WRF Hub-Height Wind Forecast Sensitivity to PBL Scheme, Grid Length, and Initial Condition Choice in Complex Terrain

DAVID SIUTA, GREGORY WEST, AND ROLAND STULL

University of British Columbia, Vancouver, British Columbia, Canada

(Manuscript received 23 June 2016, in final form 9 December 2016)

ABSTRACT

This study evaluates the sensitivity of wind turbine hub-height wind speed forecasts to the planetary boundary layer (PBL) scheme, grid length, and initial condition selection in the Weather Research and Forecasting (WRF) Model over complex terrain. Eight PBL schemes available for the WRF-ARW dynamical core were tested with initial conditions sources from the North American Mesoscale (NAM) model and Global Forecast System (GFS) to produce short-term wind speed forecasts. The largest improvements in forecast accuracy primarily depended on the grid length or PBL scheme choice, although the most important factor varied by location, season, time of day, and bias-correction application. Aggregated over all locations, the Asymmetric Convective Model, version 2 (ACM2) PBL scheme provided the best forecast accuracy, particularly for the 12-km grid length. Other PBL schemes and grid lengths, however, did perform better than the ACM2 scheme for individual seasons or locations.

1. Introduction

Decreases in renewable energy costs, increased government incentives, and demand for cleaner energy sources have led to growth of the alternative energy sector. Renewable energy sources include hydroelectric, solar, and wind. Wind energy in particular has become more competitive because of a decrease in the cost of wind turbine material production and an increase in turbine efficiency. In addition, climate change mitigation strategies mandate increased reliance on renewables such as wind to decrease greenhouse gas emissions from the energy sector (IPCC 2014).

In 2008, the U.S. Department of Energy outlined a goal of obtaining 20% of its domestic energy supply from wind energy by 2030 (Department of Energy 2008; Greene et al. 2012). As of April 2016, the United States had an installed wind capacity of 72 GW and is on track to meet this goal. Twenty states achieved 5% of their energy from wind in 2015, with 12 of those states over 10%. Likewise, the Canadian Wind Energy Association established wind energy goals (Wind Vision 2025), aiming to obtain 20% of Canada’s domestic energy from wind energy by the year 2025 (Canadian Wind Energy Association 2008).

Wind energy generation is different from many traditional energy sources in that its availability depends on the presence of wind. Since total electric supply and demand must always remain in sync, one of the keys to successfully managing wind energy is the ability to accurately forecast the expected amount of wind energy supplied to the power grid. Energy planners rely on wind forecasts, often from numerical weather prediction (NWP) models, to predict wind energy at multiple forecast horizons. Forecasts are primarily used to plan spinning and standby reserves (DeMeo et al. 2007; Corbus et al. 2009) and market trading (Draxl et al. 2014). Spinning and standby reserves are traditional energy sources that are online (spinning) or ready to be turned on (standby) so that power can be adjusted to continually match electrical demand. These sources add flexibility to the electric grid in the sense that they are available to add power to the electric grid at short notice to account for the variable nature of winds, as well as for forecast uncertainty (DeMeo et al. 2007; Monteiro et al. 2009). Reserve resources require spinup time that can range from several hours to over a day, depending on the source (Ahlstrom et al. 2013). Thus, more accurate
short-term wind forecasts lead to more efficient reserve resource planning. These reserves, however, operate at a cost, and scheduling too many reserves can result in missed market trading opportunities (Clement 2012). For market trading, errors in short-term wind forecasts can result in losses from not being able to provide the correct amount of power. In some markets, traders are able to update market bids within the hour of expected delivery (Monteiro et al. 2009; Draxl et al. 2014). Because of the approximate cubic relationship between wind speed and associated wind power, even a small forecast error of 1 m s$^{-1}$ in a wind of 10 m s$^{-1}$ results in a 33% error in wind power (Cheng et al. 2013). Given the importance of the short-term forecast horizon, it is the focus of our research. It is expected, however, that our results will also prove useful for longer planning horizons.

Many methods exist to approximate hub-height wind speeds from 10-m winds. Typical methods are the 1/7 power law and the adiabatic profile (logarithmic law) (Peterson and Hennessey 1978; Stull 1988; Arya 2001), as well as the Radix profile (Santoso and Stull 1998, 2001). However, these methods are based on implicit assumptions about the boundary layer, including atmospheric stability, that are not always valid. Siuta (2013) showed for the 1999 Cooperative Atmospheric–Surface Exchange Study (CASES-99; Poulos et al. 2002) that the assumptions in these wind-profile laws can break down in differing boundary layer static stability regimes to produce large errors in the estimated hub-height wind.

NWP models, such as the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008), have taken the forefront in wind speed forecast research and can be used to directly forecast winds at hub height to avoid vertical interpolation. WRF has two dynamical cores: the Nonhydrostatic Mesoscale Model (NMM) and the Advanced Research version of WRF (ARW). The NMM core is used by the National Centers for Environmental Prediction (NCEP) North American Mesoscale Model (NAM), while the ARW core has been primarily used by academic researchers. Both cores have several physics and dynamics options that can be individually selected by the user, which can influence the hub-height wind speed forecast and lead to improvements over standard operational models from national meteorological centers. Siuta (2013) found WRF-ARW simulations over the Wyoming wind corridor (Martner and Marwitz 1982) to be superior to those from the operational NAM forecasts over the same area. In some cases the difference in estimated wind power density varied by several hundred watts per square meter, both seasonally and daily.

Several studies have recently been done to evaluate the effects of WRF planetary boundary layer (PBL) physics parameterization scheme choice on wind forecasts. Shin and Hong (2011) evaluated five PBL schemes for a single-day case study in the CASES-99 dataset and found larger differences and higher biases in 10-m wind forecasts produced by different PBL schemes during nighttime stable regimes. Deppe et al. (2013) evaluated six PBL schemes over Iowa at wind turbine hub height and found an ensemble of schemes to outperform the deterministic forecasts from any one PBL scheme. They also found the nonlocal-mixing PBL schemes (detailed in the next section) to have better accuracy.

Draxl et al. (2014) found biases over flat terrain to be strongly dependent on atmospheric static stability, with the nonlocal schemes performing best during unstable and near-neutral conditions, and a local scheme performing best during stable conditions. Draxl et al. (2014) also highlight the strong variation in the vertical wind shear between seven PBL schemes, indicating that evaluating PBL scheme performance on the 10-m winds alone is not sufficient when trying to use forecasts for hub-height winds.

Storm et al. (2009) evaluated two PBL schemes for simulating the Great Plains nocturnal low-level jet at two locations and found different positional and magnitude errors, depending on the scheme. Finally, Hahmann et al. (2015) determine that hub-height wind forecasts over sea are most sensitive to the choice of PBL scheme. While these studies provide useful results to the wind energy and NWP communities, they leave knowledge gaps in areas of complex terrain.

Jiménez and Dudhia (2012) showed WRF-ARW to have a surface (10 m) wind bias dependent on the terrain type, highlighting important deficiencies in wind forecasts over complex terrain. However, given the Draxl et al. (2014) findings, hub-height winds should be evaluated for wind energy purposes. Yang et al. (2013) provide one such study over the Columbia River basin for three PBL schemes, with the finding that each scheme allowed winds to be too fast during ramp events under stable conditions. Yang et al. (2016) find strong sensitivity to the adjustment of 26 parameters within the Mellor–Yamada–Nakanishi–Niino PBL scheme, but do not consider other schemes.

One of the primary problems with producing hub-height wind forecasts from NWP models over complex terrain is that none of the available PBL physics parameterization schemes (in WRF) were derived from data taken over complex terrain, which we conclude from reading the source publications cited in the next section. Instead, model physics that are responsible for evolving boundary layer processes were formulated mainly from field data over flat terrain, with a heavy focus on improving precipitation forecasts. Important aspects of boundary layer simulation over complex terrain include thermally induced circulations (mountain/valley and sea breezes), terrain-forced flow (gap and drainage flows),
very long roughness lengths, and mountain waves. Thus, the assumptions made in PBL schemes may not be appropriate for complex terrain, and their performance should be evaluated (Chow et al. 2013). This is especially true for hub-height wind forecasts where small improvements (around 10%–20% improvement in mean absolute error) can lead to millions of dollars in cost savings (Marquis et al. 2011). Because of these limitations, the improvement of hub-height wind forecasts in regions of complex terrain is the goal of the Wind Forecast Improvement Project 2 (http://www.esrl.noaa.gov/gsd/renewable/wfip2.html), which recently commenced field work.

The goal of the study presented here is to quantify the sensitivity of hub-height wind speed forecasts in complex terrain to the PBL scheme, initial condition, and grid-length selection in the WRF Model. Answering this question will fill a knowledge gap concerning the use of different PBL schemes over complex terrain. It will also be of use to those who need to create WRF Model hub-height wind forecasts over mountainous areas by addressing which factors are the most important to consider during the initial stages of the modeling process (e.g., the influences of initial condition, PBL scheme, and grid-length selection on forecast accuracy). The rest of this paper is organized as follows. Section 2 describes the methodology used, including the WRF Model setup and brief descriptions of the PBL schemes, initial conditions, and grid lengths tested. Section 3 provides results from the study period, followed by a discussion in section 4. Section 5 concludes with a summary of the findings and suggestions for future work.

2. Methodology

The WRF-ARW model was used to produce 48 forecasts per day for 1 yr (June 2014–May 2015) by using different initial condition sources, horizontal grid lengths, and PBL schemes (Fig. 1, Table 1). Eight available PBL schemes, detailed later in this section, were each used in 16 daily WRF Models. Eight runs were initialized by the 32-km NAM and another eight by the 0.5° Global Forecast System (GFS). Each of these runs contained domains

| Model detail                  | Setting                                                                 |
|------------------------------|-------------------------------------------------------------------------|
| WRF core                     | ARW version 3.5.1                                                       |
| Grid dimensions              | 36 km, 100 × 76; 12 km, 136 × 103; 4 km, 103 × 103                       |
| Land surface                 | Noah                                                                    |
| Microphysics                 | WRF single-moment 5-class                                               |
| Cumulus                      | Kain–Fritsch                                                           |
| Shortwave radiation          | Dudhia                                                                  |
| Longwave radiation           | Rapid Radiative Transfer Model                                          |
| Planetary boundary layer     | YSU, MRF, ACM2, MYJ, QNSE, MYNN, GBM, and UW                           |
| Surface layer                | MM5 similarity (used with YSU, ACM2, MRF, MYNN, GBM, and UW PBL schemes) |
|                              | or Eta similarity (MYJ PBL scheme only) or QNSE surface layer (QNSE PBL scheme only) |
of three different grid lengths, yielding the 48 total forecasts. The model forecast used four, two-way nested domains: a 36-km parent domain over western Canada, a 12-km nest over British Columbia and Alberta, and two 4-km nests containing wind farms of interest (Fig. 2). Forecasts covered the important 0–24-h forecast horizon, after allowing for 9 h of model spinup. With a 24-h forecast horizon, such forecasts would be used for same-day reserve scheduling and market trading.

Spinup time is important for higher-resolution local area models like WRF because observation networks and relatively coarse initial conditions do not contain enough information to describe typical small-scale features (e.g., tight gradients around fronts, terrain-induced local flows, and boundary layer circulations) (Warner 2011). Instead, WRF must develop, or spin up, these features with time. Thus, the output from the first few hours may contain unrealistic features and should be discarded. Spinup times generally range from 6 to 12 h (Warner 2011; Skamarock 2004), during which the output should not be used.

Vertical levels in the model were tuned to have fine resolution around wind turbine hub height with 62 vertical levels total and six levels in the lowest 100 m. Forecast data were extracted without horizontal interpolation (using the nearest neighbor) or vertical interpolation (using the model level closest to hub height) for each wind farm location to allow for differences to be attributed only to the PBL scheme. The authors recognize that interpolations can further improve the hub-height wind forecasts.

a. Planetary boundary layer schemes

There are 11 PBL schemes available for use in the WRF-ARW core, version 3.5.1. Of those, eight were tested here. Three schemes were not tested (Mellor–Yamada–Nakanishi–Niino level 3, total energy–mass flux, and Boulac schemes), because they caused WRF to crash, as well as computing resource constraints.

A PBL scheme is ultimately responsible for the simulated development of the planetary boundary layer, and for subgrid-scale fluxes of heat, momentum, and moisture by vertical turbulent transport from the first model level to the model top (Yang et al. 2013; Cheng et al. 2013; Shin and Hong 2011). Wind turbine hub heights and blade spans fall within the PBL during most times of the day, thus making the PBL scheme of crucial importance in understanding WRF strengths and deficiencies in simulating the hub-height wind. Each PBL scheme is generally compatible with at least one (but sometimes up to three) surface-layer scheme(s). The surface-layer scheme was kept constant (Table 1) for consistency when possible. For more information on how the PBL and surface layer are linked, see Skamarock et al. (2008). Following are brief explanations of the main components of each PBL scheme.

1) The Medium-Range Forecast Model (MRF) scheme is a first-order closure, nonlocal eddy-diffusivity (K theory) approach that uses a countergradient correction term based on the definition of the PBL top (Hong and Pan 1996). PBL top is defined by a critical value of the bulk Richardson number ($R_i = 0.5$). The countergradient correction term accounts for contributions to the total flux by large eddies that span the boundary layer, making this a nonlocal scheme.

2) The Yonsei University (YSU) scheme is a first-order closure, nonlocal eddy-diffusivity (K theory) approach with explicit treatment of the PBL top (Hong et al. 2006). This scheme provides an updated countergradient flux term originally included in the MRF scheme. Here, the PBL top is determined by the level at which minimum turbulent flux exists (heat, momentum, moisture). The YSU scheme was updated in WRF version 3.4 to account for a bug in the code in the calculation of the momentum diffusion coefficient (Hu et al. 2013). Hu et al. (2013) claims this improves wind speed estimation in stable regimes by decreasing vertical mixing and is one of the only PBL schemes to focus improvements in wind.

3) The Mellor–Yamada–Janjic´ (MYJ) scheme is a local mixing, 1.5-order closure, prognostic turbulent kinetic energy (TKE) scheme, used operationally in the NAM. The TKE vertical distribution is controlled via a master mixing length, which is used in the TKE budget equation (Janjić 1994). Local closure implies only small turbulent eddy contributions to the TKE distribution are accounted for.
4) The Asymmetric Convective Model, version 2 (ACM2) scheme is a combination of nonlocal transient turbulence (Stull 1993) scheme during unstable conditions with local eddy diffusivity in stable conditions (Pleim 2007). A stability parameter controls the contribution of the total mixing from nonlocal and local components. Downward mixing is controlled by local K theory. Pleim (2007) showed strong agreement between the ACM2 scheme and LES. The PBL height is determined by a critical bulk Richardson number.

5) The Quasi-Normal Scale Elimination (QNSE) scheme follows the same theory as the MYJ scheme in unstable and neutral conditions but develops a spectral theory during stable periods (Deppe et al. 2013). The goal of this scheme is to more accurately develop the boundary layer during stable regimes (Sukoriansky et al. 2005). This scheme treats turbulent eddies as waves during stable regimes.

6) The Mellor–Yamada–Nakanishi–Niino level 2.5 (MYNN) scheme is also a local mixing, 1.5-order closure, prognostic TKE scheme, and a slight variation of the MYJ scheme. TKE vertical distribution is controlled via a master mixing length that differs from that in the MYJ scheme. For the MYNN scheme, the master mixing length is the sum of three individual length scales: the surface-layer length, the turbulent-layer length, and the buoyancy length (Deppe et al. 2013).

7) The Grenier–Bretherton–McCaa (GBM) scheme was formed to try and improve representations of cloud-topped marine boundary layers (Grenier and Bretherton 2001). It uses moist thermodynamics in its turbulence scheme, unlike others that use unsaturated thermodynamics. The GBM scheme is also a 1.5-order closure local TKE prognosis scheme using K theory to parameterize turbulent fluxes.

8) The University of Washington (UW) scheme is a modified GBM scheme taking into account important cloud-top processes such as cloud-top cooling, evaporation, and phase change (Bretherton and Park 2009). This scheme uses moist thermodynamic variables as in GBM and slightly varies the master mixing-length formulation.

b. Initial condition sources

Initial conditions are used as the best estimate of the initial atmospheric state in a model simulation, whereas boundary conditions provide updates to the lateral boundaries of the outermost model domain during the simulation. In this paper we refer to both generally as initial conditions. In a similar study over uniform terrain, Deppe et al. (2013) show different initial condition sources do lead to different solutions for hub-height winds after progressing forward in time.

For this study the NCEP GFS and NAM models were used for the initial conditions. Each source was tested with all PBL schemes mentioned in section 2a. GFS-initialized WRF runs used the 0000 UTC GFS output at 0.5° resolution, while NAM-initialized WRF runs used 0000 UTC 32-km NAM output. The NAM contains many different grid lengths ranging from 4 to 32 km at the time of this writing. The master grid of 32-km grid length was used for this study because it was the only NAM grid that captured the full study area. The NAM (NMM dynamical core) differs from WRF-ARW primarily in its grid configuration. The NAM is a grid-box model operating on an Arakawa B-grid arrangement whereas the WRF-ARW core used in this study uses an Arakawa C-grid arrangement.

c. Verification metrics

Murphy and Winkler (1987) detail a systematic framework for forecast verification. Namely, the joint distribution of forecasts and observations should be considered, and the forecast user or provider should pinpoint which aspects of the forecast are most relevant for their use. Since wind energy is roughly proportional to the cube of the wind speed, errors are highly sensitive to differences between the forecast and observed wind speeds, and they need to be highly correlated. State-of-the-art forecasts should capture wind speed variations, as well as forecast uncertainty due to imperfect models and initial conditions.

This paper addresses short-term, deterministic forecast performance through the following metrics: mean absolute error (MAE), root-mean-square error (RMSE), and the linear correlation coefficient $r$. The best forecasts will have small MAEs (high accuracy), small RMSEs (small amounts of large forecast errors), and $r$ closest to 1.0 (forecasts and observations trend with each other). When gauging forecast performance of a large number of forecasts (in this case 48 forecasts), a summary plot such as a Taylor diagram (Taylor 2001) is useful. It shows the centered RMSE (CRMSE), $r$, and standard deviation (the latter indicating how well the standard deviation of the forecast values matches the standard deviation of the observations).

We also use the standardized anomaly ($z$ score) to measure which model configurations are better or worse than average. In addition, the MAE skill score (MAESS) is used to quantify the amount of improvement that can be gained using one configuration over another, and through the use of bias correction (detailed later).

Finally, an analysis of variance (ANOVA) was performed to estimate the importance of three factors.
(initial condition source, grid length, and PBL scheme) on the hub-height wind forecast accuracy. An ANOVA is a method of evaluating how much of the overall variance of a quantity (in this case, MAE) can be explained by any number of factors (Wilks 2011). We use the ANOVA to pinpoint which of these factors has the biggest potential for improving short-term, hub-height wind forecast accuracy over complex terrain for the wind farms studied here.

This paper does not address aspects of ensemble forecasting or forecast uncertainty, but the authors would like to mention their importance for wind energy forecasts. Forecast uncertainty is typically derived from an ensemble prediction system, where multiple models are run producing a distribution of possible forecast solutions. The resulting forecast probability distribution provides an indication of forecast uncertainty. We discuss this aspect of the study in Siuta et al. (2017).

d. Observations

Hub-height wind speed observations used to calculate verification metrics in this paper are the confidential property of the wind farm owners. The authors have taken careful consideration to not reveal wind farm locations and observed values. We refer to the wind farm sites as stations numbered 1–4. Hub heights ranged from 80 to 95 m, depending on the wind site. All wind farms were located on mountain ridgetops with elevations ranging from approximately 500 to 1000 m above sea level. We use wind farm–averaged observations to compare with model output. Wind farm–averaged observations are derived from nacelle-based observations. The raw, hub-height wind observations from the wind farms were quality controlled by BC Hydro (a utility company independent of the private wind farm operators) before being made available to the authors.

e. Bias correction

We apply a degree of mass balance (DMB), or multiplicative, bias-correction scheme detailed in Bourdin et al. (2014). This statistical postprocessing scheme is suitable for wind forecasts because it does not allow for negative wind speeds. Bias correction effectively removes the systematic component of error, but is only possible by having access to real-time wind farm observations. A follow-up paper (Siuta et al. 2017) discusses this bias-correction method and the effects of bias correction on ensemble and probabilistic wind forecasts. We found that a running 30-day training period for bias correction provided the maximum increase in forecast skill over raw forecasts, when compared with other training periods ranging from 1 to 30 days (not shown here).

This bias-correction technique implicitly accounts for average wind turbine wake effects. However, it is possible that this method does not successfully differentiate between waked and unwaked flows, so some wake-induced biases still exist under specific regimes. We did not use the wind turbine drag parameterization scheme in WRF, which is not compatible with all the PBL schemes used here.

3. Results and discussion

Given the complexity of evaluating 48 forecasts, our discussion will be split into two subsections. First, we detail the deterministic performance of each configuration for the bias-corrected forecasts. Second, we discuss the sensitivity of the bias-corrected forecast accuracy to each of the three factors (initial condition source, grid length, and PBL scheme). The appendix contains the complete MAE dataset used in the following discussion for the bias-corrected forecasts and a summary of forecast skill gained through bias correction.

a. Deterministic verification

Annual Taylor diagrams for sites 1–4 (Fig. 3) provide a quick summary of the short-term forecast performance of the 48 forecasts. Better forecasts will have lower CRMSEs (closer to the center of the gray semicircles), a linear correlation coefficient (dashed radii) closer to 1.0 (on the abscissa), and match the standard deviation of the observations (be closer to the yellow line). In Fig. 3, similar symbols represent similar PBL schemes and similar colors represent similar grids (i.e., grid length and initial condition source).

While each Taylor diagram contains 48 forecasts, several key patterns can be found by looking at the differences in the plots between wind farm sites. At sites 1 and 4, the Taylor diagrams show grid clustering, meaning forecasts from the same grid length perform similarly, regardless of the PBL scheme or initial conditions used. In other words, differences in the forecast performance are mainly a function of grid length or of the location of the nearest-neighbor grid point. At site 1, the grid clusters are primarily differentiated by their correlation coefficients, with the 12-km grids having the highest correlation and lowest CRMSEs. At site 4, the grids are primarily differentiated by their standard deviation, with the 4-km grids able to most closely match the standard deviation of the observations (the yellow quarter-circle curve). At site 4, however, the 36-km grids have the highest correlation, while having similar accuracy to the 12- and 4-km grids.

At sites 2 and 3, the grid-clustering pattern is not so distinct. At site 2, the 12-km grids provide the closest match to the observations, while the 4-km (36 km) grids generally have too little (too much) standard deviation. Forecasts using the MRF PBL scheme, which is an older
approach, have the best correlation with the observations. At site 3, the grid clustering is evident, but the difference in standard deviation between the different grid clusters is much smaller than at sites 1 and 4. At site 3, the 12-km grids also provide the best forecasts because they have higher accuracy and correlation, while better matching the standard deviation of the observations.

While the Taylor diagrams (Fig. 3) have hinted at a large importance in the choice of grid length in complex terrain, there is also a clear difference in the performance of forecasts using different PBL schemes for the same grid length. To quantify which of the factors (grid length, PBL scheme, or initial condition source) is the most important, we later perform an ANOVA on the distribution of MAE scores.

One other finding of note is that using either the GFS or NAM as the initial conditions produced similar forecasts for the short term (see the appendix). To simplify subsequent figures, the NAM-initialized members are omitted. One potential reason for the similarity in the GFS- and NAM-produced forecasts for equivalent configurations is that both of these models are produced by NCEP and rely on the same underlying data assimilation system, the Gridpoint Statistical Interpolation analysis system (GSI; Shao et al. 2016). Thus, their initial states could be more similar than when compared with initial conditions.
produced from other agencies or data assimilation methods.

Reducing the dataset further, Fig. 4 shows the MAE standardized anomaly $z$ scores for the bias-corrected WRF forecasts at each site for the summer (June–August) and winter (December–February) seasons. Since the standardized anomaly compares the MAEs of each scheme to the mean MAE of all schemes, negative values indicate configurations that perform better than the average (lower MAEs).

Figure 4 shows two distinct patterns in forecast accuracy. PBL schemes that include nonlocal mixing (YSU, ACM2, and MRF schemes; Fig. 4, boxed) have accuracy levels that are higher than those that include only local mixing in summer, although forecasts using the MYJ scheme also generally had better than average accuracy. Nonlocal schemes treat vertical mixing in a more realistic way by recognizing that turbulent eddies can span and/or move through multiple small layers within the boundary layer. In other words, mixing is attributed to not only eddies present within a small, local layer of constant static stability, but also those that originate in other layers that move through in transit to their level of neutral buoyancy (Stull 1993). These large eddies will
have the most influence in complex terrain over the summer season where differential heating over sloped terrain can lead to important boundary layer circulation development.

The 12-km forecast using the ACM2 PBL scheme initialized off the GFS had the best accuracy averaged over all wind farms during the summer (Fig. 4, purple bars) although other forecasts had better accuracy at individual wind farms. For example, the 4-km forecast using the ACM2 PBL scheme was best at sites 2 and 3, and the 12-km GBM scheme was best at site 4, all initialized off the GFS.

During the winter, the MAE $z$ scores show a different pattern. Here, the 12-km grids for each scheme, regardless of the treatment of vertical mixing, generally perform better (Fig. 4, boxed), especially when averaged over all locations. The exceptions are the MYNN and MRF PBL schemes, which perform relatively poorly regardless of grid length. We also find several cases where the 4-km grids produce better than average forecasts.

Averaged over all locations, the best forecast for the winter was the 12-km forecast produced using the GBM PBL scheme. Again, individual sites did show slightly better accuracy with other configurations. Site 1 had the best accuracy from the 12-km grid using the GBM PBL scheme. Site 2 had the highest accuracy using the 4-km grid and ACM2 PBL scheme, site 3 with the 4-km grid and MRF PBL scheme, and site 4 with the 12-km grid and ACM2 PBL scheme. In contrast to the summer, the most accurate bias-corrected forecasts in the winter were initialized off the NAM. In winter when static stability is generally higher and wind shear–generated mechanical turbulence is greater, nonlocal mixing is likely not as important, especially at these latitudes where the sun angle is low. Low pressure systems frequently approach the coast of British Columbia and move into the interior of North America, and dynamically forced flows have more influence on wind patterns in complex terrain.

While one might expect the 4-km grid to have the best forecast verification because of the superior resolution of the topography, our results indicate the 12-km grids generally have higher accuracy (with the mentioned exceptions), particularly when averaged over all locations. Mass et al. (2002) suggest that while shorter grid lengths do simulate the structure of atmospheric phenomena better, they are subject to positional error. Mass et al. (2002) and Murphy and Winkler (1987) also mention that skill measures should ultimately be defined by the forecast user. So, while the 4-km grids may produce more realistic-looking feature structures, for the purposes of generation planning, the forecast with the least error over the long term at a wind farm location is ultimately preferred. Our results show that higher-resolution grids (those with shorter grid lengths) may provide a poorer quality forecast than coarser (12 km) grids, while being more computationally expensive to produce (although we do find several cases where the 4-km grids are the most accurate). However, we are not advocating entirely for coarse grids to be used for wind forecasts in regions of complex terrain. The grid re-sampling process that occurs with two-way nested domains could result in the 4-km grids having a positive influence on the 12-km grid forecast accuracy. Additionally, it is clear from Fig. 4 that the 36-km grids generally produce some of the least accurate forecasts.

To highlight the potential for improvement in forecast accuracy by using particular PBL schemes or grid lengths, we have calculated the MAESSs for the best forecast configuration relative to the worst for each site for both the summer and winter seasons. In summer, sites 1–3 had MAESS values (i.e., a potential improvement in MAE) of 18%, 17%, and 19%, respectively. Site 4 had a much larger MAESS of 29%. In winter, the MAESSs were 24%, 17%, 15%, and 13% for sites 1–4, respectively. Differences of this magnitude are significant to the wind energy community (Marquis et al. 2011) and highlight the importance of careful grid-length and PBL physics configuration.

Figure 5 shows the MAE $z$ score for the entire year. These results show the best forecast averaged over the entire study period and all locations to be the 12-km ACM2 forecast initialized with the GFS initial condition source (Fig. 5, purple bars). Forecasts from the 4-km grid using the ACM2 scheme, and the 12- and 4-km grids using the GBM schemes, all with the GFS initial condition source, are also some of the more accurate forecasts.

b. Sensitivity analysis

So far, this study has shown that proper selection of the PBL scheme and grid length can lead to significantly improved forecast accuracy for hub-height wind forecasts. Also, the best-performing configuration varies by location and season. To quantify which factor (PBL scheme, grid length, or initial condition) is the most important to forecast accuracy at these wind farms, a multifactor (factorial) ANOVA was done on the MAE scores.

Figure 6 shows the contribution to the overall variance in MAE scores for the bias-corrected WRF forecasts attributed to each factor at each wind site. The plot also shows how the contribution to variance changes by season. In addition, the residual column includes the variance that can be attributed to the interaction between two or more factors.

At sites I and 4, where the grid clustering was easy to distinguish in Fig. 3, the grid length is the dominant factor.
controlling forecast accuracy for all seasons of the year. Site 4 in particular has the largest influence from the grid length, contributing to 70% or greater of the variance among the MAE scores for all seasons. At site 1, the grid length explains 50% of the variance in MAE, or greater, depending on the season. Both of these locations also show a seasonal cycle, where the PBL scheme or the residual increase contribute to the variance in MAE scores. At site 1, the PBL scheme has the largest influence in the spring and summer months, although the grid length is still the dominant factor. At site 4, the influence of the PBL scheme increases during the fall and winter, but

| Forecast Name [PBL] (IC) (GRID) |
|--------------------------------|
| YSUGFS12 | YSUGFS4 | YSUGFS8 | YSUGFS12 | YSUGFS4 | YSUGFS8 | YSUGFS12 | YSUGFS4 | YSUGFS8 | YSUGFS12 |
| MYNFS4 | MYNFS8 | MYNFS12 | MYNFS4 | MYNFS8 | MYNFS12 | MYNFS4 | MYNFS8 | MYNFS12 | MYNFS4 | MYNFS8 | MYNFS12 |
| UWGFS4 | UWGFS8 | UWGFS12 | UWGFS4 | UWGFS8 | UWGFS12 | UWGFS4 | UWGFS8 | UWGFS12 | UWGFS4 | UWGFS8 | UWGFS12 |
| GBMFS4 | GBMFS8 | GBMFS12 | GBMFS4 | GBMFS8 | GBMFS12 | GBMFS4 | GBMFS8 | GBMFS12 | GBMFS4 | GBMFS8 | GBMFS12 |

**FIG. 5.** Annual MAE $z$ score for each of the 48 bias-corrected wind forecasts. Colors represent individual wind farm sites, with purple bars representing the site-averaged performance. A larger-magnitude negative $z$ score indicates forecasts that are much better than the average.

**FIG. 6.** Contributions to the variance in MAE scores for the bias-corrected wind forecasts as a percentage contribution by each factor for each site. Different colors represent different seasons, where blue is summer (JJA), red is fall (SON), orange is winter (DJF), and green is spring (MAM). The residual represents the variance attributed to factor interactions and other unaccounted differences.
explains 12% or less of the variance in forecast accuracy. At site 2, the PBL scheme is the dominant factor, along with a strong seasonal cycle. The PBL scheme explains a minimum of 36% of the variance in winter and a maximum of 62%–70% in the spring and summer months. At site 3, the PBL scheme is dominant during the fall and spring, the grid length during the winter, and the initial condition source during the summer. Site 3 is the only location where we have seen the initial condition choice explain most of the variance for an individual season (52% during the summer). Site 3 is also the location where the PBL scheme has explained the highest amount of the variance in MAE for an individual season, at 80% during the spring.

Like the initial analyses, the ANOVA also indicates that the choice of initial and boundary conditions explains little of the variance in forecast accuracy for the short term at most of the locations, with the exception being site 3 during the summer. However, there are two potential caveats of this analysis. The first is that both sources are produced by NCEP, with both relying of the GSI data assimilation system. Greater variance may be achieved by using initial condition sources produced by separate agencies and separate data assimilation methods. The second is that greater variance may be achieved at longer forecast horizons since a 24-h forecast may not have allowed enough time for the small differences in the initial conditions to grow significantly.

Since the grid length affects the exact location of the nearest-neighbor grid-cell location, we posit the physical location of the nearest-neighbor grid point could contribute to the variance in accuracy scores at some locations (especially those that have their accuracy largely explained by the grid length). To evaluate these effects, we looked at the distance between the nearest-neighbor grid point and the actual coordinates of the wind farm site. At site 1 (and site 4), where the grid length was seen to control the MAE variance for individual seasons (Fig. 6), we found horizontal distances of 14.5, 4.1, and 1.4 km (and 11.8, 5.5, and 2.1 km), for the 36-, 12-, and 4-km grids, respectively. At site 2 (and site 3), which showed the PBL scheme to be the dominant factor for most seasons, the horizontal distances were found to be 15.2, 6.5, and 2.1 km (and 19.3, 2.7, and 1.1 km), for the 36-, 12-, and 4-km grids, respectively. This indicates that the physical error in location of the nearest-neighbor grid cell may not be the cause for some locations being more largely influenced by the grid length.

Next, we checked to see if differences in the modeled terrain height at the nearest-neighbor grid points are larger between grid lengths at locations that showed a higher influence of the grid length. Again, this does not seem to explain why some locations are more affected by the grid length (sites 1 and 4; Fig. 6), as all locations had similar height variations between nested grids of approximately 100–200 m. Finally, we checked to see if there was an abrupt change in land-use category and thus roughness between the grids and came to a similar conclusion, although for one location the nearest 36-km grid cell is identified as being over water. Thus, the reason why hub-height wind forecasts at particular sites are more affected by the PBL scheme, while others by the grid length, remains unclear and is worthy of future study. One theory could be that some locations are more affected by flow channeling and blocking, which may be more largely influenced by the grid length than the PBL scheme.

Figure 7 shows the contribution to the MAE variance by time of day at each location (over the entire year). The midday [1200–1800 Pacific standard time (PST)] peak in the PBL scheme’s contribution to the variance for sites 1, 2, and 3 is consistent with the development of slope flows and boundary layer growth. Forecast accuracy at site 4 does not exhibit a diurnal cycle, with the grid length explaining the most variance throughout the entire day. Since each forecast run was initialized at the same time of day, the results in Fig. 7 also show the relative contribution of each factor by forecast horizon, although the signal of the diurnal cycle appears to be more evident than that of the forecast horizon.

4. Conclusions and future work

This study compared the sensitivity of hub-height wind speed forecasts to the PBL scheme, grid length, and initial condition source for the WRF Model over complex terrain. We have provided several important conclusions that forecasters and researchers should consider when using NWP models to produce hub-height wind speed forecasts. Namely, we found the grid length and PBL scheme were the most important factors to consider for short-term forecasts (24-h forecast horizon) in complex terrain, for our datasets. For the majority of locations, the effects of PBL scheme and grid length contribute more to differences in the bias-corrected forecast accuracy than the initial condition choice. This means that choosing the best PBL scheme and grid length over the worst could yield the largest accuracy gains. However, one caveat could be a lack of diversity in initial condition sources, since the processes used during data assimilation for the NAM and GFS are similar, and both are produced by the same agency. Initializing WRF with other initial condition sources, such as the with the Canadian Global Deterministic Prediction System (GDPS), the Fleet Numerical Meteorology and Oceanography Center Navy
Global Environmental Model (NAVGEM), the European Centre for Medium-Range Weather Forecast Integrated Forecast System (IFS or ECMWF), or the Met Office Unified Model (UKMO) could result in a larger sensitivity to initial condition choice than has been shown here, and is worthy of future study.

At sites 2 and 3, the PBL scheme was a large or dominant factor controlling forecast accuracy, regardless of the season. At the other two stations (sites 1 and 4), the grid length had the largest influence on forecast accuracy. Additionally, the influence of the PBL scheme showed a seasonal cycle at sites 1, 2, and 4. At sites 1 and 2, the influence of the PBL scheme (grid length) peaked during the spring and summer (fall and winter) months. At site 4, the influence of the PBL scheme was small, but peaked in the winter. Site 3 had a different seasonal trend where the PBL scheme and grid length were the largest influence over forecast accuracy for all seasons, except for the summer.

The influence of the PBL scheme was also largest during times of the day when boundary layer circulations are expected to develop (from midday to early afternoon at sites 1–3). Site 4 showed relatively little variation in which factor contributed the most to the variance in accuracy, regardless of the time of day.

We investigated if the reason why sites 1 and 4 were more heavily influenced by the grid length was because of differences in the location of the nearest-neighbor grid point. We concluded that the physical distance of the wind farm location to the nearest-neighbor location was not the primary cause. We also found that the differences in elevation among the nearest-neighbor grid points were similar at all locations, and that land-use types were similar between grids at three of the four locations. We concluded that the locations that were most strongly influenced by the grid length are likely more impacted by terrain-flow interactions, such as channeling and blocking, than at the other locations, but that this finding is worthy of future study.

When averaged over all locations and seasons, the ACM2 PBL scheme was the most accurate PBL scheme in complex terrain, although for some specific farms and seasons it was slightly outperformed by other schemes. Energy planners will likely benefit the most by using the ACM2 scheme in complex terrain since it had the best overall average accuracy (averaged over all locations).

Choosing the best PBL scheme and grid-length configuration resulted in reductions in the MAE of up to 29% over the poorest configuration, depending on the season, with larger improvements found during the summer.

Improvements of this magnitude can save millions of dollars (Marquis et al. 2011). Our sensitivity results provide valuable information to meteorologists, energy planners, and utilities that use wind energy forecasts.
| Initial conditions | Grid length (km) | YSU | ACM2 | MRF | MYJ | MYNN | QNSE | UW | GBM |
|-------------------|-----------------|-----|------|-----|-----|------|------|----|-----|
|                   | Annual          |     |      |     |     |      |      |    |     |
| Site 1            | 36              | 2.30| 2.16 | 2.45| 2.28| 2.12 | 2.41| 2.27| 2.20 |
| Site 2            | 12              | 1.83| 1.82 | 1.79| 1.76| 1.77 | 1.68| 1.81| 1.86 |
| Site 3            | 4               | 1.81| 1.77 | 1.78| 1.72| 1.73 | 1.73| 1.77| 1.73 |
| Site 4            | 36              | 1.87| 1.61 | 1.62| 1.85| 1.61 | 1.60| 1.86| 1.66 |
| Summer            | Site 1          | 2.04| 1.99 | 2.14| 1.95| 1.90 | 2.08| 1.97| 1.91 |
| Site 2            | 1.70            | 1.70 | 1.64 | 1.58| 1.59 | 1.53 | 1.69| 1.65| 1.58 |
| Site 3            | 1.64            | 1.61 | 1.60 | 1.56| 1.53 | 1.52 | 1.66| 1.62| 1.55 |
| Site 4            | 1.58            | 1.21 | 1.32 | 1.57| 1.21 | 1.28 | 1.64| 1.25| 1.34 |
| Fall              | Site 1          | 2.32| 2.13 | 2.36| 2.32| 2.12 | 2.34| 2.19| 2.43 |
| Site 2            | 1.79            | 1.83 | 1.82 | 1.72| 1.77 | 1.71 | 1.76| 1.92 |1.87 |
| Site 3            | 1.75            | 1.70 | 1.72 | 1.70| 1.62 | 1.66 | 1.73| 1.68 |1.68 |
| Site 4            | 1.96            | 1.84 | 1.74 | 1.97| 1.86 | 1.73 | 1.95| 1.90 |1.79 |
| Winter            | Site 1          | 2.67| 2.45 | 2.98| 2.76| 2.45 | 3.01| 2.76 |2.62 |
| Site 2            | 2.03            | 1.90 | 1.92 | 2.03| 1.93 | 1.80 | 2.05| 2.05 |2.05 |
| Site 3            | 1.96            | 1.85 | 1.79 | 2.05| 1.89 | 1.87 | 1.92 |1.85 |1.79 |
| Site 4            | 2.16            | 1.97 | 1.98 | 2.10| 1.93 | 1.94 | 2.06 |1.99 |2.02 |
| Spring            | Site 1          | 2.16| 2.06 | 2.32| 2.11| 2.00 | 2.22| 2.04 |2.07 |
| Site 2            | 1.82            | 1.84 | 1.78 | 1.74| 1.77 | 1.69 | 1.76 |1.81 |1.80 |
| Site 3            | 1.90            | 1.95 | 1.95 | 1.83| 1.87 | 1.77 | 1.75 |1.75 |1.75 |
| Site 4            | 1.68            | 1.33 | 1.35 | 1.67| 1.35 | 1.34 | 1.72 |1.41 |1.39 |

Table A1. MAEs (m{s}^{-1}) for each bias-corrected forecast initialized using the GFS. Sites 1–4 are the anonymous wind farm locations. Statistics are divided into five blocks: annual, summer (JJA), fall (SON), winter (DJF), and spring (MAM).
| Grid length (km) | YSU | ACM2 | MRF | MYJ | MYNN | QNSE | UW | GBM |
|-----------------|-----|------|-----|-----|------|------|----|-----|
| Annual          |     |      |     |     |      |      |    |     |
| Site 1          | 36  | 12   | 4   | 36  | 12   | 4    | 36 | 12  |
| Site 2          | 1.81| 1.83 | 1.76| 1.75| 1.80 | 1.70 | 1.81| 1.86| 1.79|
| Site 3          | 1.87| 1.81 | 1.80| 1.83| 1.75 | 1.76 | 1.81| 1.77| 1.72|
| Site 4          | 1.88| 1.63 | 1.64| 1.85| 1.63 | 1.63 | 1.88| 1.66| 1.67|
| Summer          |     |      |     |     |      |      |    |     |
| Site 1          | 2.07| 1.95 | 2.13| 1.99| 1.86 | 2.08 | 2.01| 1.90| 2.09|
| Site 2          | 1.73| 1.72 | 1.70| 1.63| 1.66 | 1.62 | 1.73| 1.74| 1.67|
| Site 3          | 1.80| 1.80 | 1.71| 1.70| 1.69 | 1.63 | 1.75| 1.75| 1.66|
| Site 4          | 1.60| 1.26 | 1.36| 1.58| 1.28 | 1.35 | 1.64| 1.27| 1.36|
| Fall            |     |      |     |     |      |      |    |     |
| Site 1          | 2.44| 2.19 | 2.51| 2.45| 2.15 | 2.48 | 2.37| 2.19| 2.52|
| Site 2          | 1.75| 1.83 | 1.74| 1.71| 1.82 | 1.69 | 1.74| 1.89| 1.81|
| Site 3          | 1.80| 1.73 | 1.77| 1.75| 1.67 | 1.70 | 1.80| 1.75| 1.75|
| Site 4          | 2.00| 1.86 | 1.77| 2.01| 1.86 | 1.74 | 1.98| 1.87| 1.80|
| Winter          |     |      |     |     |      |      |    |     |
| Site 1          | 2.73| 2.39 | 2.90| 2.81| 2.45 | 3.01 | 2.74| 2.56| 3.06|
| Site 2          | 1.95| 1.92 | 1.82| 1.91| 1.92 | 1.75| 1.98| 2.00| 1.92|
| Site 3          | 1.98| 1.82 | 1.76| 2.05| 1.84 | 1.84 | 1.94| 1.83| 1.74|
| Site 4          | 2.11| 1.92 | 1.95| 2.05| 1.88 | 1.90 | 2.06| 1.96| 1.97|
| Spring          |     |      |     |     |      |      |    |     |
| Site 1          | 2.32| 2.17 | 2.40| 2.25| 2.09 | 2.32 | 2.16| 2.11| 2.27|
| Site 2          | 1.83| 1.85 | 1.80| 1.76| 1.82 | 1.72 | 1.78| 1.81| 1.77|
| Site 3          | 1.89| 1.90 | 1.95| 1.80| 1.81 | 1.88 | 1.75| 1.73| 1.74|
| Site 4          | 1.73| 1.37 | 1.42| 1.70| 1.42 | 1.43 | 1.78| 1.45| 1.47|
Namely, savings can be realized through careful experimentation with their WRF configuration, focusing on PBL scheme and grid-length selection, as well as bias correction.

Because we have shown that the best PBL scheme and grid length vary by location and season, it may be beneficial to use an ensemble of multiple PBL schemes, grid lengths, and initial condition sources to improve accuracy through use of the ensemble mean forecast. This study has only evaluated the deterministic aspect of wind forecasts. Capturing forecast uncertainty is critical, especially for energy planning. The use of an ensemble consisting of all available PBL schemes, grid lengths, and initial conditions may be beneficial, especially over complex terrain where boundary layer circulations are prevalent. A forthcoming paper (Siuta et al. 2017) evaluates the bias-correction method and the use of this dataset for ensemble and probabilistic forecasts. Use of probabilistic information has been shown to provide further economic value over individual deterministic forecasts (Zhu et al. 2002; McCollor and Stull 2008).

Finally, an interesting study would be to compare the effects of one- and two-way domain nesting techniques on hub-height wind forecasts over complex terrain. While we find the 12-km grids to generally have higher accuracy, particularly in the winter, it is possible that this is a result of the resampling and averaging process that takes place between nested grids. The 12-km grid forecast at these locations might be improved over the 4-km grid through this feedback process. A useful follow-up to this study would be to rerun the same tests with one-way nesting. This would eliminate fine-grid resampling and may provide different results with respect to the performance of fine and coarse grid lengths. It would also provide insight into the benefits of one- versus two-way nested grid setups in WRF over complex terrain.

Acknowledgments. The authors acknowledge BC Hydro (Grant 00089063), Mitacs (Grant IT03826), and the Canadian Natural Science and Engineering Research Council (Grant 184017-12) for providing the funding to make this extensive project possible. We also thank GDF SUEZ Energy North America, Altagas Ltd., Alterra Power Corp., and Capital Power for allowing us access to their observations. In addition, we thank Magdalena Rucker, Doug McCollor, Thomas Nipen, and Pedro Odon for their contributions to this work. Finally, we thank the three anonymous reviewers for their comments and suggestions, which made this a much better manuscript.

APPENDIX

MAE Dataset

Tables A1 and A2 provide the full MAE dataset for wind farm sites 1–4. MAE is provided for the annual and seasonal averages, for each PBL grid combination, and for using the GFS (Table A1) and NAM (Table A2) models as initial condition sources for the bias-corrected wind forecasts.

Figure A1 shows the annual MAESS of the bias-corrected wind forecasts (skill score is relative to the corresponding raw forecast, with larger positive scores being better) at each station using the DMB post-processing method. For sites 1–3 improvement in MAE due to bias correction ranged from 0% to 12% over the MAE results of the raw forecasts. While this
REFERENCES
Ahlstrom, M., and Coauthors, 2013: Knowledge is power: Efficiently integrating wind energy and wind forecasts. IEEE Power Energy Mag., 11 (6), 45-52, doi:10.1109/MPE.2013.2277999.
Arya, S. P., 2001: Introduction to Micrometeorology. 2nd ed. Academic Press, 420 pp.
Bourdin, D. R., T. N. Nipen, and R. B. Stull, 2014: Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system. Water Resour. Res., 50, 3108-3130, doi:10.1002/2014WR015462.
Bretherton, C. S., and S. Park, 2009: A new moist turbulence parameterization in the Community Atmosphere Model. J. Climate, 22, 3422–3448, doi:10.1175/2008JCLI2556.1.
Canadian Wind Energy Association, 2008: Wind Vision 2025: Powering Canada’s future. CanWEA. Rep., 24 pp. [Available online at http://www.ipcc.ch/pdf/assessment-report/ar5/wg2/IPCC_WGIIAR5-PartA_FINAL.pdf.]
Cheng, W. Y., Y. Liu, Y. Liu, Y. Zhang, W. P. Mahoney, and T. T. Warner, 2013: The impact of model physics on numerical wind forecasts. Renewable Energy, 55, 347–356, doi:10.1016/j.renene.2012.12.041.
Chow, F. K., B. J. Snyder, and S. F. J. De Wekker, Eds., 2013: Mountain Weather Research and Forecasting: Recent Progress and Current Challenges. Springer, 750 pp.
Clement, M. A., 2012: A methodology to assess the value of integrated hydropower and wind generation. M.S. thesis, Dept. of Civil, Environmental, and Architectural Engineering, University of Colorado, 183 pp. [Available online at https://cadswes.colorado.edu/sites/default/files/Theses-PhD/Clement_Thesis2012.pdf.]
Corbus, D., D. Lew, G. Jordan, W. Winters, F. Van Hull, J. Manobianco, and B. Zavadil, 2009: Up with wind. IEEE Power Energy Mag., 7 (6), 36–46, doi:10.1109/MPE.2009.934260.
DeMeo, E., G. Jordan, C. Kalich, J. King, M. Milligan, C. Murley, B. Oakleaf, and M. Schuerverger, 2007: Accommodating wind’s natural behavior. IEEE Power Energy Mag., 5 (6), 59–67, doi:10.1109/MPE.2007.906562.
Department of Energy, 2008: 20% wind energy by 2030: Increasing wind energy’s contribution to U.S. electricity supply. Dept. of Energy Rep. DOE/GO-102008-2567, 228 pp. [Available online at http://www.nrel.gov/docs/fy08osti/41869.pdf.]
Deppe, A. J., W. A. Gallus Jr., and E. S. Takle, 2013: A WRF ensemble for improved wind speed forecasts at turbine height. Wea. Forecasting, 28, 212–228, doi:10.1175/WAF-D-11-00112.1.
Draxl, C., A. N. Hahmann, A. Peña, and G. Giebel, 2014: Evaluating winds and vertical wind shear from Weather Research and Forecasting model forecasts using seven planetary boundary layer schemes. Wind Energy, 17, 39–55, doi:10.1002/we.1555.
Greene, J. S., M. Chatelain, M. Morrisey, and S. Stadler, 2012: Projected future wind speed and wind power density trends over the western US high plains. Atmos. Climate Sci., 2, 32–40, doi:10.4246/acs.2012.21005.
Grenier, H., and C. S. Bretherton, 2001: A moist PBL parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers. Mon. Wea. Rev., 129, 357–377, doi:10.1175/1520-0493(2001)129<0357:AMPPFL><2.0.CO2.]
Hahmann, A. N., C. L. Vincent, A. Peña, J. Lange, and C. B. Hasager, 2015: Wind climate estimation using WRF model output: Method and model sensitivities over the sea. Int. J. Climatol., 35, 3422–3439, doi:10.1002/joc.4217.
Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. Mon. Wea. Rev., 124, 2322–2339, doi:10.1175/1520-0493(1996)124<2322:NBLVDI><2.0.CO2.]
Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Wea. Rev., 122, 927–945, doi:10.1175/1520-0493(1994)122<0927:TSMEMC><2.0.CO2.]
Jiménez, P. A., and J. Dudhia, 2012: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. J. Appl. Meteor. Climatol., 51, 300–316, doi:10.1175/JAMC-D-11-084.1.
Marquis, M., J. Wilczak, M. Ahlstrom, J. Sharp, A. Stern, J. C. Smith, and S. Calvert, 2011: Forecasting the wind to reach significant penetration levels of wind energy. Bull. Amer. Meteor. Soc., 92, 1159–1171, doi:10.1175/2011BAMS3033.1.
Martner, B. E., and J. D. Marwitz, 1982: Wind characteristics in southern Wyoming. J. Appl. Meteor., 21, 1815–1827, doi:10.1175/1520-0450(1982)21<1815:WCISW><2.0.CO2.]
Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? Bull. Amer. Meteor. Soc., 83, 407–430, doi:10.1175/1520-0477(2002)83<0407:DIHPRM><2.0.CO2.]
McCollor, D., and R. Stull, 2008: Hydrometeorological short-range ensemble forecasts in complex terrain. Part II: Economic evaluation. Wea. Forecasting, 23, 557–574, doi:10.1175/2007WAF2007064.1.
Monteiro, C., R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, 2009: Wind power forecasting: State-of-the-art 2009. Argonne National Laboratory Tech. Rep. ANL/DIS-10-1, 198 pp., doi:10.2172/968212.
Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. Mon. Wea. Rev., 115, 1330–1338, doi:10.1175/1520-0493(1987)115<1330:AGFFVV><2.0.CO2.]

12% improvement may seem small, the resulting monetary value can be large (Marquis et al. 2011). Site 4 saw large improvements (up to 56% for the 4-km grid using the MRF PBL scheme) as a result of large systematic biases. While the bias-correction scheme applied to some forecasts did make the MAE up to 3% worse at individual sites, the majority of the configurations showed improvement. We attribute the cause of poorer verification after bias correction at individual locations to the training dataset and to the fact that some grids (mostly of 12-km grid length) have low bias prior to bias correction (Siuta et al. 2017). When averaged over all locations, bias correction always resulted in improved accuracy skill.
Peterson, E. W., and J. P. Hennessey, 1978: On the use of power laws for estimates of wind power potential. J. Appl. Meteor., 17, 390–394, doi:10.1175/1520-0450(1978)017<0390:OTUOPL>2.0.CO;2.

Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. J. Appl. Meteor. Climatol., 46, 1383–1395, doi:10.1175/JAM2539.1.

Poulos, G. S., and Coauthors, 2002: CASES-99: A comprehensive investigation of the stable nocturnal boundary layer. Bull. Amer. Meteor. Soc., 83, 555–581, doi:10.1175/1520-0477(2002)083<0555:CACTOS>2.0.CO;2.

Santoso, E., and R. Stull, 1998: Wind and temperature profiles in the radix layer: The bottom fifth of the convective boundary layer. J. Appl. Meteor., 37, 545–558, doi:10.1175/1520-0450(1998)037<0545:WATPIT>2.0.CO;2.

——, and ——, 2001: Similarity equations for wind and temperature profiles in the radix layer, at the bottom of the convective boundary layer. J. Atmos. Sci., 58, 1446–1464, doi:10.1175/1520-0469(2001)058<1446:SEFWAT>2.0.CO;2.

Shao, H., and Coauthors, 2016: Bridging research to operations transitions: Status and plans of community GSI. Bull. Amer. Meteor. Soc., 97, 1427–1440, doi:10.1175/BAMS-D-13-00245.1.

Shin, H., and S.-Y. Hong, 2011: Intercomparison of planetary boundary-layer parameterizations in the WRF model for a single day from CASES-99. Bound.-Layer Meteor., 139, 261–281, doi:10.1007/s10546-010-9583-z.

Siuta, D., 2013: An analysis of the weather research and forecasting model for wind energy applications in Wyoming. M.S. thesis, Dept. of Atmospheric Science, University of Wyoming, 116 pp.

——, G. West, T. Nipen, and R. Stull, 2017: Calibrated probabilistic hub-height wind forecasts in complex terrain. Wea. Forecasting, doi:10.1175/WAF-D-16-0157.1, in press.

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

——, and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.

Storm, B., J. Dudhia, S. Basu, A. Swift, and I. Giammanco, 2009: Evaluation of the Weather Research and Forecasting model on forecasting low-level jets: Implications for wind energy. Wind Energy, 12, 81–90, doi:10.1002/we.288.

Stull, R. B., 1988: An Introduction to Boundary Layer Meteorology. Kluwer Academic, 666 pp.

——, 1993: Review of non-local mixing in turbulent atmospheres: Transilient turbulence theory. Bound.-Layer Meteor., 62, 21–96, doi:10.1007/BF00705546.

Sukoriansky, S., B. Galperin, and I. Staroselsky, 2005: A quasi-normal scale elimination model of turbulent flows with stable stratification. Phys. Fluids, 17, 085107, doi:10.1063/1.2009010.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., 106, 7183–7192, doi:10.1029/2000JD900719.

Warner, T. T., 2011: Numerical Weather and Climate Prediction. Cambridge University Press, 526 pp.

Wilks, D. S., 2011: Statistical Methods in the Atmospheric Sciences. 3rd ed. Elsevier, 676 pp.

Yang, B., and Coauthors, 2016: Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the Weather Research and Forecasting model. Bound.-Layer Meteor., 162, 117–142, doi:10.1007/s10546-016-0185-2.

Yang, Q., L. K. Berg, M. Pekour, J. D. Fast, R. K. Newsom, M. Stoebling, and C. Finley, 2013: Evaluation of WRF-predicted near-hub-height winds and ramp events over a Pacific Northwest site with complex terrain. J. Appl. Meteor. Climatol., 52, 1753–1763, doi:10.1175/JAMC-D-12-0267.1.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. Bull. Amer. Meteor. Soc., 83, 73–83, doi:10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2.