Improved Representation of Horizontal Variability and Turbulence in Mesoscale Simulations of an Extended Cold-Air Pool Event

ROBERT S. ARTHUR,[a] TIMOTHY W. JULIANO,[b] BIANCA ADLER,[c,d] RAGHAVENDRA KRISHNAMURTHY,[e] JULIE K. LUNDQUIST,[f,g,h] BRANKO KOSOVIĆ,[b] AND PEDRO A. JIMÉNEZ[b]

[a] *Lawrence Livermore National Laboratory, Livermore, California*
[b] *Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado*
[c] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[d] *NOAA/Physical Sciences Laboratory, Boulder, Colorado*
[e] *Pacific Northwest National Laboratory, Richland, Washington*
[f] *Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, Colorado*
[g] *National Renewable Energy Laboratory, Golden, Colorado*
[h] *Renewable and Sustainable Energy Institute, Boulder, Colorado*

(Manuscript received 23 July 2021, in final form 13 December 2021)

ABSTRACT: Cold-air pools (CAPs), or stable atmospheric boundary layers that form within topographic basins, are associated with poor air quality, hazardous weather, and low wind energy output. Accurate prediction of CAP dynamics presents a challenge for mesoscale forecast models in part because CAPs occur in regions of complex terrain, where traditional turbulence parameterizations may not be appropriate. This study examines the effects of the planetary boundary layer (PBL) scheme and horizontal diffusion treatment on CAP prediction in the Weather Research and Forecasting (WRF) Model. Model runs with a one-dimensional (1D) PBL scheme and Smagorinsky-like horizontal diffusion are compared with runs that use a new three-dimensional (3D) PBL scheme to calculate turbulent fluxes. Simulations are completed in a nested configuration with 3-km/750-m horizontal grid spacing over a 10-day case study in the Columbia River basin, and results are compared with observations from the Second Wind Forecast Improvement Project. Using event-averaged error metrics, potential temperature and wind speed errors are shown to decrease both with increased horizontal grid resolution and with improved treatment of horizontal diffusion over steep terrain. The 3D PBL scheme further reduces errors relative to a standard 1D PBL approach. Error reduction is accentuated during CAP erosion, when turbulent mixing plays a more dominant role in the dynamics. Last, the 3D PBL scheme is shown to reduce near-surface overestimates of turbulence kinetic energy during the CAP event. The sensitivity of turbulence predictions to the master length-scale formulation in the 3D PBL parameterization is also explored.

SIGNIFICANCE STATEMENT: In this article, we demonstrate how a new framework for modeling atmospheric turbulence improves cold pool predictions, using a case study from January 2017 in the Columbia River basin (U.S. Pacific Northwest). Cold pools are regions of cold, stagnant air that form within valleys or basins, and improved forecasts could help to mitigate the risks they pose to air quality, transportation, and wind energy production. For the chosen case study, our tests show a reduction in temperature and wind speed errors by up to a factor of 2–3 relative to standard model options. These results strongly motivate continued development of the framework as well as its application to other complex weather events.

KEYWORDS: Complex terrain; Turbulence; Boundary layer; Cold pools; Bias; Mesoscale models

1. Introduction

a. Background and motivation

Cold-air pools (CAPs) are stable atmospheric boundary layers that form within topographic basins. Because of their association with poor air quality (Olofson et al. 2009; Silcox et al. 2012; Lareau et al. 2013; Whiteman et al. 2014; Chemel et al. 2016), hazardous weather such as fog and freezing rain (Whiteman et al. 2001, 2014; Chachere and Pu 2016), and, more recently, low wind energy resources (Wilczak et al. 2019; McCaffrey et al. 2019; Pichugina et al. 2019), accurate forecasting of CAP events is an important challenge for numerical weather prediction (NWP). However, CAPs remain difficult to predict accurately because of their small-scale dynamics, which are typically unresolved in mesoscale forecast models (Mahrt 1998; Holtslag et al. 2013; Olson et al. 2019). This leads to a strong dependence on model parameterizations for processes such as stably stratified turbulence, land surface exchange, radiation, and cloud microphysics.

Adding to this challenge is the fact that CAPs occur in regions of complex terrain, which is a well-known source of error in many NWP models. Complex terrain-related errors stem from two primary sources. The first is grid skewness associated with the commonly used terrain-following vertical

Denotes content that is immediately available upon publication as open access.

Corresponding author: Robert S. Arthur, arthur7@llnl.gov

DOI: 10.1175/JAMC-D-21-0138.1

coordinate (Gal-Chen and Somerville 1975). A variety of approaches have been taken to address this issue, ranging from the use of hybrid sigma-pressure coordinate systems that flatten with height (Arakawa and Lamb 1977; Simmons and Burridge 1981; Schär et al. 2002; Klemp 2011) to immersed boundary methods that use a Cartesian grid (Lundquist et al. 2010, 2012; Ma and Liu 2017; DeLeon et al. 2018; Bao et al. 2018; Arthur et al. 2020; Wiersema et al. 2020). Improved finite difference stencils have also been used to reduce model errors related to grid skewness (Mahrer 1984; Dempsey and Davis 1998; Klemp 2011; Zängl 2002, 2012). Both Billings et al. (2006) and Arthur et al. (2021) used a "truly horizontal" diffusion scheme to improve CAP predictions in NWP models.

The second source of model error in regions of complex terrain, related to horizontal variability in the governing equations, has received less attention in the NWP literature. Mesoscale forecast models typically employ one-dimensional (1D) planetary boundary layer (PBL) schemes that parameterize turbulent fluxes in the vertical direction only; horizontal turbulent fluxes are assumed to be relatively small and are therefore neglected. In fact, horizontal smoothing is often employed to reduce horizontal variability and increase numerical stability (e.g., Smagorinsky 1993). The assumption of 1D turbulent fluxes is typically valid in the coarse-resolution configurations (e.g., horizontal grid spacings of 3 km or larger) for which mesoscale models have historically been used. However, as advances in computing power allow NWP models to be run at higher resolution (e.g., horizontal grid spacings of 1 km or smaller) with the goal of capturing finer-scale phenomena, this assumption can be violated (Honnert and Masson 2014; Mazzaro et al. 2017; Muñoz-Esparza et al. 2017; Doubrawa and Muñoz-Esparza 2020).

A related concern for high-resolution NWP is the *terra incognita* or "gray zone" (Wyngaard 2004), representing the range of model resolutions for which neither traditional 1D PBL schemes nor large-eddy simulation (LES) schemes are valid. This range depends on atmospheric stability conditions but can span horizontal grid spacings between roughly 1 km and 10 m (Chow et al. 2019). LES is particularly expensive in stable CAP conditions due to the restricted scale of turbulent eddies, making it impractical for NWP applications.

b. 3D PBL scheme

If CAPs, as well as other complex-terrain phenomena, are to be accurately represented in NWP models with subkilometer horizontal grid spacing, improved treatment of horizontal variability is necessary. With this in mind, Kosović et al. (2020) and Juliano et al. (2022) recently implemented a new three-dimensional (3D) PBL scheme in the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2019), version 4.0.3. The scheme, based on Mellor and Yamada (1974, 1982), calculates both horizontal and vertical turbulent fluxes, and adds the divergence of these fluxes to the governing equations for momentum, potential temperature, and moisture.

The 3D PBL scheme is similar to the standard 1D Mellor–Yamada–Nakanishi–Niino (MYNN) level-2.5 model (Nakanishi and Niino 2006) available in WRF. Both solve a prognostic equation for the unresolved (i.e., parameterized) turbulence kinetic energy (TKE),

$$\frac{\partial q^2}{\partial t} + U_k \frac{\partial q^2}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\ell q S_q \frac{\partial q^2}{\partial x_k}\right) + 2(P_s + P_b - \epsilon), \qquad (1)$$

where q^2 is 2 times the TKE, U_k is the resolved velocity vector ($k = 1, 2, 3$), ℓ is the Mellor–Yamada master length scale, S_q is a stability function, P_s is the shear production term, P_b is the buoyancy production term, and ϵ is the dissipation term [see Eq. (9) in Juliano et al. 2022].

In turn, the TKE is used to calculate the necessary turbulent flux terms in the governing equations for momentum, potential temperature, and moisture. With MYNN or other 1D PBL schemes, only the vertical fluxes are calculated, while the effects of horizontal fluxes are usually treated using a two-dimensional (2D) form of the Smagorinsky model (Smagorinsky 1963). However, with the 3D PBL scheme, the full 3D flux tensor is calculated by solving a system of 13 linear algebraic equations [see Eq. (19) in Juliano et al. 2022] at each grid point.

Because of the relatively high cost of solving the full system of equations, it can be useful to neglect horizontal derivatives within the system, allowing the turbulent flux terms to be calculated analytically [see Eq. (22) in Juliano et al. 2022]. Known as the "boundary layer approximation" in the context of Mellor and Yamada (1974, 1982), this makes the cost of the 3D PBL scheme comparable to that of a traditional 1D PBL scheme while still improving representation of horizontal variability relative to a 2D Smagorinsky approach. Moreover, this approximation increases the numerical stability of the 3D PBL scheme by eliminating the possibility of an ill-conditioned matrix system when calculating the turbulent flux tensor. Note that when the boundary layer approximation is used, horizontal derivatives are also neglected in the prognostic TKE equation [Eq. (1)]. A comparison of the 3D PBL scheme using both the full and approximate solutions is presented in Juliano et al. (2022) for several idealized convective cases. However, as discussed further in section 2a, only the approximate solution is considered in this study.

c. Case study selection

The Second Wind Forecast Improvement Project (WFIP2; Shaw et al. 2019) provides a useful test bed for exploring the importance of horizontal variability in CAP forecasts. The project included an 18-month observational campaign (Wilczak et al. 2019) in the complex terrain of the Columbia River basin, where many CAP events were captured (McCaffrey et al. 2019). This was combined with efforts to improve the operational High-Resolution Rapid Refresh (HRRR) model (Benjamin et al. 2016), which is based on the Advanced Research version of WRF, for wind energy forecasting applications (Olson et al. 2019). A special "provisional" setup of the HRRR was used during WFIP2, with a 3-km HRRR domain (d01) covering the western United States and a 750-m HRRRNEST domain (d02) centered over the Columbia River basin (see Fig. 1). Model improvements were tested in a cold-

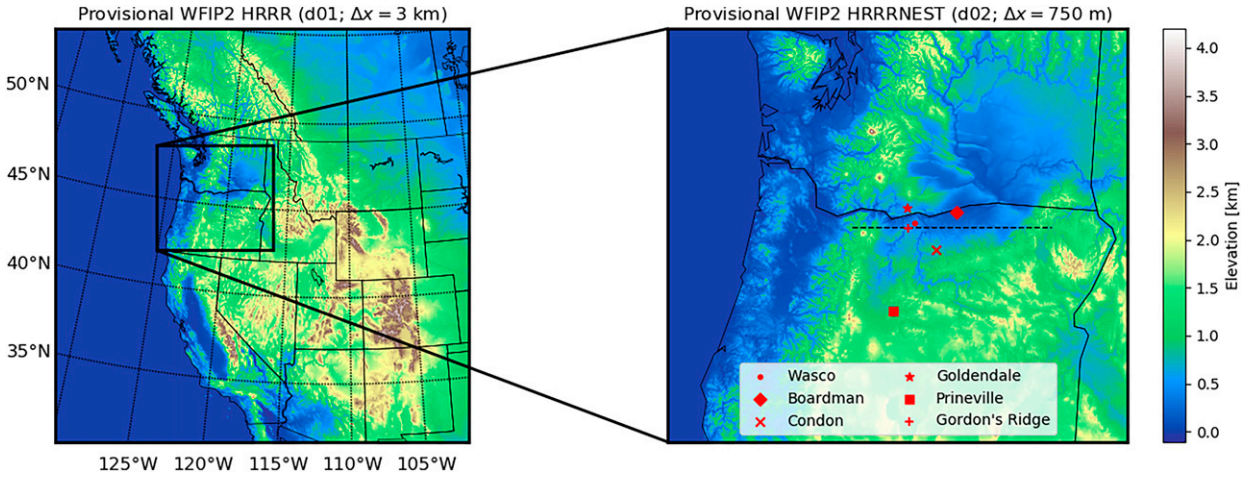

FIG. 1. WRF domain setup for 10-day cold-air pool reforecasts. The provisional WFIP2 HRRR (d01) and HRRRNEST (d02) domains are used in a one-way nesting setup following Olson et al. (2019). On d02, the locations of the observation sites used in this study are shown for reference (see Table 2; note that a symbol is not shown for the Physics site because it would essentially overlap the symbol for Wasco), and the dashed black line denotes the zonal transect shown in Figs. 10 and 12, below.

start reforecast configuration following Olson et al. (2019). The WFIP2 observation datasets and the HRRR/HRRRNEST setup files are publicly available through the U.S. Department of Energy Data Archive and Portal (DAP; http://a2e.energy.gov/data).

As noted by Wilczak et al. (2019), overestimates of near-surface wind speed, and thus of wind power generation, are a common model bias during CAP events. Numerical mixing in the model can also lead to early erosion of the CAP, accentuating this bias and leading to poor predictions of wind energy ramp-up following the event. Improved forecasting of CAP persistence and erosion thus plays an important role in the integration of wind energy into the electrical grid (Francis 2008). In this work, WRF is used with various horizontal diffusion and PBL treatments, including the 3D PBL scheme, to simulate an extended CAP event observed during WFIP2. It is hypothesized that by improving representation of horizontal variability and turbulence, these common model biases can be reduced.

The period of 10–20 January 2017 was chosen for model evaluation from the WFIP2 event log [Atmosphere to Electrons (A2e) 2017b]. This includes a persistent, multiday CAP with relatively steady synoptic conditions and low near-surface wind speeds. Both the initial development of the CAP and its ultimate erosion are also included in the 10-day period. Model results during this time are characteristic of the common biases discussed above, with overestimates of near-surface wind speeds during the CAP and difficulty capturing CAP erosion (see Fig. 2).

This event has been analyzed in a separate study by Adler et al. (2021) that details the observed dynamics of the CAP within the larger-scale meteorological context. Notably, Adler et al. (2021) used a novel algorithm known as the optimal estimation physical retrieval Tropospheric Remotely Observed Profiling via Optimal Estimation (TROPoe; Turner and Löhnert 2014; Turner and Blumberg 2019; Djalalova et al. 2022) to obtain the best possible thermodynamic profiles from available observations. As discussed below, the instruments used for vertical temperature profiling during WFIP2 were generally unreliable during CAP conditions. Thus, the optimal temperature profile dataset of Adler et al. (2021) provides a valuable contribution to the model evaluation in the present study.

Several other WFIP2 observation datasets are also used to evaluate model bias during the 10-day case study (see Table 2 below). They include wind observations between the surface and roughly 2–2.5 km AGL from various instruments, including surface stations, sodars, and radar wind profilers. These observations are detailed in section 3. Observations of TKE from a profiling lidar and a surface station are also used to evaluate the model's turbulence prediction, as detailed in section 4.

2. Model configuration

a. *Turbulence and horizontal diffusion options*

Three separate 10-day reforecasts are completed to compare the performance of the 3D PBL scheme with that of WRF's standard options for mesoscale turbulence parameterization (see Table 1). The WFIP2 "control" configuration (see Table 2 in Olson et al. 2019), which corresponds to the standard HRRR configuration, is used as a baseline in case 1. On both domains, vertical turbulent fluxes are parameterized using the MYNN level-2.5 PBL scheme (bl_pbl_opt = 5). Horizontal diffusion is treated using WRF's 2D Smagorinsky scheme (km_opt = 4), with horizontal gradients calculated along coordinate surfaces (diff_opt = 1).

As noted by Juliano et al. (2022), the 2D Smagorinsky scheme for horizontal diffusion is often used to enhance numerical stability (see Smagorinsky 1993) by smoothing the model solution rather than truly parameterizing horizontal turbulent fluxes. However, previous studies (Olson et al. 2019;

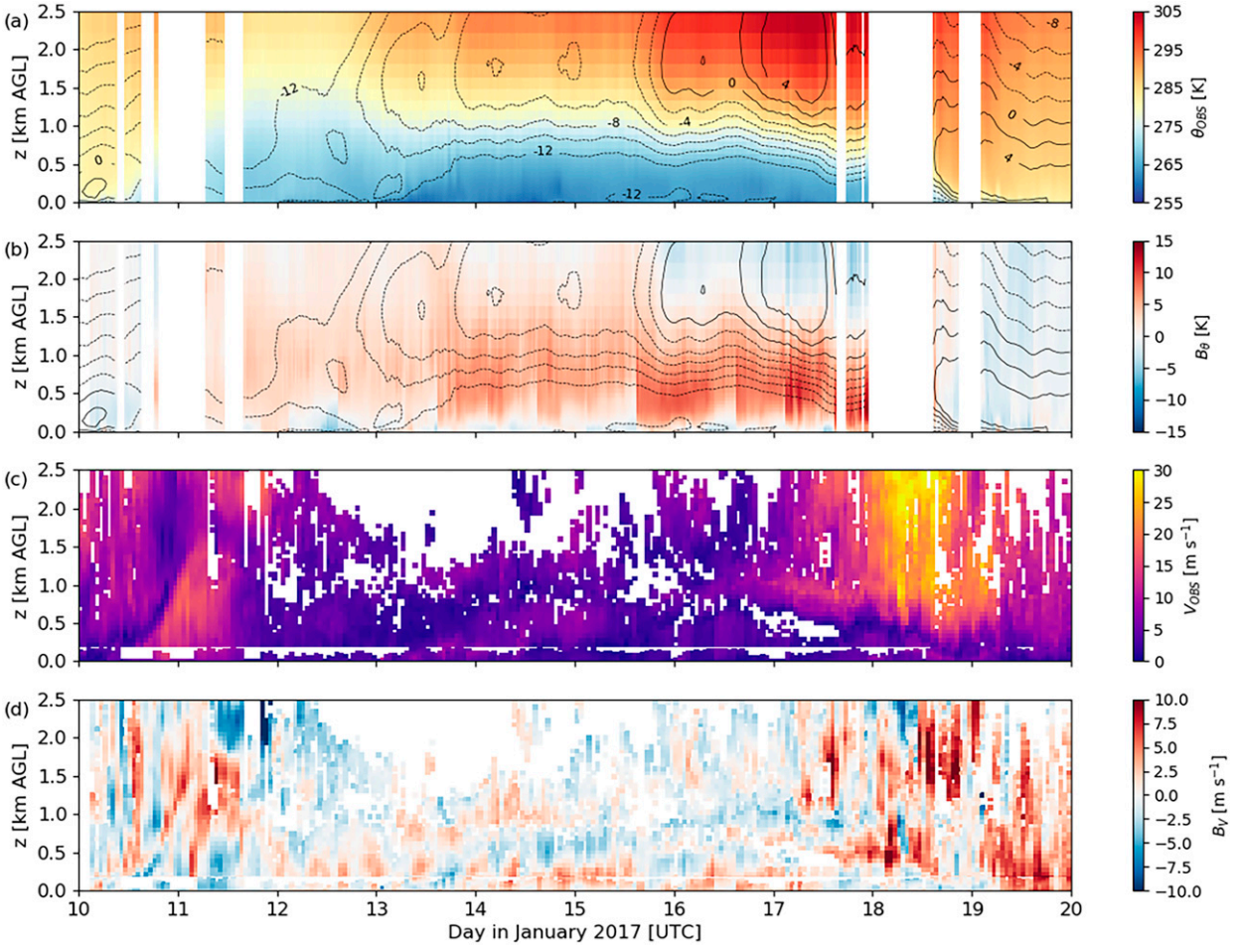

FIG. 2. Observed (a) potential temperature and (c) wind speed at Wasco for 10-day cold-air pool case study. (b),(d) Model bias [Eq. (2)] results for case 3, d02 (with the 3D PBL scheme), for the first 12 h of each reforecast (i.e., hours 3–15). Contours of observed temperature (°C) are included in (a) and (b) for reference.

Arthur et al. 2021) have demonstrated model improvement when this scheme is used with WRF's diff_opt = 2. With this option, horizontal gradients are calculated "in physical space" using local metric terms to help account for grid skewness. This setup is tested in case 2, using the same PBL and horizontal diffusion schemes as case 1 but with diff_opt = 2.

The 3D PBL scheme is tested in case 3. On d01, because resolved horizontal gradients are relatively small on the Δx = 3 km grid, the assumption of horizontal homogeneity is reasonable. Thus, this case uses the same d01 setup as case 2. On d02, however, both vertical and horizontal turbulent fluxes are treated using the 3D PBL scheme. Note that horizontal gradients of turbulent fluxes within the 3D PBL scheme are calculated in physical space, similarly to diff_opt = 2.

When calculating the turbulent flux tensor within the 3D PBL scheme, the boundary layer approximation noted in section 1b is used to reduce computational cost over the 10-day case study period and to ensure numerical stability.

TABLE 1. Turbulence and horizontal diffusion treatments used in 10-day cold-air pool reforecast cases.

	d01			d02		
Case	Vertical turbulence	Horizontal turbulence	Horizontal diffusion	Vertical turbulence	Horizontal turbulence	Horizontal diffusion
1	MYNN PBL	2D Smagorinsky	Along coordinate surfaces	MYNN PBL	2D Smagorinsky	Along coordinate surfaces
2	MYNN PBL	2D Smagorinsky	In physical space	MYNN PBL	2D Smagorinsky	In physical space
3	MYNN PBL	2D Smagorinsky	In physical space	3D PBL	3D PBL	In physical space

Use of the full matrix solution was attempted but found to be numerically unstable for the present case study. This is likely related to the turbulent length-scale formulation within the model (discussed further below), which may have limited applicability in stable atmospheric conditions over complex terrain. Continued development of the 3D PBL scheme should enable the full matrix solution to be evaluated for this case study and is thus a topic of future work.

It is important to note that both the MYNN and 3D PBL schemes used here are dependent upon several parameters defined within the Mellor–Yamada framework. These include a master length scale [i.e., ℓ in Eq. (1)], which is calculated diagnostically, as well as closure constants that relate the master length scale to other length scales used in the model. In this study, the standard MYNN level 2.5 model available in WRF is used, with the master length scale defined as a harmonic average of the surface, center-of-mass, and buoyancy length scales [see Eqs. (52)–(55) in Nakanishi and Niino 2009] and modified closure constants [see Eq. (66) in Nakanishi and Niino 2009].

For the 3D PBL scheme, the standard formulation of Mellor and Yamada (1982) is followed, with a center-of-mass approach (Mellor and Yamada 1974) applied using the Blackadar (1962) interpolation formula and the original Mellor and Yamada (1982) closure constants [see Eqs. (14)–(16) in Juliano et al. 2022]. Because this formulation is based on neutral conditions, its performance may be limited in a stable case study. However, because the development of improved length-scale formulations within the 3D PBL scheme is a topic of ongoing work (see discussion in Juliano et al. 2022), the standard formulation is used here. Sensitivities to the above parameter choices will be discussed in section 4b and are further analyzed in the appendix.

b. Model setup

The provisional WFIP2 HRRR and HRRRNEST domains, shown in Fig. 1, are initialized every 12 h between 0000 UTC 10 January 2017 and 1200 UTC 19 January 2017. A cold-start procedure is followed such that d01 is initialized using the operational Rapid Refresh (RAP) model without additional data assimilation or antecedent cycling (see Olson et al. 2019). The reforecasts are run for 24 h with one-way nesting and 3 h of spinup time on d01 prior to the initialization of d02.

WRF input, boundary, and namelist files for each reforecast period are available through the DAP (A2e 2017o). However, the basic grid configuration is summarized here. Domain d01 includes $N_x \times N_y = 889 \times 889$ grid cells in the horizontal direction; with $\Delta x = 3$ km, this gives a total domain size of $L_x \times L_y = 2667$ km $\times$ 2667 km. Domain d02 includes $N_x \times N_y = 1000 \times 900$ grid cells in the horizontal direction; with $\Delta x = 750$ m, this gives a total domain size of $L_x \times L_y = 750$ km $\times$ 675 km.

Both domains use the same vertical coordinate with $N_z = 50$ levels and $\Delta z \approx 15$ m for the first grid cell above the surface. The vertical grid spacing is stretched above the surface, as detailed in Benjamin et al. (2016), with a domain top of roughly 25 km. Note that on WRF's staggered grid, temperature and velocities are calculated at vertical half levels, which are denoted by dots in vertical profile figures that follow. The first half level (the lowest level at which temperature and velocities are calculated) is located at roughly 7.5 m AGL.

Beyond the turbulence and horizontal diffusion options discussed in section 2a, several additional changes were made to the cited namelists for the present study. Specifically, for compatibility with the 3D PBL scheme, the revised MM5 surface layer scheme (sf_sfclay_physics = 1) is used instead of the MYNN scheme (sf_sfclay_physics = 5). For consistency, this option is used for all cases on all domains. Note that for the purposes of this study, minimal differences were found between simulations using these two surface layer schemes.

Additionally, on both domains, WRF's option to add positive-definite sixth-order horizontal diffusion (diff_6th_opt = 2) is used with a factor of 0.25. However, in order to prevent overdiffusion in regions of sloping terrain, where numerical diffusion is expected to be relatively large, diff_6th_slopeopt is set to 1 with a threshold value of diff_6th_thresh = 0.05. Thus, sixth-order diffusion is linearly damped between slopes of 0 and 0.05 (2.86°) and shut off for larger slopes. At the observation sites considered here, the model terrain slope is generally below this threshold, although it reaches roughly 16° in the region surrounding the basin.

Otherwise, all cases use modules from the HRRR physics suite, as in Olson et al. (2019). These include the Rapid Update Cycle (RUC) land surface model (sf_surface_physics = 3), the Thompson aerosol-aware microphysics scheme (mp_physics = 28; Thompson and Eidhammer 2014), and the RRTMG radiation schemes (ra_sw/lw_physics = 4; Iacono et al. 2008). Following the WFIP2 control configuration, WRF's standard terrain-following vertical coordinate system (hybrid_opt = 0) is used for all cases as well. Although a hybrid sigma-pressure coordinate is available in WRF as of version 4, it is not expected to have a large effect on the results of the present study; similar grid skewness is found near the surface regardless of which vertical coordinate system is used (see, e.g., Arthur et al. 2021).

3. Model error evaluation

Model errors are evaluated using observations from the WFIP2 field campaign (see Table 2), with analysis focused first on vertical potential temperature and wind speed profiles during the 10-day event. Wind speed was measured at several sites listed in Table 2, including Wasco, Boardman, Condon, and Prineville, Oregon; and Goldendale, Washington (see locations in Fig. 1). Each site contained a radar wind profiler (RWP), measuring between roughly 100 m and 2.0–2.5 km AGL during the investigated period; a sodar, measuring between roughly 30 and 200 m AGL; and a surface station, measuring at 10 m AGL. Although each of these sites also included a radio acoustic sounding system (RASS) measuring vertical profiles of virtual temperature, RASS performance was limited during CAP conditions (see discussion in Adler et al. 2021). Temperature comparisons therefore focus on the Wasco site, where the availability of a microwave radiometer (MWR) helped to overcome the limited vertical range of the RASS. The temperature data used here for model comparison

TABLE 2. Observation datasets from the WFIP2 field campaign used for model bias evaluation. Most datasets have been quality controlled (denoted "b0" in the DAP code). Note that the raw (denoted "00" in the DAP code) Wasco MWR data are cited here because the dataset used in this work was reprocessed by Adler et al. (2021). In addition, only the quality controlled Gordon's Ridge lidar data are available through the DAP; the raw lidar data of Bodini et al. (2019), used to calculate TKE in the present study, are not currently available. See the data availability statement for additional information.

Site	Instrument	Variable used	Vertical range (AGL)	DAP code	Citation
Wasco	MWR	θ; T	21 m–15.5 km	mwr.z03.00	A2e (2017i)
	RASS	θ; T	142–1581 m	radar.z04.b0	A2e (2017l)
	RWP	V	81–2542 m	radar.z04.b0	A2e (2017l)
	Sodar	V	30–200 m	sodar.z09.b0	A2e (2017r)
	Surface station	V	10 m	met.z06.b0	A2e (2017c)
Boardman	RWP	V	124–2156 m	radar.z07.b0	A2e (2017n)
	Sodar	V	30–200 m	sodar.z16.b0	A2e (2017t)
	Surface station	V	10 m	met.z12.b0	A2e (2017d)
Condon	RWP	V	138–2511 m	radar.z06.b0	A2e (2017m)
	Sodar	V	10–250 m	sodar.z13.b0	A2e (2017s)
	Surface station	V	10 m	met.z33.b0	A2e (2017g)
Goldendale	RWP	V	94–2555 m	radar.z02.b0	A2e (2017j)
	Sodar	V	30–200 m	sodar.z03.b0	A2e (2017p)
	Surface station	V	10 m	met.z25.b0	A2e (2017f)
Prineville	RWP	V	138–2280 m	radar.z03.b0	A2e (2017k)
	Sodar	V	30–200 m	sodar.z04.b0	A2e (2017q)
	Surface station	V	10 m	met.z34.b0	A2e (2017h)
Gordon's Ridge	Profiling lidar	V; TKE	40–260 m	lidar.z02.b0	A2e (2017a); Bodini et al. (2019)
Physics site	Surface station	TKE	3 m	met.z21.b0	A2e (2017e)

were processed using the aforementioned TROPoe algorithm that combines RASS, MWR, and surface station data to obtain an optimal thermodynamic profile (Adler et al. 2021). Note that for instruments other than surface stations, the vertical ranges listed in Table 2 represent the maximum possible based on instrument settings. As seen in Fig. 2, the actual range varies with meteorological conditions and is often less than the maximum.

The model bias B_ϕ is defined as

$$B_\phi = \phi_{\mathrm{WRF}} - \phi_{\mathrm{OBS}}, \tag{2}$$

where ϕ is the meteorological variable, either potential temperature θ or wind speed V. Thus, a positive bias indicates an overestimate by the model, whereas a negative bias indicates an underestimate. To calculate the bias, model variables are interpolated to the latitude/longitude and vertical levels of the observations, with nearest-neighbor interpolation in the horizontal followed by linear interpolation in the vertical. The reader is referred to Pichugina et al. (2020) for a discussion of interpolation methods used for model bias calculations within the WFIP2 study region. Here, potential temperature bias is calculated at 15-min intervals (corresponding to the frequency of the optimal observed temperature profiles calculated by Adler et al. 2021), while wind speed bias is calculated using hourly averages (corresponding to the RWP data, noting that higher-frequency sodar and surface wind data are hourly averaged to match the RWP).

The 10-day event is summarized in Fig. 2 using vertical profiles of observed potential temperature (Fig. 2a) and wind speed (Fig. 2c) at Wasco. Following relatively high winds and temperatures on 10–12 January 2017, a CAP forms on 13 January. The CAP persists for roughly 6 days, through the end of 18 January. Starting on 16 January, warm air and relatively high wind speeds aloft begin to penetrate into the basin. This, along with subsequent cold-air advection aloft, contribute to the gradual erosion of the CAP, with the strongest erosion on 17–18 January. By the end of 19 January, the CAP is fully eroded.

Figures 2b and 2d also highlight important model biases. Most notably, there is a warm bias during the persistent CAP period above roughly 200 m AGL, with some instances of cold bias near the surface. This is associated with a predominantly positive wind speed bias within the CAP, although some negative bias is also seen between roughly 0.5–1.0 km AGL. Toward the end of the event, the model tends to overestimate both the high wind speeds aloft and their descent into the basin, affecting the timing of the modeled CAP erosion. These biases, which generally lead to overpredictions of wind turbine power output during the persistent CAP, as well as mistimed ramp-up of wind power following CAP erosion, have been noted as particular challenges for mesoscale wind energy forecasting in this region (Wilczak et al. 2019; Olson et al. 2019). Addressing these challenges is a focus of WFIP2 model development efforts (summarized in Olson et al. 2019) and therefore strongly motivates the present work.

Bias results in Fig. 2 are shown for case 3, d02, which uses the 3D PBL scheme. However, it is important to note that these features are relatively consistent across cases and domains. This is because CAP dynamics are highly dependent on model parameterizations for processes such

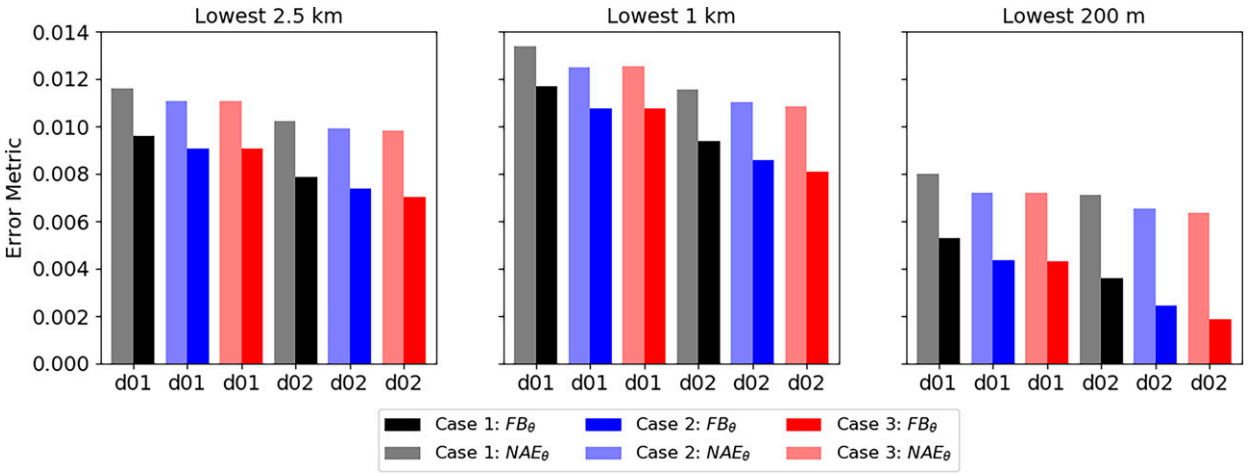

FIG. 3. Event-averaged potential temperature error metrics at Wasco for each case and domain shown in Table 1. Results are shown for the (left) lowest 2.5 km AGL, (middle) lowest 1 km AGL, and (right) lowest 200 m AGL. The first 12 h of each reforecast (i.e., hours 3–15) are used in the calculation. Note that because cases 2 and 3 have the same setup on d01 their error metrics on d01 are identical.

as radiation and land surface fluxes, which are not varied between cases. For example, additional spinup of the meteorology and/or land surface model (see, e.g., Jerez et al. 2020) could improve cold pool predictions overall but is not explored here. In the model evaluation that follows, differences in performance among the three tested cases are driven by the treatment of horizontal variability and turbulence.

a. *Event-averaged error*

The overall performance of the model is evaluated using two event-averaged error metrics (following e.g., Chang and Hanna 2004; Smith et al. 2018; Wiersema et al. 2020), the fractional bias

$$\mathrm{FB}_\phi = \frac{\overline{B_\phi}}{0.5\left(\overline{\phi_{\mathrm{WRF}}} + \overline{\phi_{\mathrm{OBS}}}\right)}, \tag{3}$$

and the normalized absolute error

$$\mathrm{NAE}_\phi = \frac{\overline{|B_\phi|}}{0.5\left(\overline{\phi_{\mathrm{WRF}}} + \overline{\phi_{\mathrm{OBS}}}\right)}. \tag{4}$$

where the overbar denotes an average over all available observations. FB_ϕ captures whether the model, on average, over or underestimates the observed values, while NAE_ϕ captures the average magnitude of the difference between the modeled and observed values. Because reforecasts begin every 12 h, with 3 h of spinup time on d01 prior to initializing d02, hours 3–15 of each reforecast are used in the calculation of the error metrics. This allows the results from both domains to be compared over the same time window. Note that the bias calculation is cut off at 0000 UTC 20 January 2017, such that only hours 3–12 are used from the 1200 UTC 19 January 2017 reforecast.

The event-averaged error metrics are shown in Figs. 3 and 4 for each case and domain over three vertical regions, focusing on the boundary layer. The lowest 2.5 km AGL captures the full range of the RWP instruments; the lowest 1 km AGL, used by McCaffrey et al. (2019) to identify CAPs in the WFIP2 dataset, generally captures the full depth of the CAP in the present case study; the lowest 200 m AGL, which roughly aligns with the range of the sodar instruments, captures the region of interest for wind energy forecasts. These regions are represented in the model by roughly 13, 9, and 4 vertical levels, respectively. Note that because cases 2 and 3 have the same setup on d01, their error metrics on d01 are identical.

Improved representation of horizontal variability and turbulence in the model leads to a reduction in the potential temperature bias at Wasco during the 10-day event (Fig. 3). In an event-averaged sense, a gradual reduction in the warm bias is seen both as the resolution is increased (from d01 to d02) and as more advanced turbulent diffusion treatments are used. On d02 specifically, the bias reduction from case 1 to case 2 demonstrates the effect of changing WRF's diff_opt = 1, which does not account for grid skewness in the horizontal diffusion calculation, to diff_opt = 2, which does. Further bias reduction from case 2 to case 3 demonstrates the effect of the 3D PBL scheme, which calculates horizontal turbulent fluxes explicitly while also applying a different turbulent length-scale formulation.

Related gains in model performance are seen in wind speed predictions (Fig. 4). As for the potential temperature, both increased resolution and more advanced turbulent diffusion treatments tend to improve model performance. In general at each site, and in an average across all sites (see bottom row of Fig. 4), gradual reductions in the positive wind speed bias are seen on d02 from case 1 to case 2, and from case 2 to case 3. The trends in wind speed bias at Boardman, however, differ from those at other sites (see second row of Fig. 4). This may be due to the relative flatness of the terrain there, as seen in

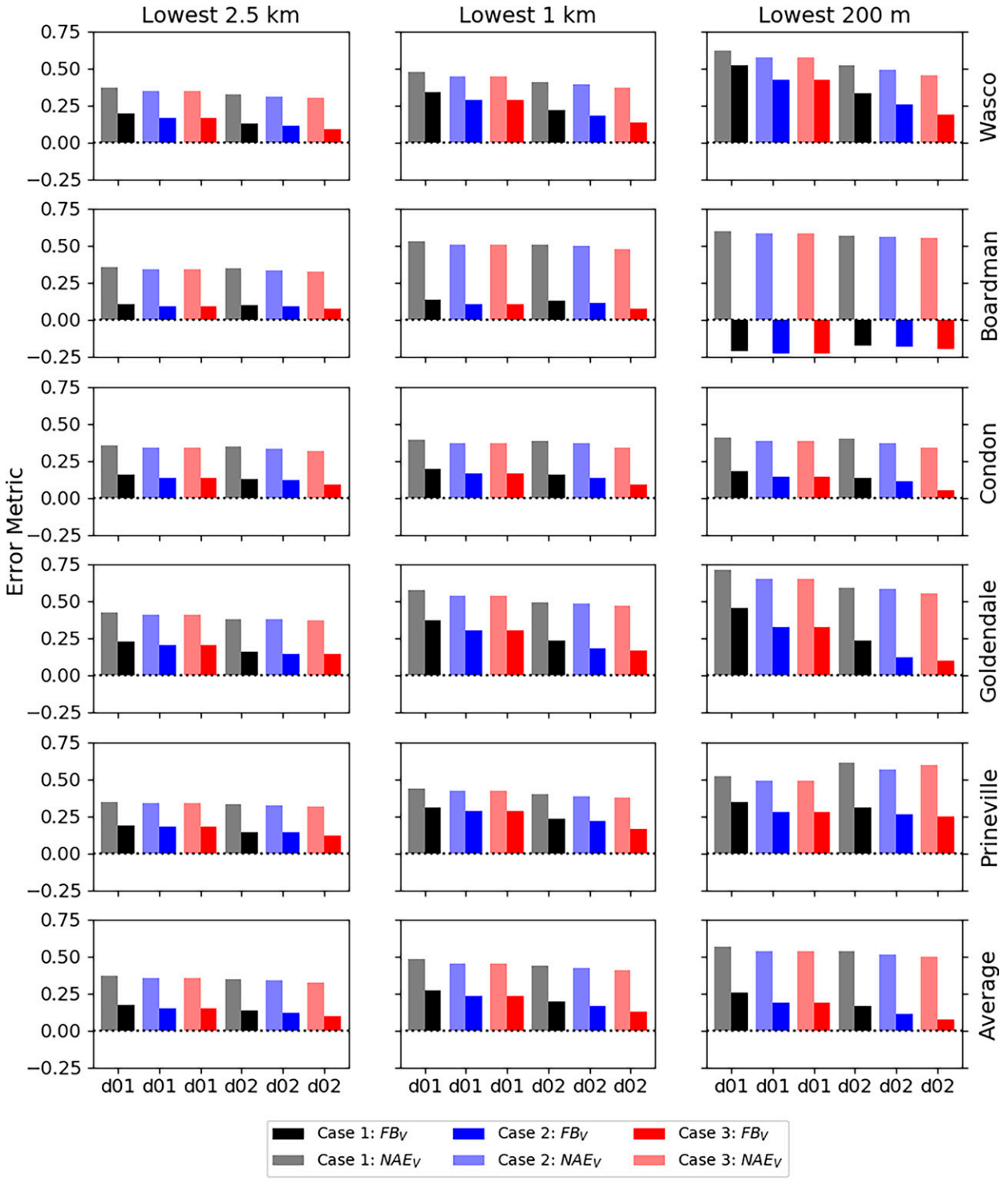

FIG. 4. Event-averaged wind speed error metrics for each case and domain shown in Table 1. Results are shown for the (left) lowest 2.5 km AGL, (middle) lowest 1 km AGL, and (right) lowest 200 m AGL; each wind speed observation site is shown individually, followed by an average over all five sites. The first 12 h of each reforecast (i.e., hours 3–15) are used in the calculation. Note that because cases 2 and 3 have the same setup on d01 their error metrics on d01 are identical.

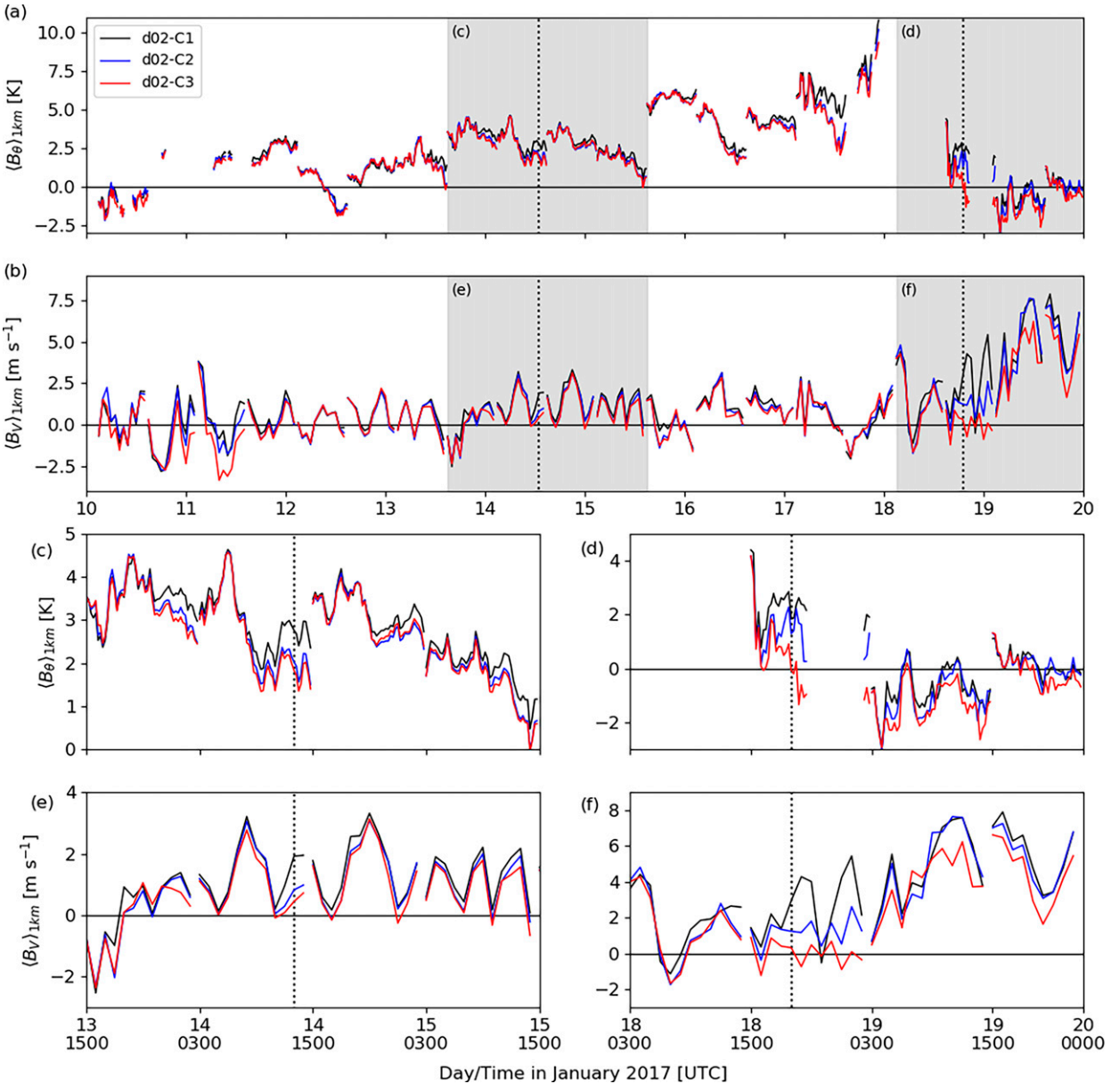

FIG. 5. Time-dependent (a) potential temperature and (b) wind speed bias at Wasco for each case on d02. Results are vertically averaged over the lowest 1 km AGL. Dotted lines denote the times of the profiles shown in Fig. 6, below. (c)–(f) A zoomed-in view of the gray-shaded regions in (a) and (b).

Fig. 1, which could limit the influence of horizontal variability relative to other physical processes.

Overall, the results in Figs. 3 and 4 show a robust reduction in event-averaged model error as the treatment of horizontal variability and turbulence is improved. In an average over all sites, the most advanced treatment (case 3, d02, using the 3D PBL scheme) leads to the smallest error. In particular, the positive wind speed bias during the event is reduced by a factor of roughly 2–3 from the simplest treatment (case 1, d01). The reduction in absolute error follows the same trend but is more modest because directionality is not included in the calculation. In what follows, this performance improvement is examined in more detail, through calculations of time-dependent bias and vertical profile analysis (section 3b), as well as turbulence comparisons (section 4).

b. *Time-dependent bias*

As the atmospheric dynamics evolve over course of the extended CAP event, model performance varies. This variation is summarized in Fig. 5, which shows the time-dependent potential temperature and wind speed bias at Wasco over the 10-day study period. Two shorter time windows are also shown, highlighting the cold pool and erosion periods. At the

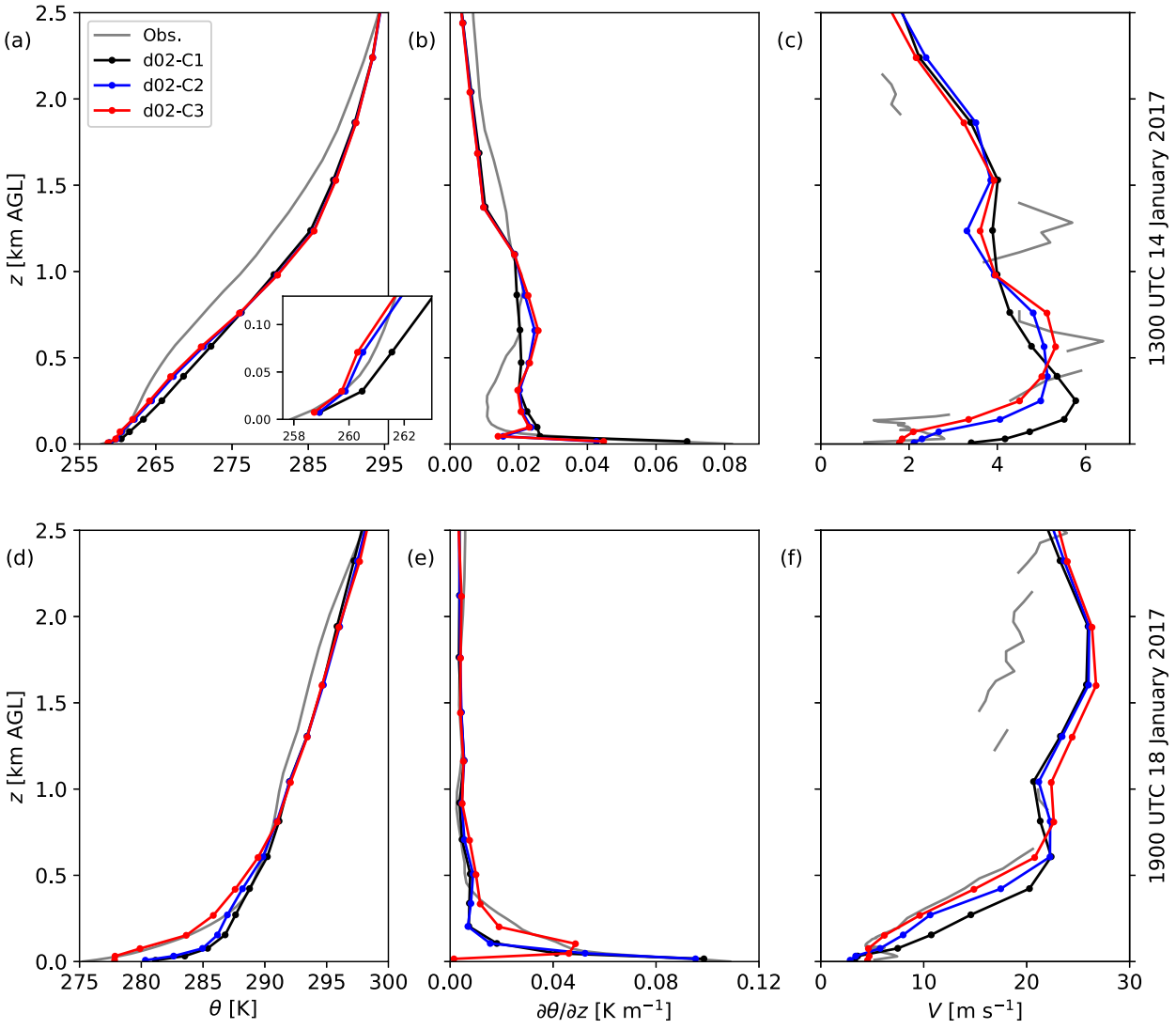

FIG. 6. Vertical profiles of potential temperature θ, the vertical potential temperature gradient $\partial\theta/\partial z$, and wind speed V for each case on d02 in comparison with observations at Wasco. The times shown, (a)–(c) 1300 UTC 14 Jan 2017 and (d)–(f) 1900 UTC 18 Jan 2017, are denoted by vertical dotted lines in Fig. 5. Potential temperature profiles are hourly averaged to match the velocity data.

time of each available observation, the model bias is vertically averaged over the bottom 1 km AGL, denoted $\langle B_\phi \rangle_{1\text{km}}$.

Results in Fig. 5 are shown for d02, which captures more horizontal variability, and thus greater differences among the three model treatments. As in the event-averaged error calculations above, the first 12 h of model data (hours 3–15) from each reforecast are used in the calculation. For this reason, there are discontinuities in the time series as the model results shift from one reforecast to the next. These discontinuities occur (here and in other time series figures) because each reforecast is initialized independently in a cold-start configuration, rather than as a restart from the previous end state.

During the cold pool portion of the event (roughly 13–15 January 2017, highlighted in Figs. 5c,e), the potential temperature bias is generally between 0 and 5 K. The corresponding wind speed bias is generally between 0 and 3 m s^{-1}, with some instances of negative bias. Small increases in model performance are seen as the horizontal diffusion and turbulence treatments are improved. Differences between cases are generally less than 1 K or 1 m s^{-1} but are occasionally larger. This modest improvement is expected during relatively steady, low wind conditions, during which cold pool dynamics are highly dependent on model physics, especially the surface layer parameterization, which does not vary between cases. The model improvement captured here is likely a result of reduced numerical mixing, which helps to preserve the cold pool.

Model performance during the cold pool portion of the event is further evaluated using example vertical profiles of potential temperature and its vertical gradient, as well as wind speed, in comparison with observations at Wasco (Figs. 6a–c). While the observed surface temperature is

roughly captured in each case, the model generally overpredicts atmospheric stability within the bottom 1 km AGL (Figs. 6a,b). This discrepancy is likely related the presence of low-level clouds, which enhance mixing between the surface and the cloud level (Adler et al. 2021, see Fig. 9 therein), but are not well represented in the model. Despite this limitation, cases 2 and 3 still show a reduction in the warm bias of up to several kelvins within the CAP (Fig. 6a). The example profiles also show evidence of near-surface cold bias, as seen in Fig. 2b, and underestimated stability for cases 2 and 3. This will be discussed in the context of modeled TKE in section 4a.

Overall, improvements in the modeled temperature profile lead to more accurate wind speed predictions at this time during the cold pool period (Fig. 6c). Case 1 shows a positive wind speed bias within the CAP and a negative wind speed bias just above it. Both biases are reduced by several m s^{-1} in cases 2 and 3, with case 3 showing the best agreement with observations.

Model improvements are accentuated during the erosion portion of the event, when turbulent mixing becomes a more dominant forcing mechanism. Larger differences between the three cases begin to arise on 16 January and increase over the next several days as erosion intensifies (Fig. 5). The largest consistent improvements in the modeled wind speed are seen from the end of 18 January through 19 January (see Figs. 5d,f), when high wind speeds from aloft penetrate closest to the surface.

Example profiles of potential temperature, its vertical gradient, and velocity are shown during the erosion period in Figs. 6d–f. The cold pool erodes early in cases 1 and 2, with increased temperatures between the surface and roughly 0.25 km. Using the 3D PBL scheme, case 3 maintains cooler temperatures in this region (Fig. 6d). Furthermore, although cases 1 and 2 capture the near-surface stability at the first grid level above the surface, they underestimate stability within most of the cold pool layer, up to roughly 0.5 km AGL (Fig. 6e). Case 3 more accurately captures the stability throughout most of the cold pool layer, despite underestimating the stability at the surface. The latter is due to a brief local change in the sign of the surface sensible heat flux in case 3 from negative (cooling) to positive (warming) at the time shown in Fig. 6.

As seen during the cold pool period, improvements in the modeled temperature profile lead to more accurate wind speed predictions at this time during the erosion period, most notably a reduction in the positive wind speed bias of as much as 5 m s^{-1} within the bottom 1 km AGL (see Fig. 6f). This reduction in near-surface wind speed bias is especially important for wind energy forecasting. Overall, the model improvements presented here could reduce the overestimation of available wind power during CAP conditions, while also leading to better predictions of the timing of cold pool erosion.

The results in this section also highlight the limitations of mesoscale models. For example, both of the observed velocity profiles in Fig. 6 contain a near-surface jet, which is a common feature of stable boundary layers. The dynamics of the jet have not been fully analyzed, but observations show a relatively consistent easterly direction and lack of diurnal variability, suggesting that it is driven by a large-scale zonal pressure gradient (for additional discussion of the pressure-driven flow, see Adler et al. 2021) rather than thermal/slope effects (as in, e.g., Grachev et al. 2016). Nevertheless, the model cases are not able to fully capture the jet, primarily because its vertical scale is too small to be resolved on the computational grid. As mentioned previously, model dependence on the surface layer parameterization also likely contributes to this issue. The accurate representation of small-scale features in stable boundary layers is a continued challenge for wind energy forecasting, and for mesoscale forecast models in general (e.g., Jia and Zhang 2020).

Additionally, all three model cases demonstrate relatively consistent bias values aloft (above roughly 1–2 km AGL). At these altitudes, the influence of complex terrain, and thus horizontal variability, becomes less important, and the model treatments discussed here have less of an effect on the dynamics.

4. Turbulence characterization

Based on the ability of the 3D PBL scheme to reduce potential temperature and wind speed bias during the 10-day case study, its effect on turbulence characterization is evaluated. Turbulence is generally difficult to predict in stable conditions because it can be weak and/or intermittent, posing a challenge for most turbulence closure models (Mahrt 1998; Zhou and Chow 2014; Chow et al. 2019). However, it is hypothesized that the 3D PBL scheme can better represent the impact of unresolved turbulence on the mean flow relative to a standard PBL approach.

Because both of the turbulence schemes considered here include a prognostic equation for unresolved TKE [Eq. (1)], it provides a useful metric for model comparison. Furthermore, observed profiles of TKE can be calculated from the profiling lidar instruments deployed during WFIP2 (see Fig. 7a). Here, the WindCube v2 profiling lidar at Gordon's Ridge is used with data at 20 m increments between 40 and 260 m AGL (see Table 2). This lidar measured line-of-site velocities along four cardinal-direction beams with a 28° zenith angle, as well as a fifth vertically staring central beam. The temporal resolution was 1.25 Hz along each beam direction, and a carrier-to-noise ratio of −22 dB was used to filter the raw radial velocity measurements.

The TKE is calculated from the lidar measurements as

$$\mathrm{TKE} = \frac{1}{2}\left(\langle u^2\rangle + \langle v^2\rangle + \langle w^2\rangle\right), \tag{5}$$

where angle brackets denote a time average and u, v, and w represent turbulent velocity fluctuations relative to that average. In particular, the horizontal wind vector is calculated at ~1 Hz using the raw radial velocity measurements from the four cardinal-direction beams and then averaged for 5 min. To calculate 5-min averaged TKE, the horizontal velocity covariances $\langle u^2\rangle$ and $\langle v^2\rangle$ are estimated using the ~1-Hz horizontal velocity components, while the vertical velocity covariance $\langle w^2\rangle$ is estimated at ~1 Hz using the vertical beam. Five-

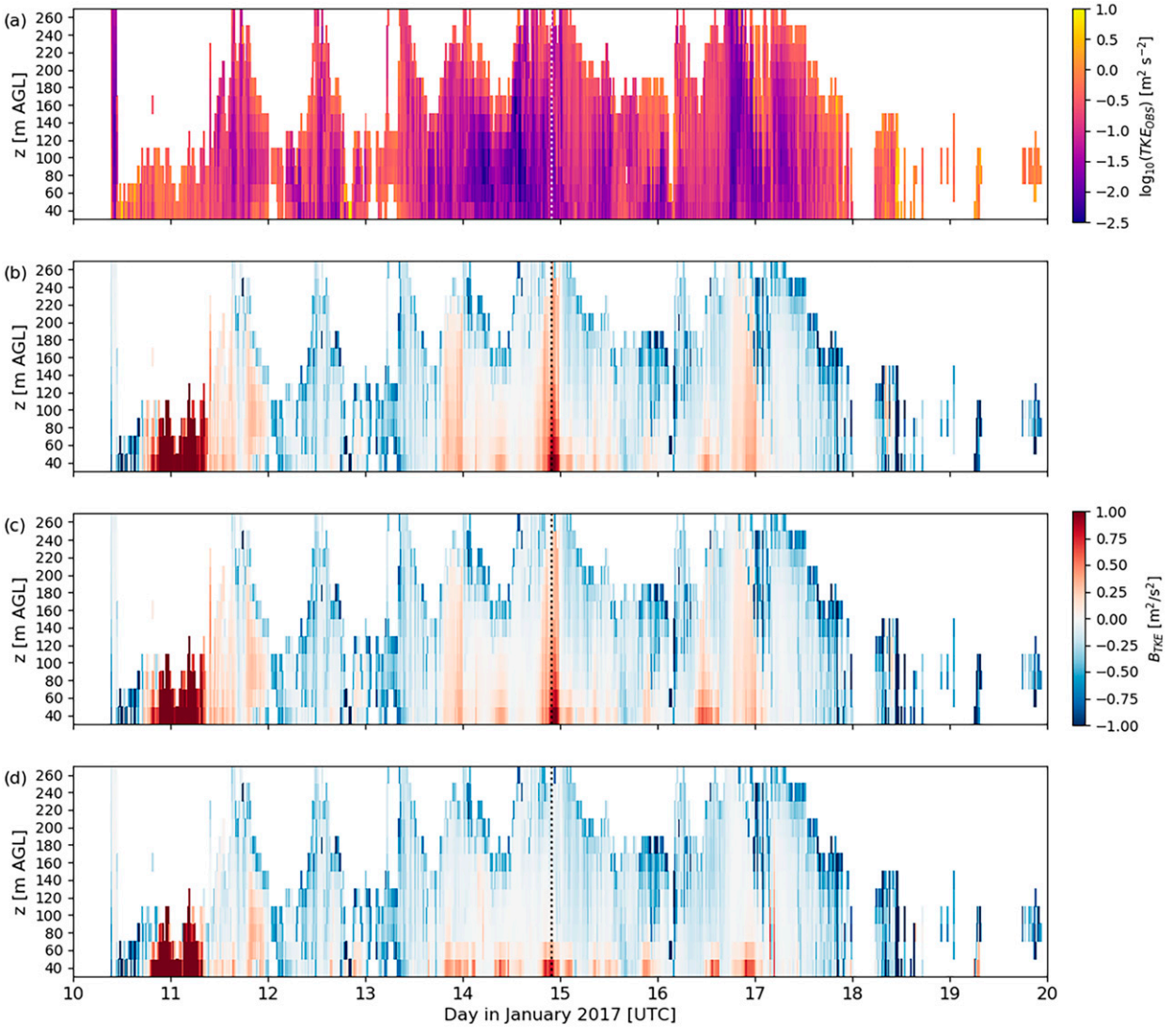

FIG. 7. Turbulence kinetic energy bias results for 10-day cold-air pool case study. (a) Observed TKE from the profiling lidar at Gordon's Ridge, shown on a $\log_{10}$ scale. Model bias [Eq. (2)] results for cases (b) 1, (c) 2, and (d) 3, shown on a linear scale. The bias is calculated for d02 using the first 12 h of each reforecast (i.e., hours 3–15). Vertical dotted lines denote the time of the profile shown in Fig. 10 below, i.e., 2200 UTC 14 Jan 2017.

minute averaging periods with less than 90% data availability are excluded. Note that a profiling lidar was also deployed at Wasco during WFIP2 but unfortunately did not collect data on 16–18 January 2017. Nearby Gordon's Ridge (see location in Fig. 1) is thus the best available profiling lidar site available for analyzing TKE in the present case study.

Under stable CAP conditions, during which the physical length scale of turbulent eddies is limited by buoyancy effects, the amount of resolved TKE in the model is assumed to be negligible. This is reasonable because the horizontal grid spacing (Δx = 750 m on d02) is generally greater than or equal to the boundary layer depth (roughly 0.5–1 km; see potential temperature profiles in Figs. 2a and 6a,c). Thus, in what follows, the parameterized TKE values calculated by the PBL schemes in the model are compared directly with observed TKE values.

a. *TKE bias*

The TKE bias for each model configuration in Table 1 is calculated relative to the observations at Gordon's Ridge (Figs. 7b–d). As in section 3, hours 3–15 of each reforecast are used to cover the 10-day case study period. The bias is calculated using Eq. (2) every 15 min (the output frequency of the model), and results are focused on d02 in order to examine the effect of the PBL scheme.

The model TKE bias is generally negative (blue colors in Figs. 7b–d), with some positive bias values occurring mostly near the surface (red colors in Figs. 7b–d). These near-surface

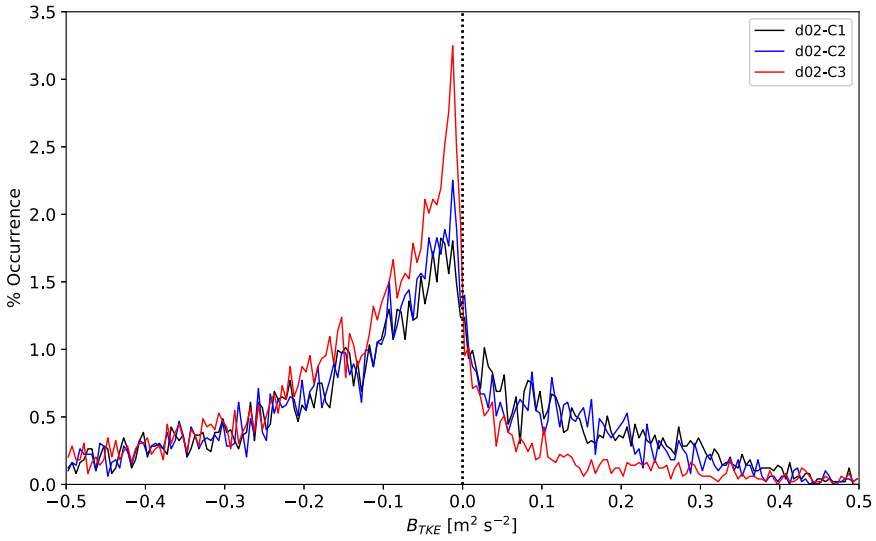

FIG. 8. A histogram of the turbulence kinetic energy bias values in Figs. 7b–d.

overestimates of TKE occur despite a cold bias within the vertical range of the lidar during the cold pool period (see Fig. 2b), which might be expected to suppress TKE (leading to a near-surface underestimate). However, there is also evidence of underestimated stability in this layer (see Figs. 6a,b), which could lead to the positive TKE bias seen here. Recall that temperature observations are not available at the Gordon's Ridge site, only at nearby Wasco. Outside of the cold pool period, there is evidence of large overestimates of TKE during the cold pool formation on 11 January, and underestimates during the erosion period after 17 January, although the lidar coverage was limited during these times.

Note that wind observations from profiling lidars can show differences relative to in situ measurements in complex terrain. This is generally an effect of volume averaging over the lidar cone (Bingöl et al. 2009) but depends on the surrounding terrain and atmospheric conditions. Analysis and mediation of these differences, for both mean and turbulence quantities, is an active area of research (e.g., Pauscher et al. 2016; Kelberlau and Mann 2020). However, the model TKE bias presented here is generally expected to be larger than the error in the lidar observations themselves (Bingöl et al. 2009, for example, found maximum wind speed errors on the order of 10%). Thus, this issue should not substantially change the results below.

The most striking results in the model TKE bias are the large overestimates of TKE that extend vertically throughout the lidar range for cases 1 and 2. These occur several times during the cold pool period of the event (14–17 January), and to a lesser extent during the cold pool formation on 11 January (see Figs. 7b,c). Moreover, TKE bias results for cases 1 and 2 are very similar, indicating that for this particular case and location, horizontal diffusion likely plays a minor role in the TKE prediction. Use of the 3D PBL scheme greatly reduces these overestimates, although large positive bias values can still occur between 40 and 60 m AGL (Fig. 7d). This improvement is likely related to differences between the MYNN and 3D PBL schemes, including the calculation of horizontal flux divergence terms, the length-scale formulation, and the model constants, and will be examined further in section 4b.

The TKE is highly variable over the case study period, ranging roughly thee orders of magnitude in the observations (Fig. 7a), and even more in the model. An event-averaged bias metric, as calculated previously for the potential temperature and wind speed, is therefore less informative for the TKE. Instead, a histogram (Fig. 8) is used to summarize the distribution of B_{TKE} over the 10-day event. As expected from Figs. 7b,c, the bias distributions for cases 1 and 2 are very similar. Furthermore, the histogram clearly shows a reduction in TKE overestimates when the 3D PBL scheme is used, with a corresponding increase in small TKE underestimates.

Model comparisons at a nearby surface station help to further explain the near-surface TKE overestimates seen at Gordon's Ridge. Figure 9 shows 2-min averaged turbulent velocity covariances, as well as the resulting TKE [as in Eq. (5)], observed at a tower located within the WFIP2 Physics site (see Table 2), which is near Wasco (see Fig. 1; note that a marker is not shown for the Physics site because it would essentially overlap the marker for Wasco). Since the 3D PBL scheme diagnoses the turbulent velocity covariances and includes them in the modeled flux divergence, model comparisons focus on d02 case 3, using data from the first 12 h of each reforecast (i.e., hours 3–15). Note that in the model, angle brackets in Eq. (5) denote an ensemble average, rather than an explicit time average (see Juliano et al. 2022).

Over the course of the 10-day event, the horizontal turbulent velocity covariances $\langle u^2 \rangle$ and $\langle v^2 \rangle$ are generally predicted within one order of magnitude (Figs. 9a,b). Overestimates tend to occur during the cold pool period (roughly 14–18 January), while the development and erosion periods show better agreement. Conversely, the vertical covariance $\langle w^2 \rangle$ is

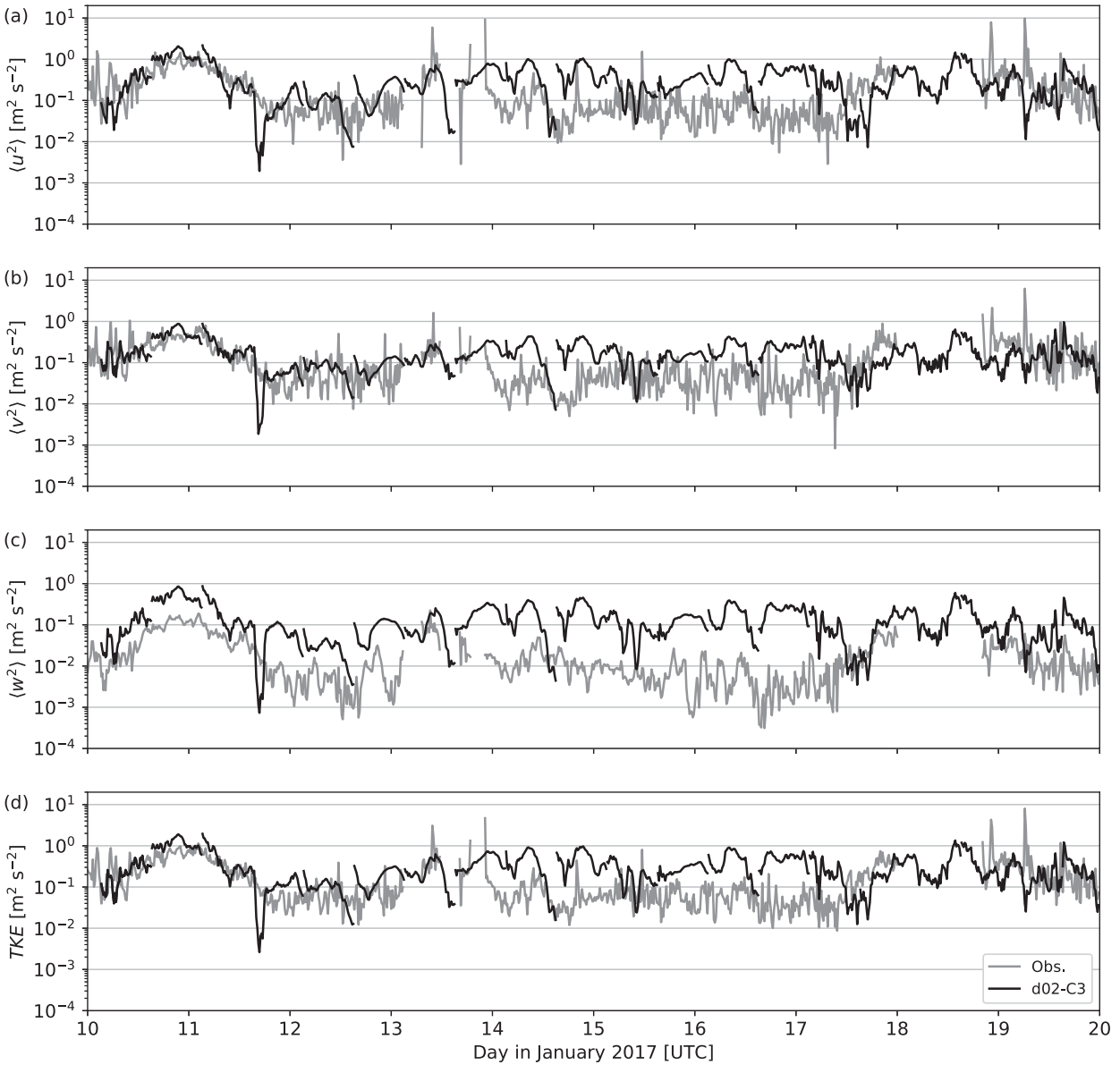

FIG. 9. Comparison of turbulent velocity covariances (a) $\langle u^2 \rangle$, (b) $\langle v^2 \rangle$, (c) $\langle w^2 \rangle$, and the resulting (d) TKE from a surface tower located at the WFIP2 Physics site (near Wasco; see Fig. 1). Observations are taken at 3 m AGL, whereas model values for d02 case 3, using the 3D PBL scheme, are taken at the surface and are shown during the first 12 h of each reforecast (i.e., hours 3–15).

overestimated throughout the case study, especially during the cold pool period (Fig. 9c). This overestimate of $\langle w^2 \rangle$ is likely related to reduced atmospheric stability in the model (see, e.g., Figs. 6a,b), which leads to larger vertical velocity fluctuations relative to the observations.

Because the horizontal covariances are generally larger than the vertical covariance, the resulting TKE comparison mostly reflects that of the horizontal components (Fig. 9d). However, the model's consistent overestimate of $\langle w^2 \rangle$ leads to a slight increase in the overestimate of TKE throughout the 10-day period. This highlights the limited ability of the 3D PBL model, as presently constructed, to represent turbulence anisotropy accurately. Scaling relationships for anisotropic turbulence in the atmospheric boundary layer (e.g., Stiperski and Calaf 2018; Stiperski et al. 2019) could be utilized to improve model performance in the future.

This analysis confirms that limiting overestimates of TKE, especially during stable conditions, is of primary importance for CAP forecasts. A reduction in TKE would, generally, lead to reduced vertical mixing and thus more pronounced decoupling between the surface layer and the flow aloft. This, in turn, would allow the model to maintain increased stability and lower boundary layer wind speeds, reducing the common temperature and wind speed biases discussed in section 3.

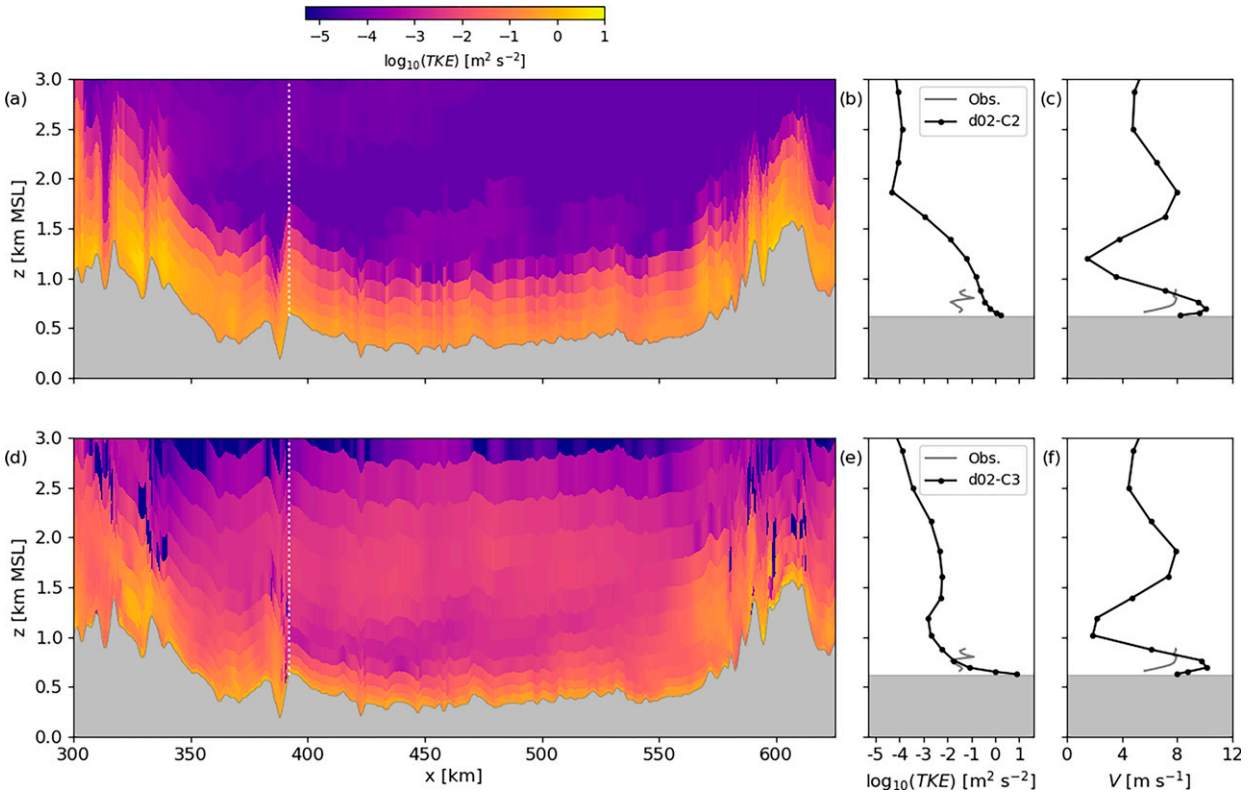

FIG. 10. Spatial variability of TKE on d02 for (top) case 2 and (bottom) case 3: (a),(d) Zonal slices of TKE through the Gordon's Ridge site, denoted by the vertical white dotted line. The transect is marked by a dashed line in Fig. 1. (b),(e) Vertical profiles of TKE in comparison with lidar observations, shown on a $\log_{10}$ scale. (c),(f) Vertical profiles of wind speed in comparison with lidar observations. Results are shown at 2200 UTC 14 Jan 2017, as denoted by the vertical dotted lines in Fig. 7.

b. *Spatial variability of turbulence*

A closer look at the spatial variability of TKE in the model helps to explain the differences between the turbulence and horizontal diffusion treatments used in this study. Figure 10 shows zonal slices of the modeled TKE field across the Columbia River basin (intersecting the Gordon's Ridge site), as well as profiles of TKE and wind speed in comparison with lidar observations. The time depicted is 2200 UTC 14 January 2017, as indicated by the dotted lines in Fig. 7. Because of the similarity in TKE between the 1D PBL cases 1 and 2, only case 2 is shown in comparison with case 3, which uses the 3D PBL scheme.

Despite the similarities in the wind speed prediction between cases 2 and 3 at this time (Figs. 10c,f), there are substantial differences in the TKE field. Most notably in case 2, the TKE increases monotonically from roughly 1–2 km AGL toward the surface, while maintaining a uniformly low value aloft (Figs. 10a,b). This is associated with an overestimate of the near-surface TKE (Fig. 10b), as compared with the lidar observation at Gordon's Ridge. There is also very little horizontal variability in TKE across the basin (Fig. 10a; $x \approx 350$–600 km).

On the other hand, case 3 displays better agreement with the observed near-surface TKE, as well as increased vertical and horizontal variability (Figs. 10d,e). Although the available TKE observations are limited in their spatial coverage, the increased TKE variability seen in the model is expected in this complex terrain region, especially in the presence of a strongly sheared wind profile with multiple jetlike features (see Figs. 10c,f as well as Fig. 6c, which shows a similar time during the cold pool period).

Analysis of the modeled TKE budget provides additional context for the difference in turbulence predictions between cases 2 and 3. Figure 11 shows the tendency terms from the prognostic TKE equation [Eq. (1)] at the Gordon's Ridge site on d02 at 2200 UTC 14 January 2017, thus corresponding to the TKE profiles shown in Figs. 10b and 10e. In particular, the shear production term $P_s = -\langle u_i u_j \rangle (\partial U_i / \partial x_j)$ can be viewed as a proxy for the parameterized vertical turbulent fluxes in the model, whereas $D_v = (\partial / \partial x_k)\left[\ell q S_q (\partial q^2 / \partial x_k)\right]$ represents vertical diffusion. Recall that in the present study, horizontal gradients are not included in Eq. (1).

The TKE budget highlights the differences in vertical mixing between cases 2 and 3. Shear production is elevated throughout the cold pool layer in case 2 (see black dashed line in Fig. 11), as is vertical diffusion (red dashed line), indicative of the "overmixing" phenomenon observed in CAP forecasts during WFIP2 that used the MYNN PBL scheme (see discussion in Wilczak et al. 2019). Conversely, in case 3, shear production is elevated only near the surface (see black

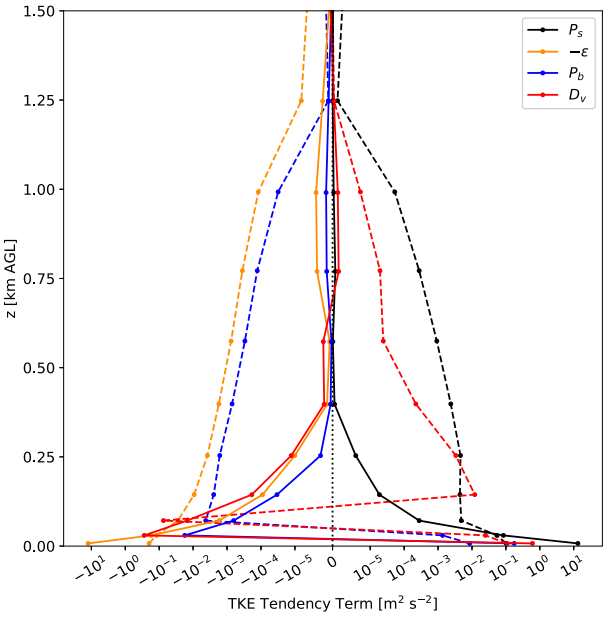

FIG. 11. TKE tendency terms for case 2 (dashed lines) and case 3 (solid lines). Results are shown at the Gordon's Ridge site on d02 at 2200 UTC 14 Jan 2017 and thus correspond to the TKE profiles shown in Figs. 10b and 10e. The plotted tendency terms correspond to those in Eq. (1) but are multiplied by the model time step Δt to give units of meters squared per second squared and are shown on a symmetric $\log_{10}$ scale that is linear between $\pm 10^{-5}$ m^2 s^{-2}; D_v represents the vertical diffusion of TKE, the first term on the right-hand side of Eq. (1).

solid line in Fig. 11), and vertical diffusion (red solid line) provides a much smaller positive contribution to the TKE aloft.

c. *Sensitivity of turbulence predictions*

For both cases 2 and 3, the vertical and horizontal TKE structure depicted in Fig. 10 is common during times of positive TKE bias (not shown), and therefore represents an important difference between the standard MYNN and 3D PBL schemes. It should be noted, however, that the TKE prediction in any Mellor–Yamada-based scheme is sensitive to the model parameters discussed in section 2a. With this in mind, an additional benefit of the 3D PBL scheme is that it provides a flexible framework for exploring model sensitivities.

Here, a brief sensitivity study is conducted using two new cases that are based on case 3 (see Table 1), but with modified parameters within the 3D PBL scheme. The first case (denoted case 3a) replaces the standard Mellor and Yamada (1982) length-scale formulation, which uses only a center-of-mass approach, with the formulation used by the MYNN scheme, which also includes a surface and buoyancy length scale (Nakanishi and Niino 2009). The second case (denoted case 3b) uses WRF's 2D Smagorinsky scheme for horizontal diffusion rather than horizontal fluxes calculated by the 3D PBL scheme. This leads to a more standard 1D PBL treatment, similar to the 1D case in Juliano et al. (2022). In both cases, all other model configuration options remain the same as in the original case 3.

Sensitivity results within the main text focus on TKE predictions during the cold pool period, using output from the 1200 UTC 14 January 2017 reforecast. In the appendix, this sensitivity analysis is extended to include potential temperature and wind speed bias results for several additional reforecasts, including 1200 UTC 14 January 2017, during the cold pool period, and 1200 UTC 18 January 2017, during the erosion period. These periods were chosen to highlight known sources of model variability during the 10-day case study, with a goal of informing future development and testing of the 3D PBL scheme.

The first sensitivity case demonstrates the importance of the length-scale formulation to the TKE calculation in the model. When the MYNN length scale is used within the 3D PBL scheme in case 3a, TKE is concentrated near the surface with very small values aloft. As in case 2, the TKE tends to increase (roughly) monotonically toward the surface and displays limited horizontal variability across the basin (see Figs. 12a–c, as compared with Figs. 10a–c). Note that there are differences between the PBL treatments in case 3a and case 2, including the Mellor–Yamada closure constants, other constants associated with the length-scale calculation, and the depth of integration used in the center-of-mass length scale.

It is also important to note that in both cases 2 and 3a, the TKE values above roughly 1–1.5 km AGL reach the minimum (floor) value specified in the code (see Figs. 10a,b and 12a,b, respectively). This occurs in case 3 as well, but only sporadically within the region of interest (see Figs. 10d,e). The TKE floor is defined in each PBL scheme independently; in the MYNN scheme, it is $q^2 = 10^{-4}$ (TKE $= 5 \times 10^{-5}$) m^2 s^{-2}, whereas in the 3D PBL scheme, it is $q^2 = 10^{-5}$ (TKE $= 5 \times 10^{-6}$) m^2 s^{-2}. Such TKE "clipping" is a known issue with TKE-based PBL schemes, and previous studies (e.g., Muñoz-Esparza et al. 2018) have lowered the floor value to improve model predictions during stable, low TKE conditions. However, this is not attempted here, because TKE values near the surface are generally well above the floor.

The second sensitivity case tests the effect of horizontal diffusion on the TKE predicted by the model. When horizontal diffusion within the 3D PBL scheme is replaced by 2D Smagorinsky-type diffusion in case 3b, the TKE field remains similar to that in the original case 3 (see Figs. 12d–f, as compared with Figs. 10d–f). This result, along with the TKE budget analysis above, suggests that in the present case, the treatment of horizontal diffusion does not play a large role in the TKE prediction. This is likely due in part to the use of the boundary layer approximation within the present 3D PBL formulation.

5. Summary and conclusions

This study has examined the effects of horizontal variability and turbulence treatment on mesoscale simulations of an extended cold-air pool event within the WRF Model. Model potential temperature and wind speed errors were shown to decrease both with increased horizontal grid resolution, and

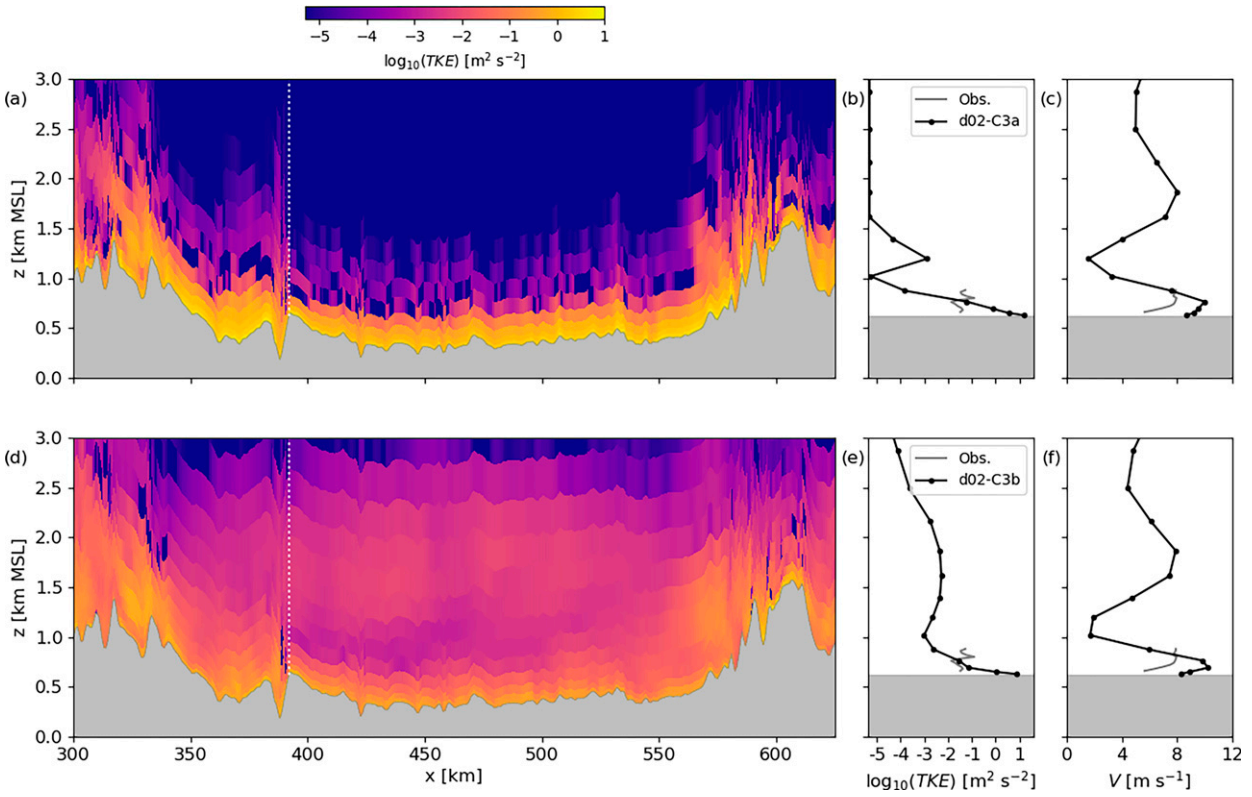

FIG. 12. As in Fig. 10, but for different options within the 3D PBL scheme. (a)–(c) Results from case 3a, using the MYNN length-scale formulation within the 3D PBL scheme. (d)–(f) Results from case 3b, using WRF's 2D Smagorinsky scheme for horizontal diffusion rather than horizontal fluxes calculated by the 3D PBL scheme. In both cases, all other model configuration options remain the same as in the original case 3. Results are shown at 2200 UTC 14 Jan 2017, as denoted by the vertical dotted lines in Fig. 7.

with a more advanced treatment of horizontal diffusion that uses local metric terms to account for grid skewness over steep terrain (WRF's diff_opt = 2 for calculating horizontal diffusion in physical space). Furthermore, the application of a new 3D PBL scheme following the Mellor–Yamada framework (Kosović et al. 2020; Juliano et al. 2022) was shown to reduce potential temperature and wind speed errors relative to a standard 1D PBL approach (WRF's MYNN level-2.5 scheme) with horizontal Smagorinsky-like diffusion. Model runs with the 3D PBL scheme resulted in the lowest overall bias values, with reductions of up to a factor of 2–3 relative to standard model options.

The model improvements tested here, including both advanced horizontal diffusion and modified turbulence treatment within the PBL scheme, lead to reduced numerical mixing in the boundary layer. This helps to preserve the near-surface inversion during the extended cold pool, leading to better predictions of potential temperature and wind speed. The effect is accentuated during the erosion of the cold pool, when turbulent mixing becomes a more dominant forcing mechanism. More accurate prediction of these dynamics represents an important step forward in mesoscale forecasting, especially wind energy forecasting, for which the persistence and erosion of cold-air pools have been noted as particular challenges (Wilczak et al. 2019; Olson et al. 2019).

In addition to potential temperature and wind speed, the 3D PBL scheme was shown to improve predictions of near-surface TKE. The TKE is important to predict accurately because it is used to estimate turbulent fluxes in the Mellor–Yamada framework, upon which both the MYNN and the 3D PBL schemes are based. Improved TKE predictions would also be useful in several numerical weather prediction applications, including Lagrangian particle dispersion modeling, in which TKE can be used to parameterize unresolved turbulent transport of pollutant particles (e.g., Brioude et al. 2013), or wind energy forecasting, in which TKE is an important component of mesoscale wind farm parameterizations (Fitch et al. 2012; Siedersleben et al. 2020).

The latter would be especially interesting to explore in the context of WFIP2 due to the presence of wind farms within the Columbia River basin study region, the effects of which were not included in the model runs considered here. Moreover, Redfern et al. (2019) updated WRF's mesoscale wind farm parameterization to include the rotor-equivalent wind speed rather than the hub-height wind speed. The impact of this change was largest for a test case based on a WFIP2 cold pool mix-out event.

In comparison with profiling lidar observations, the 3D PBL scheme limited overpredictions of near-surface TKE relative to the standard MYNN approach. The 3D PBL scheme

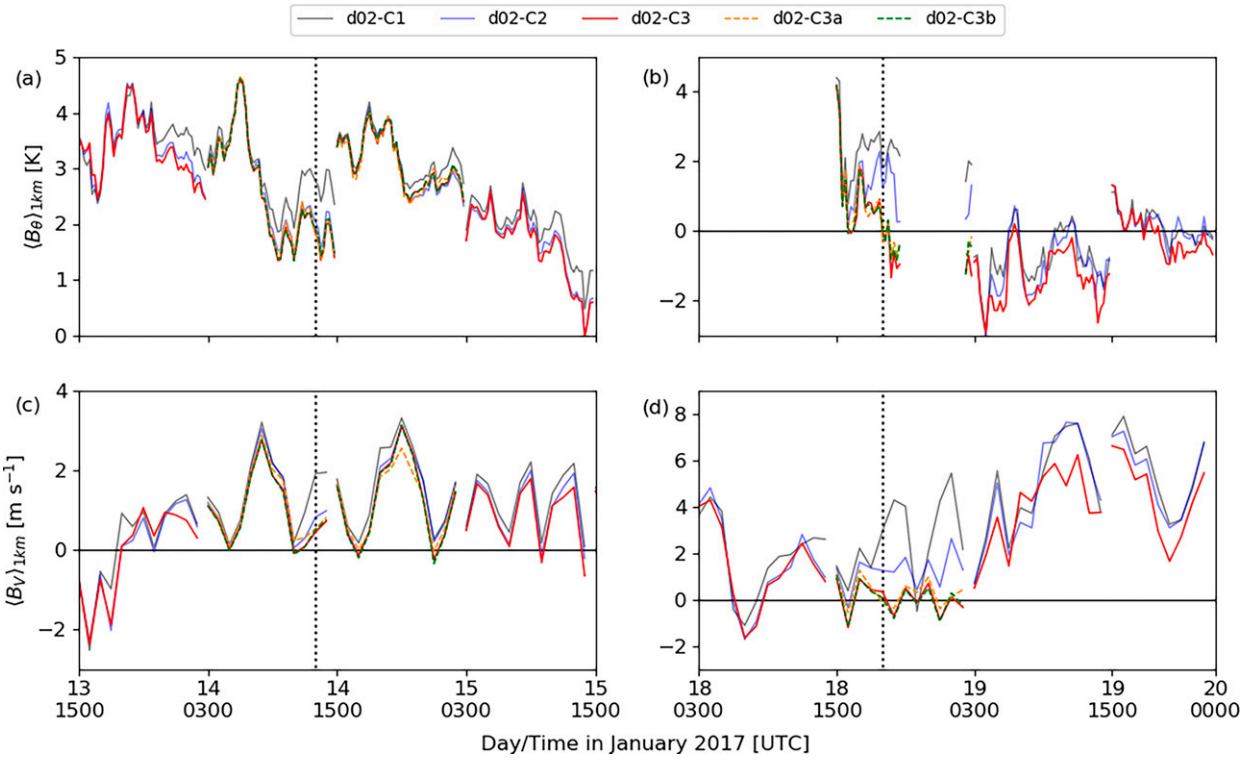

FIG. A1. As in Figs. 5c–f, but including sensitivity results from cases (left) 3a and (right) 3b. Zoomed-in views of (a),(b) potential temperature and (c),(d) wind speed bias time series are shown at Wasco for each case on d02. Results are vertically averaged over the lowest 1 km AGL. Vertical dotted lines denote the times of the profiles shown in Fig. A2.

also calculates the turbulent velocity covariances that make up the TKE, and these were compared with observations from a surface station. The modeled horizontal covariances were generally well predicted (within an order of magnitude), with more overestimates occurring during the cold pool period. The vertical covariance was overestimated throughout the case study, likely due to reduced stability in the model, highlighting the need to better capture anisotropic turbulent flows in the 3D PBL framework.

Modeled TKE predictions within the 3D PBL scheme were shown to vary with the master length-scale parameter, which is a component of any Mellor–Yamada-based turbulence parameterization. In the preliminary sensitivity study presented here, variation with the length scale was shown to outweigh variation due to the treatment of horizontal diffusion. In future work, the flexibility of the 3D PBL scheme will enable further investigation into the effect of model choices on turbulence characterization. For one, the model formulation could likely be improved for stable atmospheric conditions, perhaps by including scaling relationships for anisotropic turbulent flows (e.g., Stiperski and Calaf 2018; Stiperski et al. 2019). Initial development of the 3D PBL scheme has focused on convective conditions (see Juliano et al. 2022), and the present study is the first examination of its performance under stable conditions. Additionally, use of the full matrix solution for turbulent flux terms within the 3D PBL scheme, rather than the boundary layer approximation used here, will likely affect turbulence predictions and model bias.

As the 3D PBL framework is refined, it can be further developed and tested in realistic scenarios. For example, cloud prediction was a focus of WFIP2 model development efforts. While Adler et al. (2021) explored the role of low-level clouds on cold pool dynamics, these effects were not considered here. Future work could combine the subgrid-scale cloud parameterization (Chaboureau and Bechtold 2002, 2005) added to the MYNN scheme during WFIP2 (Olson et al. 2019) with the 3D PBL scheme for improved cold-air pool predictions. The performance of the 3D PBL scheme could also be explored for additional complex terrain flows, such as the gap flows and mountain waves that were consistently observed during WFIP2 (Wilczak et al. 2019).

Acknowledgments. This work was performed under the auspices of the U.S. Department of Energy (DOE) by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the DOE under Contract DE-AC36-08GO28308.

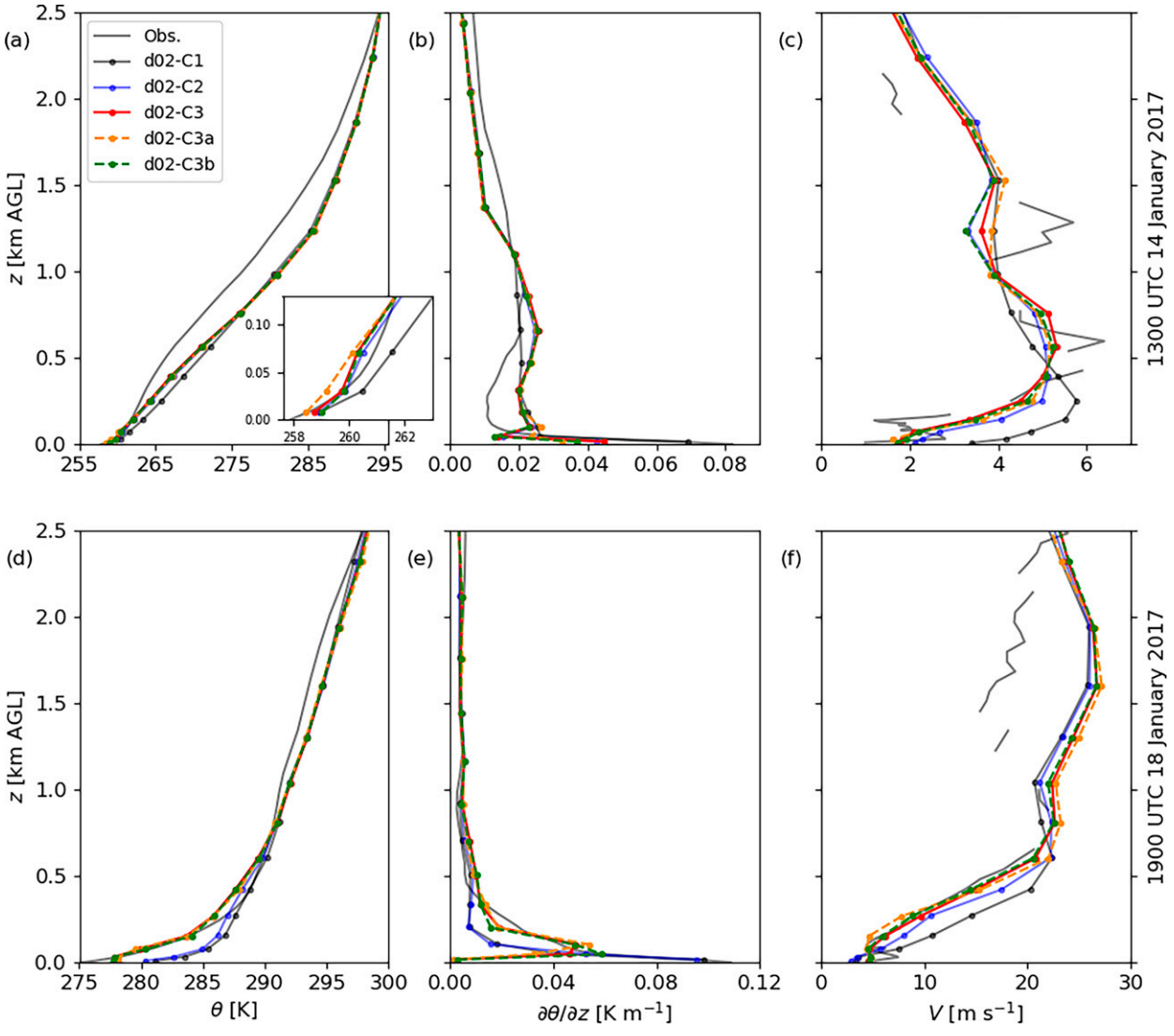

FIG. A2. As in Fig. 6, but including sensitivity results from cases 3a and 3b. Vertical profiles of potential temperature θ, the vertical potential temperature gradient $\partial\theta/\partial z$, and wind speed V are shown for each case on d02 in comparison with observations at Wasco. The times shown, (a)–(c) 1300 UTC 14 Jan 2017 and (d)–(f) 1900 UTC 18 Jan 2017, are denoted by vertical dotted lines in Fig. A1. Potential temperature profiles are hourly averaged to match the velocity data.

Funding was provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. government. The U.S. government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work or allow others to do so, for U.S. government purposes. Author Lundquist's effort was partially supported by the National Science Foundation under CAREER Grant AGS-1554055. Author Adler's effort was supported in part by the NOAA Cooperative Agreement with CIRES, NA17OAR4320101. This work was also supported in part by the NASA University Leadership Initiative (ULI) under Award 80NSSC20M0162. The National Center for Atmospheric Research is sponsored by the National Science Foundation. The authors thank Jim Wilczak for helpful discussions about the case study and three anonymous reviewers for their thoughtful comments and suggestions.

Data availability statement. All WRF input/boundary files and most observation datasets used in this work are publicly available through the U.S. Department of Energy Data Archive and Portal (DAP; http://a2e.energy.gov/data), as cited in the text. For additional information about the DAP, see Shaw et al. (2019). As mentioned in the caption of Table 2, only the quality controlled Gordon's Ridge lidar data are

available through the DAP. The raw data of Bodini et al. (2019), used to calculate TKE in the present work, are available by request.

APPENDIX

Sensitivity of Potential Temperature and Wind Speed Bias Results to 3D PBL Parameters

In section 4b, two new versions of case 3 (using the 3D PBL scheme; see Table 1) were introduced to explore the sensitivity of TKE predictions to certain model parameters. Here, the sensitivity analysis is extended to include potential temperature and wind speed bias results. To limit computational expense, only a subset of the reforecasts within the 10-day case study period are included: the 0000 and 1200 UTC 14 January 2017 reforecasts during the cold pool period and the 1200 UTC 18 January 2017 reforecast during the erosion period.

Recall that in case 3a, the standard Mellor and Yamada (1982) length-scale formulation, which uses only a center-of-mass approach, is replaced with the formulation used by the MYNN scheme, which also includes a surface and buoyancy length scale (Nakanishi and Niino 2009). In case 3b, WRF's 2D Smagorinsky scheme is used for horizontal diffusion rather than horizontal fluxes calculated by the 3D PBL scheme, leading to a more standard 1D PBL treatment. In both cases, all other model configuration options remain the same as in the original case 3.

Potential temperature and wind speed bias results are shown for the Wasco observation site in Figs. A1 and A2. Figure A1 shows the hourly bias within the bottom 1 km AGL and corresponds to Figs. 5c–f in the main text. Figure A2 shows vertical profiles during the cold pool period (1300 UTC 14 January 2017), as well as the erosion period (1900 UTC 18 January 2017), and corresponds to Fig. 6 in the main text. As seen in the TKE results in section 4b, the 3D PBL scheme shows more sensitivity to the length-scale formulation (case 3a, see dashed orange lines in Figs. A1 and A2), than to the horizontal diffusion treatment (case 3b, see dashed green lines in Figs. A1 and A2). However, variability in the potential temperature and wind speed bias between cases 3, 3a, and 3b is relatively minimal, and each generally performs as well or better than the standard MYNN PBL treatment in case 2. This suggests that the 3D PBL framework is robust to model parameter choices, although a wider range of sensitivity tests would be necessary to confirm this.

The similarity between cases 3 and 3b is likely related to the use of the boundary layer approximation within the 3D PBL scheme in the present work. However, it is also possible that the differences in horizontal fluxes calculated by the 2D Smagorinsky scheme and the 3D PBL scheme are small at the current model resolution (Δx = 750 m on d02), especially under stable conditions. As the numerical stability of the 3D PBL scheme with the full matrix solution is improved, the importance of horizontal fluxes can be explored more thoroughly under various atmospheric conditions and grid configurations.

REFERENCES

A2e, 2017a: wfip2/lidar.z02.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 1 May 2020, https://doi.org/10.21947/1410570.

——, 2017b: wfip2/log.z01.00. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 10 July 2019, https://doi.org/10.21947/1523403.

——, 2017c: wfip2/met.z06.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1412521.

——, 2017d: wfip2/met.z12.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1356316.

——, 2017e: wfip2/met.z21.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 22 February 2021, https://doi.org/10.21947/1523495.

——, 2017f: wfip2/met.z25.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1425897.

——, 2017g: wfip2/met.z33.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 12 October 2020, https://doi.org/10.21947/1401997.

——, 2017h: wfip2/met.z34.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 12 October 2020, https://doi.org/10.21947/1401998.

——, 2017i: wfip2/mwr.z03.00. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 2 June 2020, https://doi.org/10.21947/1412525.

——, 2017j: wfip2/radar.z02.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1412527.

——, 2017k: wfip2/radar.z03.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 12 October 2020, https://doi.org/10.21947/1412531.

——, 2017l: wfip2/radar.z04.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 16 July 2020, https://doi.org/10.21947/1412526.

——, 2017m: wfip2/radar.z06.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 12 October 2020, https://doi.org/10.21947/1412528.

——, 2017n: wfip2/radar.z07.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 16 July 2020, https://doi.org/10.21947/1412529.

——, 2017o: wfip2/refcst.coldstart.icbc.02. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 29 June 2020, https://doi.org/10.21947/1405131.

——, 2017p: wfip2/sodar.z03.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1491475.

——, 2017q: wfip2/sodar.z04.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 12 October 2020, https://doi.org/10.21947/1356329.

——, 2017r: wfip2/sodar.z09.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1356333.

——, 2017s: wfip2/sodar.z13.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 12 October 2020, https://doi.org/10.21947/1418162.

——, 2017t: wfip2/sodar.z16.b0. A2e Data Archive and Portal, Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy, accessed 17 July 2020, https://doi.org/10.21947/1356340.

Adler, B., J. W. Wilczak, L. Bianco, I. Djalalova, J. B. Duncan Jr., and D. Turner, 2021: Observational case study of a persistent cold pool and gap flow in the Columbia River basin. *J. Appl. Meteor. Climatol.*, **60**, 1071–1090, https://doi.org/10.1175/JAMC-D-21-0013.1.

Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *General Circulation Models of the Atmosphere*, J. Chang, Ed., Vol. 17, Academic Press, 173–265.

Arthur, R. S., K. A. Lundquist, D. J. Wiersema, J. Bao, and F. K. Chow, 2020: Evaluating implementations of the immersed boundary method in the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, **148**, 2087–2109, https://doi.org/10.1175/MWR-D-19-0219.1.

——, ——, and J. B. Olson, 2021: Improved prediction of cold-air pools in the weather research and forecasting model using a truly horizontal diffusion scheme for potential temperature. *Mon. Wea. Rev.*, **149**, 155–171, https://doi.org/10.1175/MWR-D-20-0234.1.

Bao, J., F. K. Chow, and K. A. Lundquist, 2018: Large-eddy simulation over complex terrain using an improved immersed boundary method in the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, **146**, 2781–2797, https://doi.org/10.1175/MWR-D-18-0067.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Billings, B. J., V. Grubišić, and R. D. Borys, 2006: Maintenance of a mountain valley cold pool: A numerical study. *Mon. Wea. Rev.*, **134**, 2266–2278, https://doi.org/10.1175/MWR3180.1.

Bingöl, F., J. Mann, and D. Foussekis, 2009: Conically scanning lidar error in complex terrain. *Meteor. Z.*, **18**, 189–195, https://doi.org/10.1127/0941-2948/2009/0368.

Blackadar, A. K., 1962: The vertical distribution of wind and turbulent exchange in a neutral atmosphere. *J. Geophys. Res.*, **67**, 3095–3102, https://doi.org/10.1029/JZ067i008p03095.

Bodini, N., J. Lundquist, R. Krishnamurthy, M. Pekour, L. Berg, and A. Choukulkar, 2019: Spatial and temporal variability of turbulence dissipation rate in complex terrain. *Atmos. Chem. Phys.*, **19**, 4367–4382, https://doi.org/10.5194/acp-19-4367-2019.

Brioude, J., and Coauthors, 2013: The Lagrangian particle dispersion model FLEXPART-WRF version 3.1. *Geosci. Model Dev.*, **6**, 1889–1904, https://doi.org/10.5194/gmd-6-1889-2013.

Chaboureau, 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.

——, 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.

Chachere, C. N., and Z. Pu, 2016: Connections between cold air pools and mountain valley fog events in Salt Lake City. *Pure Appl. Geophys.*, **173**, 3187–3196, https://doi.org/10.1007/s00024-016-1316-x.

Chang, J. C., and S. R. Hanna, 2004: Air quality model performance evaluation. *Meteor. Atmos. Phys.*, **87**, 167–196, https://doi.org/10.1007/s00703-003-0070-7.

Chemel, C., G. Arduini, C. Staquet, Y. Largeron, D. Legain, D. Tzanos, and A. Paci, 2016: Valley heat deficit as a bulk measure of wintertime particulate air pollution in the Arve River Valley. *Atmos. Environ.*, **128**, 208–215, https://doi.org/10.1016/j.atmosenv.2015.12.058.

Chow, F. K., C. Schär, N. Ban, K. A. Lundquist, L. Schlemmer, and X. Shi, 2019: Crossing multiple gray zones in the transition from mesoscale to microscale simulation over complex terrain. *Atmosphere*, **10**, 274, https://doi.org/10.3390/atmos10050274.

DeLeon, R., M. Sandusky, and I. Senocak, 2018: Simulations of turbulent flow over complex terrain using an immersed-boundary method. *Bound.-Layer Meteor.*, **167**, 399–420, https://doi.org/10.1007/s10546-018-0336-8.

Dempsey, D., and C. Davis, 1998: Error analysis and tests of pressure gradient force schemes in a nonhydrostatic, mesoscale model. *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 236–239.

Djalalova, I. V., D. Turner, L. Bianco, J. M. Wilczak, J. Duncan, B. Adler, and D. Gottas, 2022: Improving thermodynamic profile retrievals from microwave radiometers by including Radio Acoustic Sounding System (RASS) observations. *Atmos. Meas. Tech.*, **15**, 521–537, https://doi.org/10.5194/amt-15-521-2022.

Doubrawa, P., and D. Muñoz-Esparza, 2020: Simulating real atmospheric boundary layers at gray-zone resolutions: How do currently available turbulence parameterizations perform? *Atmosphere*, **11**, 345, https://doi.org/10.3390/atmos11040345.

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.

Francis, N., 2008: Predicting sudden changes in wind power generation. *North Amer. Windpower*, **5**, 58–60.

Gal-Chen, T., and R. C. J. Somerville, 1975: On the use of a coordinate transformation for the solution of the Navier-Stokes

equations. *J. Comput. Phys.*, **17**, 209–228, https://doi.org/10.1016/0021-9991(75)90037-6.

Grachev, A. A., L. S. Leo, S. Di Sabatino, H. J. S. Fernando, E. R. Pardyjak, and C. W. Fairall, 2016: Structure of turbulence in katabatic flows below and above the wind-speed maximum. *Bound.-Layer Meteor.*, **159**, 469–494, https://doi.org/10.1007/s10546-015-0034-8.

Holtslag, A. A. M., and Coauthors, 2013: Stable atmospheric boundary layers and diurnal cycles: Challenges for weather and climate models. *Bull. Amer. Meteor. Soc.*, **94**, 1691–1706, https://doi.org/10.1175/BAMS-D-11-00187.1.

Honnert, R., and V. Masson, 2014: What is the smallest physically acceptable scale for 1D turbulence schemes? *Front. Earth Sci.*, **2**, 27, https://doi.org/10.3389/feart.2014.00027.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Jerez, S., J. M. López-Romero, M. Turco, R. Lorente-Plazas, J. J. Gómez-Navarro, P. Jiménez-Guerrero, and J. P. Montávez, 2020: On the spin-up period in WRF simulations over Europe: Trade-offs between length and seasonality. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001945, https://doi.org/10.1029/2019MS001945.

Jia, W., and X. Zhang, 2020: The role of the planetary boundary layer parameterization schemes on the meteorological and aerosol pollution simulations: A review. *Atmos. Res.*, **239**, 104890, https://doi.org/10.1016/j.atmosres.2020.104890.

Juliano, T. W., B. Kosović, P. A. Jiménez, M. Eghdami, S. E. Haupt, and A. Martilli, 2022: "Gray zone" simulations using a three-dimensional planetary boundary layer parameterization in the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, https://doi.org/10.1175/MWR-D-21-0164.1, in press.

Kelberlau, F., and J. Mann, 2020: Cross-contamination effect on turbulence spectra from Doppler beam swinging wind lidar. *Wind Energy Sci.*, **5**, 519–541, https://doi.org/10.5194/wes-5-519-2020.

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.

Kosović, B., P. J. Munoz, T. W. Juliano, A. Martilli, M. Eghdami, A. P. Barros, and S. E. Haupt, 2020: Three-dimensional planetary boundary layer parameterization for high-resolution mesoscale simulations. *J. Phys. Conf. Ser.*, **1452**, 012080, https://doi.org/10.1088/1742-6596/1452/1/012080.

Lareau, N. P., E. Crosman, C. D. Whiteman, J. D. Horel, S. W. Hoch, W. O. J. Brown, and T. W. Horst, 2013: The Persistent Cold-Air Pool Study. *Bull. Amer. Meteor. Soc.*, **94**, 51–63, https://doi.org/10.1175/BAMS-D-11-00255.1.

Lundquist, K. A., F. K. Chow, and J. K. Lundquist, 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:AINAOT>2.0.CO;2.

Mahrt, L., 1998: Stratified atmospheric boundary layers and breakdown of models. *Theor. Comput. Fluid Dyn.*, **11**, 263–279, https://doi.org/10.1007/s001620050093.

Mazzaro, L. J., D. Muñoz-Esparza, J. K. Lundquist, and R. R. Linn, 2017: Nested mesoscale-to-LES modeling of the atmospheric boundary layer in the presence of underresolved convective structures. *J. Adv. Model. Earth Syst.*, **9**, 1795–1810, https://doi.org/10.1002/2017MS000912.

McCaffrey, K., and Coauthors, 2019: Identification and characterization of persistent cold pool events from temperature and wind profilers in the Columbia River basin. *J. Appl. Meteor. Climatol.*, **58**, 2533–2551, https://doi.org/10.1175/JAMC-D-19-0046.1.

Mellor, G. L., and T. Yamada, 1974: A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.*, **31**, 1791–1806, https://doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2.

——, and ——, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, https://doi.org/10.1029/RG020i004p00851.

Muñoz-Esparza, D., J. K. Lundquist, J. A. Sauer, B. Kosović, and R. R. Linn, 2017: Coupled mesoscale-LES modeling of a diurnal cycle during the CWEX-13 field campaign: From weather to boundary-layer eddies. *J. Adv. Model. Earth Syst.*, **9**, 1572–1594, https://doi.org/10.1002/2017MS000960.

——, R. D. Sharman, and J. K. Lundquist, 2018: Turbulence dissipation rate in the atmospheric boundary layer: Observations and WRF mesoscale modeling during the XPIA field campaign. *Mon. Wea. Rev.*, **146**, 351–371, https://doi.org/10.1175/MWR-D-17-0186.1.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

——, and ——, 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.

Olofson, K. F. G., P. U. Andersson, M. Hallquist, E. Ljungström, L. Tang, D. Chen, and J. B. Pettersson, 2009: Urban aerosol evolution and particle formation during wintertime temperature inversions. *Atmos. Environ.*, **43**, 340–346, https://doi.org/10.1016/j.atmosenv.2008.09.080.

Olson, J. B., and Coauthors, 2019: Improving wind energy forecasting through numerical weather prediction model development. *Bull. Amer. Meteor. Soc.*, **100**, 2201–2220, https://doi.org/10.1175/BAMS-D-18-0040.1.

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

Pichugina, Y. L., and Coauthors, 2019: Spatial variability of winds and HRRR–NCEP model error statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia River basin. *J. Appl. Meteor. Climatol.*, **58**, 1633–1656, https://doi.org/10.1175/JAMC-D-18-0244.1.

——, and Coauthors, 2020: Evaluating the WFIP2 updates to the HRRR model using scanning Doppler lidar measurements in

the complex terrain of the Columbia River Basin. *J. Renewable Sustainable Energy*, **12**, 043301, https://doi.org/10.1063/5.0009138.

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., O. 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.

Shaw, W. J., 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.

Siedersleben, S. K., and Coauthors, 2020: Turbulent kinetic energy over large offshore wind farms observed and simulated by the mesoscale model WRF (3.8.1). *Geosci. Model Dev.*, **13**, 249–268, https://doi.org/10.5194/gmd-13-249-2020.

Silcox, G. D., K. E. Kelly, E. T. Crosman, C. D. Whiteman, and B. L. Allen, 2012: Wintertime PM2.5 concentrations during persistent, multi-day cold-air pools in a mountain valley. *Atmos. Environ.*, **46**, 17–24, https://doi.org/10.1016/j.atmosenv.2011.10.041.

Simmons, A. J., and D. M. Burridge, 1981: An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, **109**, 758–766, https://doi.org/10.1175/1520-0493(1981)109<0758:AEAAMC>2.0.CO;2.

Skamarock, W. C., and Coauthors, 2019: A description of the Advanced Research WRF Model version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp., https://doi.org/10.5065/1dfh-6p97.

Smagorinsky, J., 1963: General circulation experiments with the primitive equations: I. The basic experiment. *Mon. Wea. Rev.*, **91**, 99–164, https://doi.org/10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2.

——, 1993: Some historical remarks on the use of nonlinear viscosities. *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, B. Galperin and S. A. Orszag, Eds., *Proceedings of an International Workshop in Large Eddy Simulation*, Cambridge University Press, 3–36.

Smith, E. N., J. A. Gibbs, E. Fedorovich, and P. M. Klein, 2018: WRF Model study of the Great Plains low-level jet: Effects of grid spacing and boundary layer parameterization. *J. Appl. Meteor. Climatol.*, **57**, 2375–2397, https://doi.org/10.1175/JAMC-D-17-0361.1.

Stiperski, I., and M. Calaf, 2018: Dependence of near-surface similarity scaling on the anisotropy of atmospheric turbulence. *Quart. J. Roy. Meteor. Soc.*, **144**, 641–657, https://doi.org/10.1002/qj.3224.

——, ——, and M. W. Rotach, 2019: Scaling, anisotropy, and complexity in near-surface atmospheric turbulence. *J. Geophys. Res. Atmos.*, **124**, 1428–1448, https://doi.org/10.1029/2018JD029383.

Thompson, T., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

Turner, D. D., and U. Löhnert, 2014: Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteor. Climatol.*, **53**, 752–771, https://doi.org/10.1175/JAMC-D-13-0126.1.

——, and W. G. Blumberg, 2019: Improvements to the AERIoe thermodynamic profile retrieval algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **12**, 1339–1354, https://doi.org/10.1109/JSTARS.2018.2874968.

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

——, S. W. Hoch, J. D. Horel, and A. Charland, 2014: Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. *Atmos. Environ.*, **94**, 742–753, https://doi.org/10.1016/j.atmosenv.2014.06.012.

Wiersema, D. J., K. A. Lundquist, and F. K. Chow, 2020: Mesoscale to microscale simulations over complex terrain with the immersed boundary method in the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, **148**, 577–595, https://doi.org/10.1175/MWR-D-19-0071.1.

Wilczak, J. M., and Coauthors, 2019: The Second Wind Forecast Improvement Project (WFIP2): Observational field campaign. *Bull. Amer. Meteor. Soc.*, **100**, 1701–1723, https://doi.org/10.1175/BAMS-D-18-0035.1.

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:TNMITT>2.0.CO;2.

Zängl, G., 2002: An improved method for computing horizontal diffusion in a sigma-coordinate model and its application to simulations over mountainous topography. *Mon. Wea. Rev.*, **130**, 1423–1432, https://doi.org/10.1175/1520-0493(2002)130<1423:AIMFCH>2.0.CO;2.

——, 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.

Zhou, B., and F. K. Chow, 2014: Nested large-eddy simulations of the intermittently turbulent stable atmospheric boundary layer over real terrain. *J. Atmos. Sci.*, **71**, 1021–1039, https://doi.org/10.1175/JAS-D-13-0168.1.