviations (Fig. 9c) differ the most during the period with wind ramps and stronger winds because the forecasted erosion, maintenance, and reestablishment of the cold pools can differ from reality, resulting in significant model errors. The control HRRR (red lines) clearly shows higher biases and standard deviations than the experimental HRRR, suggesting that much of these improvements in wind speeds come from better simulated cold pool depths.

The model components that are primarily

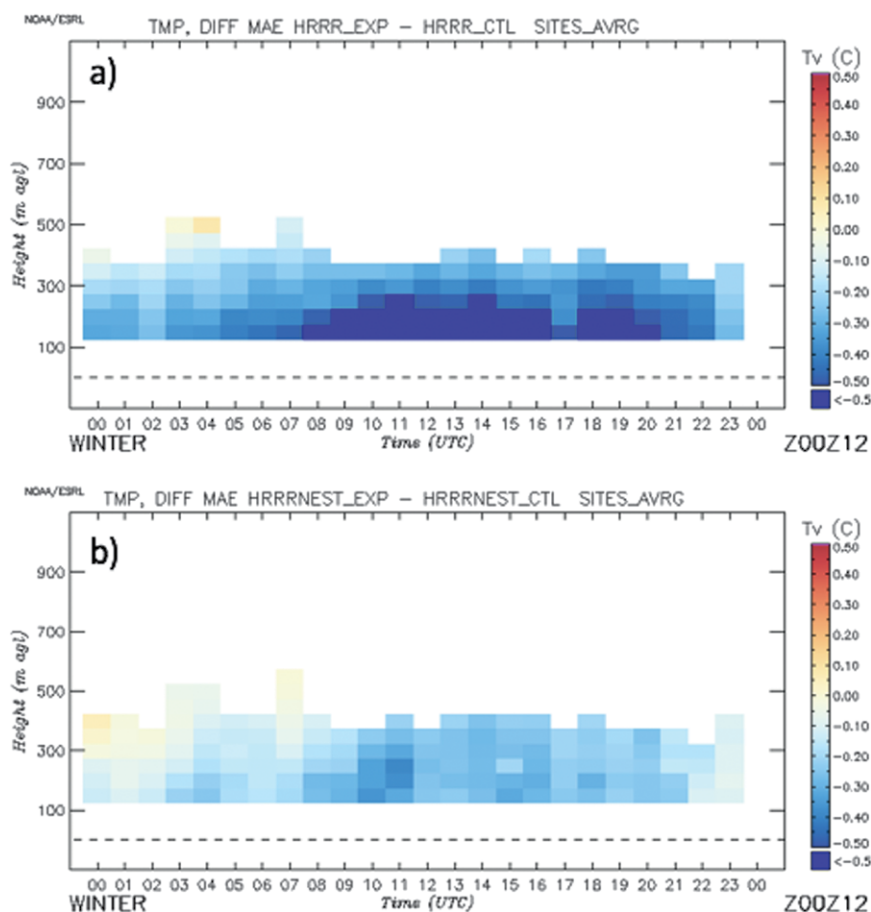
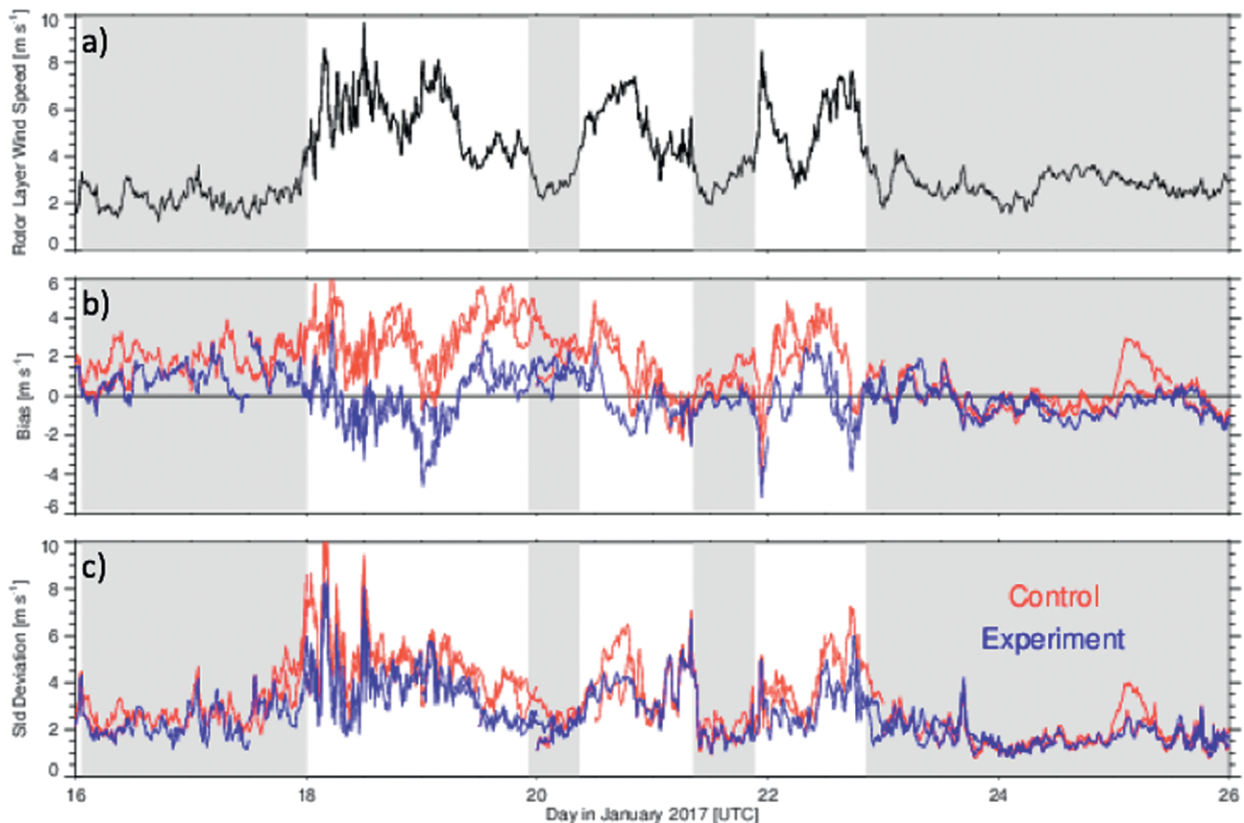
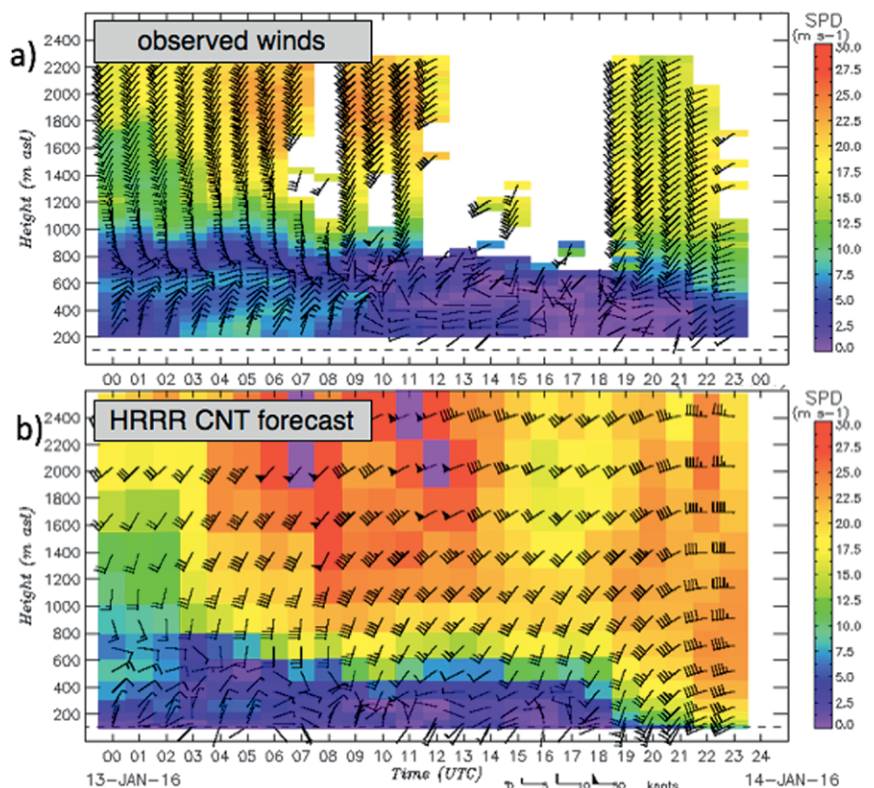


FIG. 8. Diurnal composite time–height sections depicting the change in MAE of virtual temperature forecasts (color fills, K) between the experimental and control for the (a) HRRR ($\Delta x = 3\text{ km}$) and (b) HRRRNEST ($\Delta x = 750\text{ m}$) configurations for the winter reforecast periods, as verified against RASS. Shades of blue (red) correspond to MAE improvements (degradations) in the experimental configuration over the control configuration. RASS measurements are performed during the last 5 min of each hour, with lowest measurements around 100 m and highest measurement around 1,500 m AGL (height coverage of the instrument varies with atmospheric conditions). The range gate spacing was about 60 m (varied slightly at each sites). Previous deployments of RASS, radiosonde, and in situ temperature sensors showed a mean absolute error of $<1\text{ K}$ in temperature (Bianco et al. 2017).



▲ FIG. 9. (a) Mean wind speed from 40 to 140 m (rotor layer) for all 19 sodars for a 10-day period in January 2017. (b) Mean bias for the control HRRR (red) and experimental HRRR (blue), both with $\Delta x = 3$ km. Note that there are two traces, as these are the 0–24 h forecasts for each model initialization (which occurred every 12 h). (c) Mean standard deviation for the HRRR control run (red) and the experimental HRRR (blue). Statistics were computed over the rotor layer for three sodars located in the CB (Wasco, Arlington, and Boardman). The gray shade represents the period when cold pools were present in the CB.



► FIG. 10. (a) Time–height plot of winds at Boardman, Oregon, on 13 Jan 2016, as measured by a 915-MHz wind profiling radar. Wind speeds (m s^{-1}) are shown as color fills according to scale at right, and wind vectors (kt) are shown with conventional barbs. (b) As in (a), but that forecast winds are plotted from the HRRR ($\Delta x = 3$ km) control, initialized at 0000 UTC 13 Jan 2016.

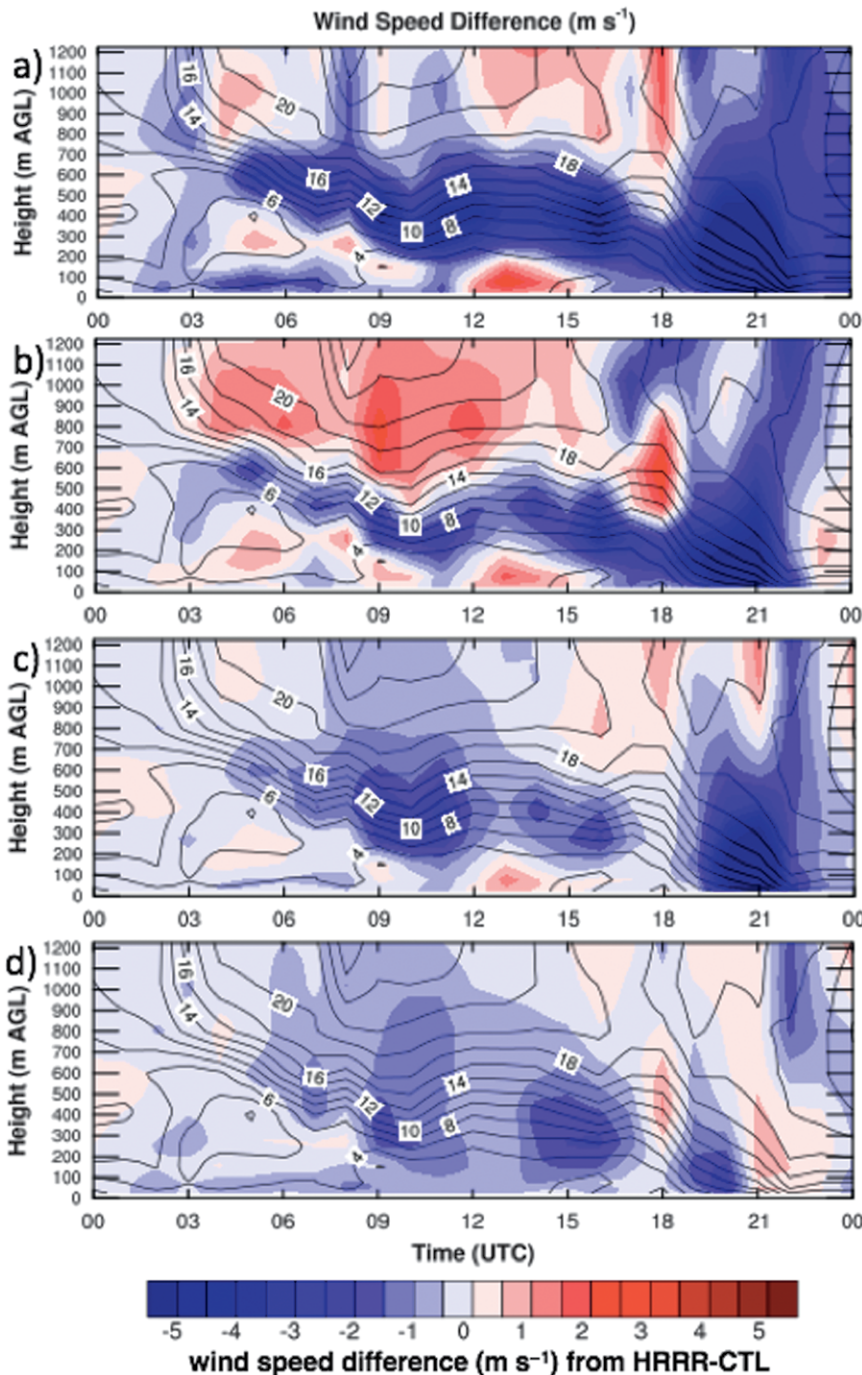


FIG. 11. Time–height plots of forecast wind speed at Boardman, Oregon, for a series of four HRRR ($\Delta x = 3$ km) forecasts initialized at 0000 UTC 13 Jan 2016. In all panels, wind speed (m s^{-1}) is shown as black solid lines, and the wind speed difference (m s^{-1}) from the HRRR control forecast (i.e., the forecast shown in Fig. 10b) is shown as color fills according to scale at bottom. In (a), the forecast uses the HRRR experimental configuration. In subsequent panels, the HRRR forecasts use (b) the mixing-length revision only, (c) modified horizontal diffusion only, and (d) the small-scale gravity wave drag only, but are otherwise identical to the HRRR control configuration.

responsible for the improved cold pool simulations are showcased in the 13 January 2016 case, which was a poorly forecasted cold pool mix-out event. The observed wind speeds (Fig. 10a) show stagnant winds below 1,000 m ASL due to the stable stratification at the top of the cold pool. The overlying winds at 1,500 m ASL strengthen from 11 to 16 m s^{-1} between 0000 and 0800 UTC and slowly erodes the cold pool to less than 300 m in depth by 0000 UTC 14 January. The control HRRR (Fig. 10b) began with too shallow of a cold pool and eroded it much faster, completely mixing it out by 2000 UTC, resulting in a high-wind-speed bias in the layer between the forecasted cold pool top and the actual cold pool top. The difference in wind speeds between the experimental HRRR and the control HRRR shows a large reduction in wind speeds near the top of the cold pool (Fig. 11a), due to the improved maintenance of the cold pool. The primary model components responsible for this improvement, tested individually, show that the mixing length changes (Fig. 11b) have the largest reduction in wind speeds near the top of the cold pool, followed by the modified horizontal diffusion (Fig. 11c), and the small-scale gravity wave drag (Fig. 11d). Note that since not all cold pools have been resimulated to test the impact of each individual model component, the generality of these results are still unknown.

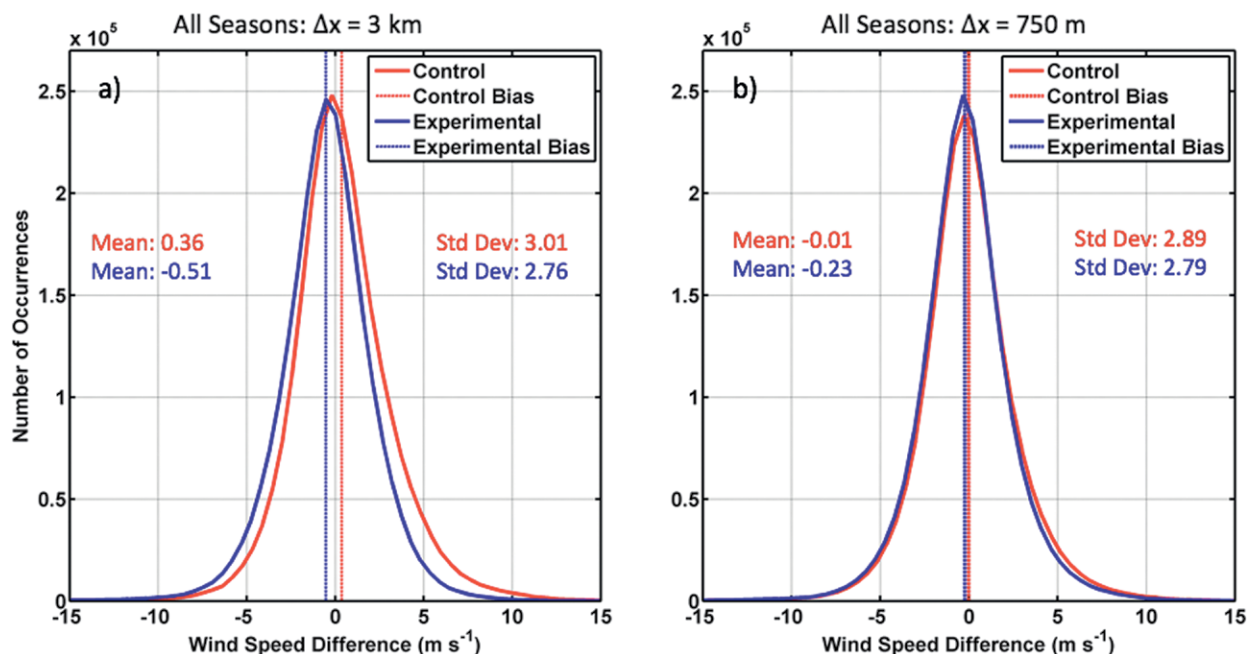


FIG. 12. Histograms of model errors for all hourly forecast hours from all seasons of the reforecast for control (red) and experiment (blue) runs for the (a) HRRR ($\Delta x = 3$ km) and (b) HRRRNEST ($\Delta x = 750$ m). Analysis of the statistical significance of the changes in the mean and standard deviation for both the 3-km and 750-m versions is tested using the Student's *t* test and the *F* test of equality of variance, respectively. In both cases, the null hypothesis is rejected at significance level of 0.05 with a *p* value of zero or near zero due to sheer number of data points ($n > 3$ million).

Large-error forecasts. A primary goal of WFIP2 is to reduce large forecast errors, which are defined as errors greater than two standard deviations from zero, where the standard deviation is taken from the set of control simulations. Large forecast errors of wind resources wreak havoc on electrical grid operators trying to balance the load on the grid. Histograms of the model wind speed errors reveal important changes in characteristics due to the collective model physics changes (Fig. 12). For the HRRR (Fig. 12a), the histogram of model errors becomes slightly thinner for the experimental HRRR (blue), with a standard deviation of $\sigma = 2.76$ m s⁻¹ compared to $\sigma = 3.01$ m s⁻¹ in control HRRR (red), but the mean bias is shifted from positive in the control HRRR (red dashed line) to negative in the experimental HRRR (blue dashed line). This overall shift in the mean bias to become slightly negative comes with a pronounced reduction in the right-side tail, representing large overforecasted wind speed, but also comes with a slight increase in the total number of large underforecasted wind speeds. The experimental HRRRNEST (blue, Fig. 12b) shows a taller peak in the small errors range ($|\text{error}| < 1$ m s⁻¹), with a smaller standard deviation of $\sigma = 2.79$ m s⁻¹ compared to $\sigma = 2.89$ m s⁻¹ in the control (red). There is also a reduction in the large high-wind-speed

forecast error tail, while only a negligible increase in the large low wind speed forecast error tail. The larger shift toward negative wind speed biases in the HRRR is caused by the additional orographic drag employed at $\Delta x = 3$ km, which is not activated when $\Delta x < 1$ km. Future revisions of the orographic drag in the HRRR will address this issue.

The frequency of large forecast errors plotted as a function of forecast length for both the HRRR (Fig. 13a) and HRRRNEST (Fig. 13b) shows a systematic reduction in large forecast errors in the experimental runs (blue) compared to the control (red), even at longer forecast times (>12 h). The reduction is much larger in the winter for both the HRRR and HRRRNEST. The overall mean reductions in the frequency of large forecast errors for all seasons are 30.4% in the HRRR (3.7% in control, 2.6% in experimental) and 11.7% in the HRRRNEST (3.5% in control, 3.1% in experimental). Note that the model improvements in the experimental HRRR result in a lower mean frequency of large forecast errors than that found in the experimental HRRRNEST in the winter and all seasons combined. This may be due to the problem of more detailed wind features at higher resolution being penalized by objective point-based model validation (Mass et al. 2002; Done et al. 2004).

Wind ramps. Another primary goal of WFIP2 is to

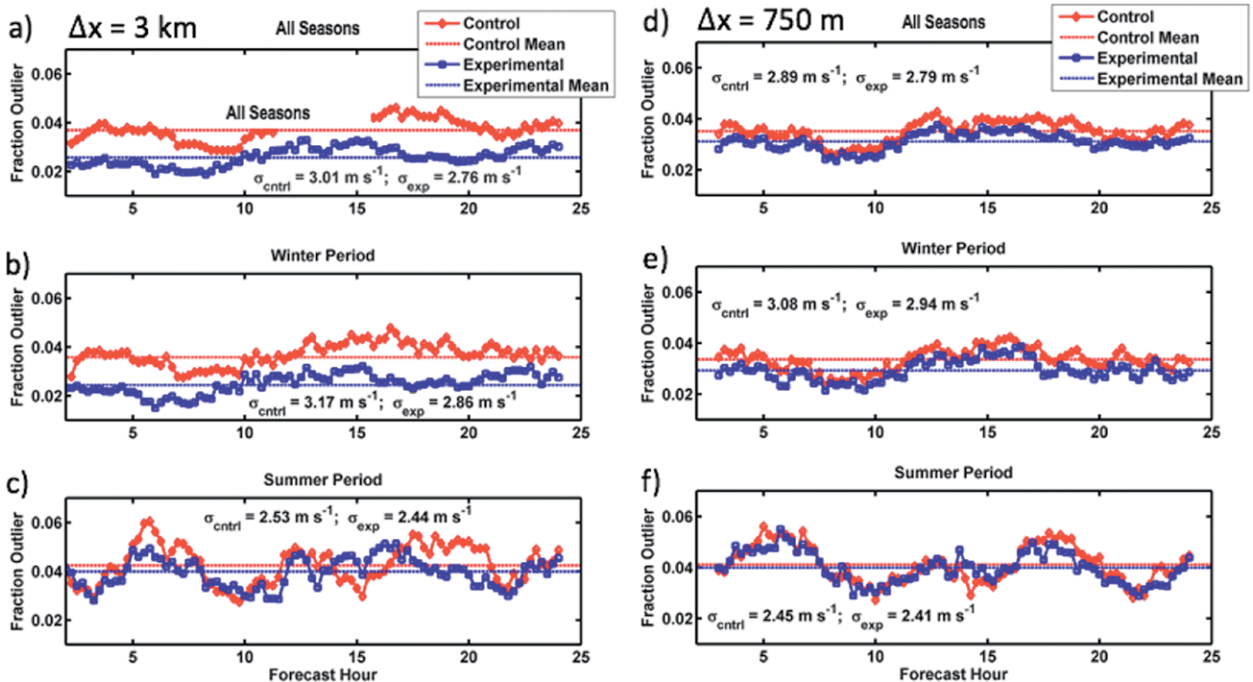


FIG. 13. (a) Forecast-length composite of the fraction of rotor layer wind speed forecasts verifying as outliers (i.e., $|\text{error}| > 2\sigma$ from the observed wind speed) for both the HRRR control configuration (red solid line) and HRRR experimental configuration (blue solid line), both with $\Delta x = 3$ km, aggregated across all four reforecast periods, as verified against lidar measurements at two locations. The mean outlier fractions for the control and experimental configurations are shown by horizontal red and blue lines, respectively. (b), (c) As in (a), but for the winter and summer reforecast periods, respectively. (d)–(f) As in (a)–(c), but for the HRRRNEST ($\Delta x = 750$ m).

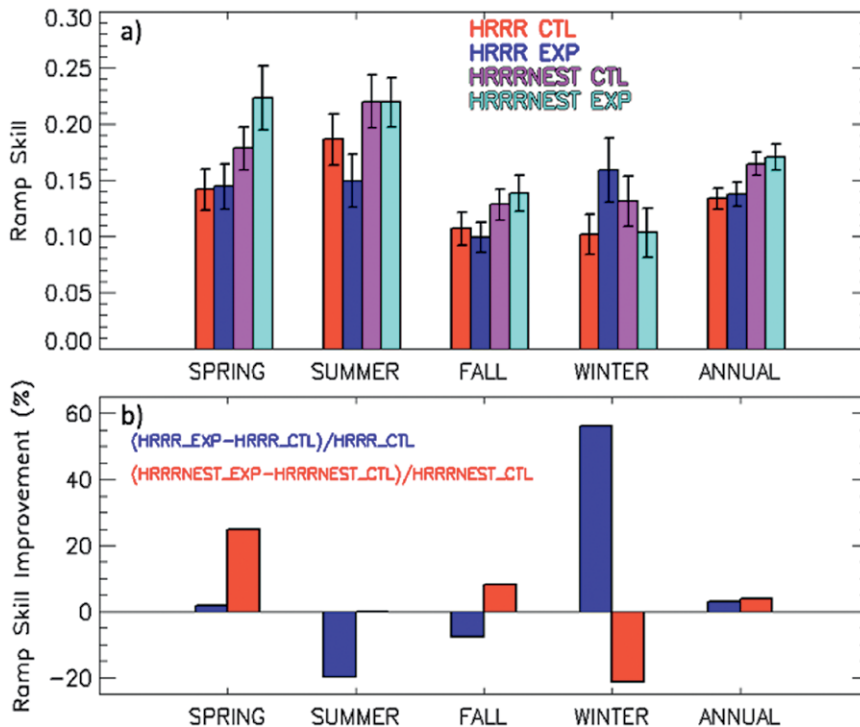


FIG. 14. (a) The ramp skill score for each model seasonally and annually with error bars indicating statistical significance $[\sigma/(N - 1)^{1/2}]$ and (b) the percentage improvement in the forecast skill due to improved physics for the HRRR (blue; $\Delta x = 3$ km) and HRRRNEST (red; $\Delta x = 750$ m) models.

improve the forecast skill of wind ramps. Wind ramps are large changes in wind speeds (i.e., $\Delta U > 3 \text{ m s}^{-1}$) over short periods of time (minutes to a few hours) that make wind power generation extremely volatile. By use of the ramp metric tool developed in the first WFIP (Bianco et al. 2016), ramp skill was diagnosed for each season and model (Fig. 14, upper panel) using the full set of 80-m wind speeds from 19 sodars. A positive ramp skill is found for every model seasonally and annually. The relative improvement/degradation (Fig. 14, lower panel) varies significantly with each season with a positive annual improvement due to the improved physics in the

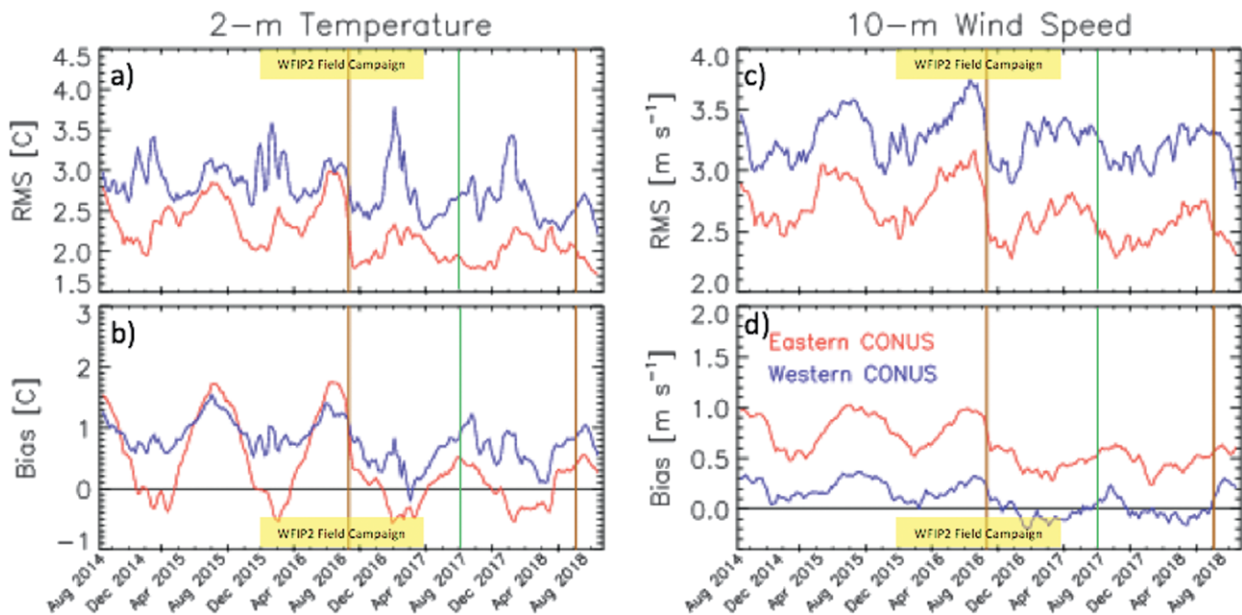


FIG. 15. Multiyear validation of the 12-h forecasts from the operational HRRR ($\Delta x = 3$ km) against METAR data over the western (blue) and eastern (red) CONUS areas for (a),(b) 2-m temperature and (c),(d) 10-m wind speed. Panels (a) and (c) show the RMS and panels (b) and (d) show the bias. The brown vertical lines represent the operational transitions between HRRRv1 to HRRRv2 and HRRRv2 to HRRRv3. The yellow boxes represent the time period of the WFIP2 field campaign and the green lines represent the time of the model development code freeze to begin the experimental simulations.

HRRR and HRRRNEST. The largest improvement of nearly 60% is found in winter for the HRRR. The winter improvement is statistically significant as seen by the nonoverlapping error bars in the upper panel between the control and experimental HRRR models.

Improvements extend to standard forecast metrics. It is important that the improvements to the RAP/HRRR physics suite can be transferred to future operational versions. However, the operational model upgrade process is contingent upon proving that these modifications do not negatively impact other key variables important for general weather forecasts (i.e., 2-m temperature, 10-m wind speed, precipitation, cloud ceilings, etc.). An example that these changes can maintain (and even improve) the forecasts of 2-m temperature and 10-m wind speed as they are integrated into each successive operational HRRR upgrade is shown in Fig. 15. Early versions of the subgrid clouds and mixing length revision were integrated into HRRRv2 (August 2016) and the mass-flux scheme, horizontal diffusion, small-scale gravity wave drag, and further refinements to the subgrid clouds and mixing lengths were integrated into HRRRv3 (July 2018). With each upgrade, the biases and RMS error show improved skill over east and west CONUS, demonstrating that the model improvements from WFIP2-led efforts

were successfully integrated into the operational HRRR along with other improvements from non-WFIP2 efforts.

SUMMARY. Throughout the 18-month WFIP2 field study, several modes of model forecast error were identified in the operational real-time HRRR by way of cross-institutional coordinated efforts to compare model forecasts with observations. Valuable insight from the private sector participants informed model developers of key forecast challenges that were specific to the wind energy industry. Meteorological characteristics of each day within the 18-month period were captured in an event log and important case studies were selected to focus model development efforts around. A set of model physical parameterizations in the RAP/HRRR physics suite were targeted for development based on previously known deficiencies and on the findings from the efforts above. The set of model components under development which showed improvements in particular case studies were promoted to an experimental physics suite, replacing the preexisting components or added as new components. Multiseasonal reforecasts with the control and experimental versions of the HRRR and HRRRNEST were performed to assess the impact of the model physics changes on the forecast skill of rotor-layer winds.

The success of model development efforts within WFIP2 is demonstrated by comparisons of the control and experimental versions of the HRRR and HRRRNEST over the multiseason reforecast simulations. The average 80-m wind speed MAE at all 19 sodar sites within the WFIP2 region has been reduced by 4%–10% averaged over all seasons, and by 20%–30% over the winter. This translates into reduced MAE of power forecasts of about 5%–12% averaged over all seasons. The largest reduction in power (and wind speed) MAE are found within the stable PBL irrespective of the forecast length. The increased skill in forecasting 80-m wind speed was greater for model grid spacing of $\Delta x = 3,000$ m than $\Delta x = 750$ m in the winter, due to 1) more physical parameterization development work completed for mesoscale modeling applications ($\Delta x > 1,000$ m) compared to terra incognita applications ($500 \text{ m} < \Delta x < 1,000 \text{ m}$) and 2) the original model wind speed errors at $\Delta x = 3,000$ m being generally larger than at $\Delta x = 750$ m, making more room for improvement.

In addition to improvements found in the mean statistics, we also demonstrate improvements to other practical forecast needs, such as reduction in the frequency of very large forecast errors and improvements to wind ramp forecasts. Mean frequency of large errors, defined as errors greater than two standard deviations from zero, are reduced by about 30% and 12% averaged over all four seasons at grid spacing of $\Delta x = 3,000$ m and $\Delta x = 750$ m, respectively. Regionally aggregated wind ramp forecasts validated against the full set of 80-m level wind speeds from 19 sodars, show a 10% and 2% improvement averaged over four seasons in the HRRR and HRRRNEST, respectively, with the most positive improvement up to 60% in winter for the HRRR. These results are encouraging because operational forecasts at $\Delta x = 750$ m for the entire operational HRRR domain are not feasible at this time, but the implemented physics changes help the 3-km model be more competitive with the high-resolution runs, at least in the study area.

Despite the positive results demonstrated in this extensive model development effort, this only represents a first-step in a much longer-term challenge. Further analysis of the diverse multiscale set of observation data and comparisons with model reforecasts is in progress. The goal is to further improve our understanding of the physical processes important in regulating the rotor-layer winds, as well as refine our characterization of the model errors in order to direct further model development efforts. Analysis of other quantities, such as simulated PBL height, cloud cover, surface fluxes, and radiation are uncovering new

model development opportunities that may indirectly provide feedback improvements to low-level winds. Other ongoing HRRR model development focused on general weather improvement have proceeded since the WFIP2 project ended and must eventually be examined in the WFIP2 context. Their impact on rotor-layer winds is still unknown.

Revisiting the extensive WFIP2 observational dataset many times over the following decades will be an advantage to any model physics developer or scientist working on research in complex terrain, and may yield new and interesting science updates. For example, the WFIP2 observations and improved HRRR model output is used successfully in the Mesoscale–Microscale Coupling project (Haupt et al. 2019) with the aim of providing the boundary conditions and a validation dataset for realistic high-resolution large-eddy simulations through wind farms. We anticipate that the diverse multiscale set of observations captured within and surrounding the CB will inspire many future process studies and model development efforts that can extend upon the forecast improvements already achieved during WFIP2.

ACKNOWLEDGMENTS. Funding for this work was provided by the U.S. DOE Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. A portion of this work was supported by NOAA's Atmospheric Science for Renewable Energy (ASRE) program. This work was authored in part by NREL, operated by the Alliance for Sustainable Energy, LLC, for the U.S. DOE, under Contract DE-AC36-08GO28308. A portion of this work was prepared by LLNL under Contract DE-AC52-07NA27344. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. We are grateful to the National Center for Atmospheric Research Mesoscale and Microscale Meteorology Laboratory (www.mmm.ucar.edu/wrf/users), which is responsible for the Weather Research and Forecasting Model. The views expressed in the article do not necessarily represent the views of the U.S. DOE or the U.S. government.

REFERENCES

- Angevine, W. M., J. B. Olson, J. S. Kenyon, W. Gustafson, S. Endo, and K. Sušelj, 2018: Shallow cumulus in a mesoscale model evaluated with the LASSO framework. *Mon. Wea. Rev.*, **146**, 4303–4322, <https://doi.org/10.1175/MWR-D-18-0115.1>.
- AWEA Data Services, 2017: U.S. Wind Industry Fourth Quarter 2017 Market Report. 17 pp., www.awea

- .org/resources/publications-and-reports/market-reports/2017-u-s-wind-industry-market-reports.
- Baker, R. W., E. W. Hewson, N. G. Butler, and E. J. Warchol, 1978: Network of wind power potential in the Pacific Northwest. *J. Appl. Meteor.*, **17**, 1814–1826, [https://doi.org/10.1175/1520-0450\(1978\)017<1814:WP P1TP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1978)017<1814:WP P1TP>2.0.CO;2).
- Beljaars, A. C., A. R. Brown, and N. Wood, 2004: A new parametrization of turbulent orographic form drag. *Quart. J. Roy. Meteor. Soc.*, **130**, 1327–1347, <https://doi.org/10.1256/qj.03.73>.
- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Bianco, L., I. V. Djalalova, J. M. Wilczak, J. Cline, S. Calvert, E. Konopleva-Akish, C. Finley, and J. Freedman, 2016: A wind energy ramp tool and metric for measuring the skill of numerical weather prediction models. *Wea. Forecasting*, **31**, 1137–1156, <https://doi.org/10.1175/WAF-D-15-0144.1>.
- , K. Friedrich, J. M. Wilczak, D. Hazen, D. Wolfe, R. Delgado, S. Oncley, and J. K. Lundquist, 2017: Assessing the accuracy of microwave radiometers and radio acoustic sounding systems for wind energy applications. *Atmos. Meas. Tech.*, **10**, 1707–1721, <https://doi.org/10.5194/amt-10-1707-2017>.
- Boutle, I. A., J. E. Eyre, and A. P. Lock, 2014: Seamless stratocumulus simulation across the turbulent gray zone. *Mon. Wea. Rev.*, **142**, 1655–1668, <https://doi.org/10.1175/MWR-D-13-00229.1>.
- Chaboureaud, J.-P., and P. Bechtold, 2002: A simple cloud parameterization derived from cloud resolving model data: Diagnostic and prognostic applications. *J. Atmos. Sci.*, **59**, 2362–2372, [https://doi.org/10.1175/1520-0469\(2002\)059<2362:ASCPDF>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<2362:ASCPDF>2.0.CO;2).
- , and —, 2005: Statistical representation of clouds in a regional model and the impact on the diurnal cycle of convection during Tropical Convection, Cirrus and Nitrogen Oxides (TROCCINOX). *J. Geophys. Res.*, **110**, D17103, <https://doi.org/10.1029/2004JD005645>.
- Daniels, M. H., K. A. Lundquist, J. D. Mirocha, D. J. Wiersema, and F. K. Chow, 2016: A new vertical grid nesting capability in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 3725–3747, <https://doi.org/10.1175/MWR-D-16-0049.1>.
- Department of Energy, 2015: Wind Vision: A new era for wind energy in the United States. DOE Rep., accessed 22 September 2018, 350 pp., www.energy.gov/sites/prod/files/2015/03/f20/wv_full_report.pdf.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, <https://doi.org/10.1002/asl.72>.
- Epifanio, C. C., 2007: A method for imposing surface stress and heat flux conditions in finite-difference models with steep terrain. *Mon. Wea. Rev.*, **135**, 906–917, <https://doi.org/10.1175/MWR3297.1>.
- Fitch, A. C., J. B. Olson, J. K. Lundquist, J. Dudhia, A. K. Gupta, J. Michalakes, and I. Barstad, 2012: Local and mesoscale impacts of wind farms as parameterized in a mesoscale NWP model. *Mon. Wea. Rev.*, **140**, 3017–3038, <https://doi.org/10.1175/MWR-D-11-00352.1>.
- , J. K. Lundquist, and J. B. Olson, 2013a: 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>.
- , J. B. Olson, and J. K. Lundquist, 2013b: Parameterization of wind farms in climate models. *J. Climate*, **26**, 6439–6458, <https://doi.org/10.1175/JCLI-D-12-00376.1>.
- Global Wind Energy Council, 2018: Global wind statistics 2017. Global Wind Energy Council, 4 pp., http://gwec.net/wp-content/uploads/vip/GWEC_PRstats2017_EN-003_FINAL.pdf.
- Haupt, S. E., and Coauthors, 2019: On bridging a modeling scale gap: Mesoscale to microscale coupling for wind energy. *Bull. of the Amer. Meteor. Soc.*, in press.
- International Electrotechnical Commission, 2007: Wind energy generation systems - Part 12-1: Power performance measurements of electricity producing wind turbines. IEC 61400-12-1, 90 pp., <https://webstore.iec.ch/publication/26603>.
- Klemp, J. B., 2011: A terrain-following coordinate with smoothed coordinate surfaces. *Mon. Wea. Rev.*, **139**, 2163–2169, <https://doi.org/10.1175/MWR-D-10-05046.1>.
- Leuenberger, D., M. Koller, O. Fuhrer, and C. Schär, 2010: A generalization of the SLEVE vertical coordinate. *Mon. Wea. Rev.*, **138**, 3683–3689, <https://doi.org/10.1175/2010MWR3307.1>.
- Lundquist, K. A., 2018: Truly horizontal finite difference scheme in the WRF model. Tech. Rep. LLNL-TR-947197, 8 pp., www.osti.gov/servlets/purl/1476236.
- , —, and —, 2010: An immersed boundary method for the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, **138**, 796–817, <https://doi.org/10.1175/2009MWR2990.1>.
- , —, and —, 2012: An immersed boundary method enabling large-eddy simulations of flow over complex terrain in the WRF model. *Mon. Wea. Rev.*, **140**, 3936–3955, <https://doi.org/10.1175/MWR-D-11-00311.1>.
- Ma, Y., and H. Liu, 2017: Large-eddy simulations of atmospheric flows over complex terrain using the

- immersed-boundary method in the Weather Research and Forecasting model. *Bound.-Layer Meteor.*, **165**, 421–445, <https://doi.org/10.1007/s10546-017-0283-9>.
- Mahrer, Y., 1984: An improved numerical approximation of the horizontal gradients in a terrain-following coordinate system. *Mon. Wea. Rev.*, **112**, 918–922, [https://doi.org/10.1175/1520-0493\(1984\)112<0918:AI NAOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<0918:AI NAOT>2.0.CO;2).
- Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2).
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875, <https://doi.org/10.1029/RG020i004p00851>.
- Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, <https://doi.org/10.2151/jmsj.87.895>.
- Neggess, R. A. J., 2015: Exploring bin-macrophysics models for moist convective transport and clouds. *J. Adv. Model. Earth Syst.*, **7**, 2079–2104, <https://doi.org/10.1002/2015MS000502>.
- Olson, J. B., J. S. Kenyon, W. M. Angevine, J. M. Brown, M. Pagowski, and K. Sušelj, 2019: A description of the MYNN-EDMF scheme and coupling to other components in WRF-ARW. NOAA Tech. Memo. OAR GSD 61, 37 pp., <https://doi.org/10.25923/n9wm-be49>.
- Redfern, S., J. B. Olson, J. K. Lundquist, and C. T. M. Clack, 2019: Incorporation of the rotor-equivalent wind speed into the Weather Research and Forecasting model's wind farm parameterization. *Mon. Wea. Rev.*, **147**, 1029–1046, <https://doi.org/10.1175/MWR-D-18-0194.1>.
- Schär, C., D. Leuenberger, O. Fuhrer, D. Lüthi, and C. Girard, 2002: A new terrain-following vertical coordinate formulation for atmospheric prediction models. *Mon. Wea. Rev.*, **130**, 2459–2480, [https://doi.org/10.1175/1520-0493\(2002\)130<2459:ANTFVC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2459:ANTFVC>2.0.CO;2).
- 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., and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): General overview. *Bull. Amer. Meteor. Soc.*, **100**, 1687–1699, <https://doi.org/10.1175/BAMS-D-18-0036.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Steenefeld, G. J., A. A. M. Holtslag, C. J. Nappo, B. J. H. van de Wiel, and L. Mahrt, 2008: Exploring the possible role of small-scale terrain drag on stable boundary layers over land. *J. Appl. Meteor. Climatol.*, **47**, 2518–2530, <https://doi.org/10.1175/2008JAMC1816.1>.
- Sušelj, K., T. F. Hogan, and J. Teixeira, 2014: Implementation of a stochastic eddy-diffusivity/mass-flux parameterization into the Navy Global Environmental Model. *Wea. Forecasting*, **29**, 1374–1390, <https://doi.org/10.1175/WAF-D-14-00043.1>.
- Tsirngakis, A., G. J. Steenefeld, and A. A. M. Holtslag, 2017: Small-scale orographic gravity wave drag in stable boundary layers and its impact on synoptic systems and near-surface meteorology. *Quart. J. Roy. Meteor. Soc.*, **143**, 1504–1516, <https://doi.org/10.1002/qj.3021>.
- Wilczak, J. M., and Coauthors, 2015: The Wind Forecast Improvement Project (WFIP): A public–private partnership addressing wind energy forecast needs. *Bull. Amer. Meteor. Soc.*, **96**, 1699–1718, <https://doi.org/10.1175/BAMS-D-14-00107.1>.
- , and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): Observational field campaign. *Bull. Amer. Meteor. Soc.*, **100**, 1701–1723, <https://doi.org/10.1175/BAMS-D-18-0035.1>.
- Wyngaard, J. C., 2004: Toward numerical modeling in the “terra incognita.” *J. Atmos. Sci.*, **61**, 1816–1826, [https://doi.org/10.1175/1520-0469\(2004\)061<1816:TN MITT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1816:TN MITT>2.0.CO;2).
- Zängl, G., 2012: Extending the numerical stability limit of terrain-following coordinate models over steep slopes. *Mon. Wea. Rev.*, **140**, 3722–3733, <https://doi.org/10.1175/MWR-D-12-00049.1>.