Sensitivity Analysis of Wind and Turbulence Predictions With Mesoscale-Coupled Large Eddy Simulations Using Ensemble Machine Learning

Colleen M. Kaul1, Zhangshuan Jason Hou1, Huifen Zhou1, Raj K. Rai1, and Larry K. Berg1

1Pacific Northwest National Laboratory, Richland, WA, USA

Abstract  Coupling between mesoscale models and large-eddy simulation (LES) models is increasingly used to more realistically represent the wide range of scales of atmospheric motions affecting boundary layer winds and turbulence that need to be simulated accurately for applications such as wind energy. However, such mesoscale-to-microscale coupled modeling frameworks are potentially affected by a large number of uncertain closure parameters. Here, we investigate the sensitivity associated with six closure parameters related to a 1.5-order subgrid-scale turbulence closure for an ensemble of mesoscale-coupled LES. The simulations are performed using the Weather Research and Forecasting model nested from horizontal resolutions of greater than a kilometer down to tens of meters. Closure parameters are varied to generate perturbed parameter ensembles for two case studies of highly sheared, convective boundary layers observed in the Columbia Basin of Oregon and Washington during the Second Wind Forecast Improvement Project. Machine learning algorithms are used to explore the sensitivity of LES predictions, considering the effects of the perturbed physical parameters alongside categorical factors such as the case study identity, measurement location, and LES resolution. For the conditions we examine, a single parameter, the eddy viscosity coefficient, is the dominant source of parametric sensitivity and its importance is comparable to the categorical factors for several of the simulation response variables we examine.

1. Introduction

The capability to realistically simulate turbulence and winds within the atmospheric boundary layer across a range of scales is important for a number of purposes, such as predictions of air quality and pollutant dispersal (Kadaverugu et al., 2019; Silveira et al., 2019) and modeling for renewable energy systems, notably wind energy. Design and control of wind turbines and siting, layout, and operation of wind farms are all influenced by atmospheric processes with scales ranging across several orders of magnitude (Haupt et al., 2019). Novel approaches are needed for capturing the range of scales that control evolution and transport of momentum, heat, and moisture (and possibly many additional scalars, in the context of air quality modeling), as doing so is intractable within a single modeling framework. Rather, it is generally necessary to employ mesoscale-to-microscale coupling techniques to link models that use a mesoscale modeling approach (e.g., the Weather Research and Forecasting [WRF] model) to large eddy simulation (LES) models that can explicitly simulate fine scales of turbulence. Thus, each model type can be used within its range of applicability, while the coupled modeling system is able to yield predictions that reflect the full range of relevant scales of atmospheric variability.

A concomitant disadvantage of this coupled modeling framework is its complexity, since it depends on the interplay of two types of sophisticated models as well as details of the coupling approach itself. Gaining an understanding of the sensitivities and uncertainties of such coupled modeling approaches is an essential, but challenging, step in establishing their usefulness for applications. In recent years, significant progress has been made toward developing best practices for mesoscale-to-microscale coupling for wind energy (Haupt et al., 2019; Rodrigo et al., 2020), but work remains to evaluate and improve the reliability of such coupled model systems.

A key challenge is the high computational cost of LES, in contrast to (relatively) computationally inexpensive Reynolds Average Navier-Stokes (RANS) models which are often used in practice for wind energy modeling. While a wide array of uncertainty quantification (UQ) methodologies have been developed for RANS simulations (Xiao & Cinnella, 2019) and some have been applied to modeling for wind energy (García-Sánchez & Gorlé, 2018; Gorlé et al., 2015; Hornshøj-Møller et al., 2021), UQ studies for LES remain limited and typically
consider far more idealized flow scenarios than a realistic atmospheric boundary layer flow (Lucor et al., 2007; Safta et al., 2017).

Another challenge is the multiplicity of sources of uncertainty for mesoscale-to-microscale coupled modeling. Many of these sources of uncertainty, for example, specification of boundary and initial conditions and numerical discretization choices, are structural in nature. However, parametric uncertainties associated with poorly constrained parameter values within model closures are also important. Both mesoscale models and LES make use of parameterizations to represent unresolved scales of turbulent motions, surface layer interactions, and other physical processes. In a holistic quantification of parametric uncertainty, sensitivities propagating from both mesoscale and microscale modeling components would be taken into account in order to quantify the overall coupled modeling uncertainty. However, the large number of poorly constrained and possibly influential parameters (plus the relatively high computational cost of LES) renders such an approach infeasible. This implies the need for an approach that first identifies the influential parameters in each of the mesoscale and LES closures, and then performs combined mesoscale and LES uncertainty analyses considering only a significantly reduced but representative subset of parameters.

From the mesoscale modeling perspective, there has been significant progress made in examining parametric sensitivities of hub-height wind predictions. For example, using the WRF model (Skamarock et al., 2019) in a mesoscale configuration to simulate wintertime conditions in the complex terrain of the Columbia Basin region, Yang et al. (2017) analyzed the sensitivity of predicted turbine-height wind speeds to 12 parameters of the Mellor-Yamada-Nakanishi-Niino (MYNN) Level 2.5 planetary boundary-layer (PBL) scheme (Nakanishi & Niino, 2006, 2009) and 14 parameters of the revised MM5 surface-layer scheme (Jiménez et al., 2012). Promisingly, this study found that most of the uncertainty in predicted wind speeds was attributable to just a few of the parameters in each of the boundary- and surface-layer schemes. Berg et al. (2019) confirmed this finding for wintertime conditions, while Yang et al. (2019) showed that the sensitivity of WRF’s Yonsei University PBL scheme (Hong et al., 2006) is also largely attributable to just a few parameters. Smith and Ancell (2019) used a similar methodology to examine sensitivity of wind ramp event predictions to initial conditions and parameters of the MYNN scheme. Importantly, these studies show that, even within a single region, the details of the sensitivity show diurnal (Yang et al., 2017, 2019) and seasonal variations (Berg et al., 2019) and can differ even between events of nominally similar types (Smith & Ancell, 2019). Here, we aim at understanding sensitivities of LES modeling for the same region.

In this work, we examine sensitivities to parameters of a widely-used 1.5-order closure for subgrid-scale (SGS) turbulent fluxes of momentum and scalars in simulations that use WRF’s grid nesting capability to span horizontal grid resolutions from 1.35 km to 50 m. This closure solves a prognostic equation for turbulent (alternatively, “turbulence”) kinetic energy (TKE) in order to provide a velocity scale for estimating an eddy diffusivity. In Section 2, we review this SGS closure, identify the parameters, and explain their prior uncertainty ranges. In Section 4 we describe the construction of a perturbed parameter ensemble (PPE) of simulations of two case studies with similarities in the large-scale flow scenario, but differences in the structure of fine-scale turbulence. Results of these simulations are qualitatively presented in Section 5. To quantitatively assess the relative significance of the SGS closure parameters for predictions of wind speed, direction, and turbulence levels, we construct random forest (RF) models of the PPE and calculate feature importance (FI) (Antoniadis et al., 2021; Genuer et al., 2010). These findings are presented in Section 6. Finally, we discuss our conclusions in Section 7.

2. Parameter Identification for a 1.5-Order SGS Turbulence Model

LESs typically employ a closure to represent turbulent fluxes associated with scales too fine to be resolved on the computational mesh. A number of such closures have been developed, but one of the most commonly used formulations for LES applications of WRF (as well as within many other LES models for atmospheric boundary layers) is based on the closure described by Deardorff (1980) (see, e.g., Gibbs and Fedorovich (2016) and references therein). This closure type is often referred to as a 1.5-order scheme because it solves a single prognostic equation for the SGS TKE $e = \frac{1}{2}\overline{u'v'}$ rather than a full set of second-moment equations. SGS momentum (scalar) fluxes are then computed using an eddy viscosity (eddy diffusivity) approach. Here we do not undertake to thoroughly review such turbulence modeling approaches but rather aim to highlight the roles of specific parameters.
within these closures and how we determine their range of values to test. In our discussion, we largely follow the notation used in documentation for WRF except as noted.

2.1. Description of the Model

The SGS TKE $e$ provides a velocity scale which is combined with a length scale $l$ and a coefficient $c_k$ to compute the eddy viscosity $K_m$ according to

$$K_m = c_k l e^{1/2}$$

Assuming homogeneous, isotropic turbulence and a sharp spectral cut-off filter, a theoretical value for $c_k$ can be obtained as

$$c_k = \left( \frac{2}{3\alpha} \right)^{3/2} \frac{1}{\pi}$$

where $\alpha$ is the Kolmogorov coefficient (Lilly, 1967). Taking $\alpha = 1.5$ as in Lilly (1967) yields $c_k = 0.094$, however other sources assume somewhat different values of $\alpha$ with correspondingly different values of $c_k$. For example, Schumann (1991) uses $\alpha = 1.6$. The limitations of such estimates are important to recognize, especially when we try to configure SGS closures for simulating non-homogeneous flows with a poorly defined, implicit, grid-based filter, and we note that WRF’s default specification of this parameter, $c_k = 0.15$, has a larger value than the theoretical estimates.

The length scale $l$ in Equation 1 is an important aspect of the closure that helps it to remain applicable under different states of stratification. When the local grid-scale stratification is neutral or unstable, $l = \Delta$, where $\Delta$ is the filter scale set by the grid resolution, that is, $\Delta = (\Delta_x \Delta_y \Delta_z)^{1/3}$. Under local stable stratification, where the scales of turbulence are suppressed, $l$ is limited by

$$l = c_n e^{1/2} \frac{1}{N}$$

where $N$ is the Brunt-Väisälä frequency and $c_n$ is a model coefficient. This closure can be understood by considering $c_n e^{1/2}$ as an approximation to the standard deviation of the SGS vertical velocity fluctuations. Different values for $c_n$ can be obtained by assuming different levels of suppression of vertical velocity fluctuations relative to horizontal velocity fluctuations. WRF’s default, which is also the value used by Deardorff (1980), is $c_n = 0.76$. Nevertheless, smaller values, which indicate stronger suppression of vertical variations, are also plausible. For example, $c_n = 0.5$ is suggested by Nieuwstadt (1990) and used by Gibbs and Fedorovich (2016). A reasonable upper bound is given by $c_n = 0.82$, which results from the assumption that the velocity fluctuations are isotropic.

SGS scalar fluxes can be computed using $K_m$ and a closure for the turbulent Prandtl number $Pr_t$. Deardorff (1980) proposes

$$Pr_t = \left( 1 + 2 \frac{l}{\Delta} \right)^{-1}$$

so that $Pr_t$ increases from 1/3 under neutral or unstable stratification to 1 under very stable stratification. While these limits are widely accepted, the details of how $Pr_t$ varies with stratification are far less clear (Gibbs & Fedorovich, 2016). To capture this uncertainty, we introduce an additional parameter into Equation 4 that varies the rate at which $Pr_t$ changes with decreasing $l/\Delta$. Our modified expression is

$$Pr_t = \left( 1 + 2 \left( \frac{l}{\Delta} \right)^{n_{Pr}} \right)^{-1}$$

and the effect of the introduced parameter $n_{Pr}$ is illustrated in Figure 1a. Note we are not suggesting Equation 5 be used as a permanent modification to the SGS scheme, but rather employed here as a means to identify possible sensitivity to the turbulent Prandtl number closures that would otherwise be difficult to isolate.

Production (or destruction) of SGS TKE due to shear and buoyancy effects can be modeled self-consistently using the eddy viscosity/eddy diffusivity-based SGS flux closures. However, additional assumptions are needed.
to account for the dissipation of TKE associated with molecular viscosity. This quantity, which is denoted by $D_\epsilon$, is parameterized as

$$D_\epsilon = c_\epsilon e^{3/2}$$

(6)

where $c_\epsilon$ is a model coefficient. At the first vertical grid level above the surface, an enhanced value $c_\epsilon = 3.9$ is used; at all other levels $c_\epsilon$ is modified in response to local stability. We express this stability dependence in the form

$$c_\epsilon = (c_{\epsilon,N} - c_{\epsilon,S}) \frac{\Delta}{c_k^2 n} + c_{\epsilon,S}$$

(7)

which differs slightly from WRF’s default expression that includes a dependence on $c_k$. Here $c_{\epsilon,N}$ is the value of $c_\epsilon$ when turbulent length scales are not limited by stability. As for $c_k$, a theoretical estimate can be made by assuming homogeneous isotropic turbulence and a sharp spectral filter, giving

$$c_{\epsilon,N} = \left(\frac{2}{3\alpha}\right)^{3/2}$$

(8)

Moeng and Wyngaard (1988) discuss a multiplicative correction factor of $3^{1/4}$ that can be applied to such estimates to better account for the filter shape (spherical rather than cubical in wave number space).

In Equation 7, $c_{\epsilon,S}$ is the minimum coefficient obtained under very stable conditions. Following de Roode et al. (2017), we can determine a critical subgrid-scale (SGS) Richardson number $Ri_c$, at which $c_\epsilon$ vanishes. In our notation, $Ri_c$ depends on the closure parameters as

$$Ri_c = \frac{1}{1 + c_{\epsilon,S}/(c_k^2 \alpha^2)}$$

(9)

which we can rearranged to yield

$$c_{\epsilon,S} = c_k^2 \alpha^2 (Ri_c^{-1} - 1)$$

(10)

Figure 1b illustrates the values of $c_{\epsilon,S}$ obtained with some different combinations of $c_k$, $c_N$, and $Ri_c$.

Finally, although it is not a parameter of the SGS turbulence closure per se, we are interested in checking how uncertainty in the surface flux parameterization might impact the LES solutions. For this purpose, we apply a multiplicative factor $z_f$ to the baseline roughness length used within the MYNN surface layer scheme using the stability functions of Dyer and Hicks (1970).
3. Determination of Parameter Ranges

As discussed by Yang et al. (2017), some subjectivity is involved in determining the ranges of parameter values to be tested in a sensitivity analysis. The prior ranges should be sufficiently inclusive of plausible values without arbitrarily enhancing the importance of a parameter with over-relaxed bounds. The parameter ranges and WRF’s default values (defined either explicitly or implicitly in the code) are summarized in Table 1. These ranges are inputs to a Latin hypercube sampling scheme (Helton & Davis, 2003; Stein, 1987).

A lower limit to the eddy viscosity coefficient \( c_k \) can be developed from the theoretical estimate (Equation 2) by varying the Kolmogorov coefficient \( \alpha \) by 10 percent around a nominal value of 1.5. This gives a minimum value of \( c_k = 0.08 \). The upper limit is set by doubling the default value used by WRF to give \( c_k = 0.3 \).

The theoretical estimate expressed as Equation 8 is similarly used to obtain a range between 0.8 and 1.1 for the neutral flow TKE dissipation coefficient, with application of the filter shape correction factor of Moeng and Wyngaard (1988) increasing the upper bound of values to 1.5.

The parameter \( c_n \) is limited between 0.5 and 0.82 as explained in the discussion following Equation 3, and the critical SGS gradient Richardson number \( R_{ic} \) is allowed to vary between 0.1 and 1. Observational analyses (e.g., Grachev et al., 2013) commonly indicate values of about 0.2–0.25, which is consistent with WRF’s assumptions.

The turbulent Prandtl number exponential factor \( n_{pr} \) is allowed to span between 0.01 and 100 to obtain limiting behaviors of \( Pr_t \) between 1/3 and 1 (see Figure 1a) and the roughness length factor \( z_f \) ranges between 1 and 2 following Yang et al. (2017).

Using these bounds, we generate 64 combinations of perturbed parameters using Latin hypercube sampling (Helton & Davis, 2003; Stein, 1987). Note we actually compute sample values of \( \log(n_{pr}) \) and \( \log(z_f) \). Each of these 64 parameter combinations is used to generate one ensemble member of a PPE of nested WRF/WRF-LES simulations for each of two case studies, yielding a total of 128 such simulations. Additional details of the case studies and simulation setups are provided in the next section.

4. Case Selection and Simulation Configurations

We investigate two case studies, each spanning multiple hours, on July 22, 2016 and August 21, 2016. These cases were selected from periods observed during the Second Wind Forecast Improvement Project (WFIP2) campaign, which was carried out in the Columbia Basin region of Washington and Oregon (Shaw et al., 2019; Wilczak et al., 2019). Both of these days were identified in the WFIP2 event log (Wilczak et al., 2019) as being associated with marine push events. Marine pushes or marine intrusions (Januzzi, 1993; Mass et al., 1986) are characterized by an onshore surge of cool maritime air, producing high wind speeds and strong surface heat fluxes. Strong surges can penetrate far inland from the coastal regions of Oregon and Washington to affect flow east of the Cascade Mountains, where there are many wind farms. Nonetheless, because these marine intrusion events are influenced by phenomena occurring at different spatial scales, individual episodes can exhibit diverse features. This, in turn, can affect parametric sensitivities when such events are simulated (Smith, 2018; Smith & Ancell, 2019).

### Table 1

| Parameter | Meaning                                      | Range          | WRF default |
|-----------|----------------------------------------------|----------------|-------------|
| \( c_k \) | Eddy viscosity coefficient                   | 0.08–0.3       | 0.15        |
| \( c_{\epsilon,N} \) | Neutral flow TKE dissipation coefficient  | 0.8–1.5        | 0.93        |
| \( R_{ic} \) | Critical SGS Richardson number             | 0.1–1          | 0.23313     |
| \( c_n \) | Length scale coefficient                    | 0.5–0.82       | 0.76        |
| \( n_{pr} \) | Turbulent Prandtl number factor            | 0.01–100       | 1           |
| \( z_f \) | Roughness length factor                     | 1–2            | 1           |

*Note:* The WRF default is the default value used in WRF v4.1.2 (either implicitly or explicitly defined).
The simulations are performed in WRF version 4.1.2 (with minor modifications to implement the perturbed parameters in the code) using three levels of nested domains, which are configured based on the findings of Rai et al. (2019). The domain configuration is illustrated in Figure 2. The outermost domain has a horizontal grid spacing of 1.35 km, slightly in excess of the boundary layer height, and simulation of the flow on this domain is performed in a mesoscale modeling mode using the MYNN level 2.5 PBL and surface scheme (Nakanishi & Niino, 2006, 2009). Initial and boundary conditions are specified from Global Forecast System reanalysis. This domain is treated identically among all ensemble members. For both case studies, the mesoscale domain is initialized at 12 UTC and the solution is integrated for 6 hr before initiating solutions on two nested domains with horizontal grid spacing of 150 and 50 m. Both of these nests are treated as LES and the same set of perturbed parameter values is used on each. All three domains are advanced for at least four additional hours, with a 2 hr period selected for detailed analysis. Specifically, we examine 20:20–22:20 UTC for July 22 and 20:00–22:00 UTC for August 21. These time intervals were chosen to obtain periods when the simulated wind speeds were fairly uniform in time while also allowing spin up of the nested domains.

Observations were collected during WFIP2 using a variety of instruments at a number of different sites covering a broad swath of the Columbia Basin (Wilczak et al., 2019). However, the limited extent of the LES domains means that only a subset of these can be encompassed. Therefore, the simulation domains are positioned so that all three include the Physics Site-12 (PS-12) tower that had sonic anemometers positioned at 50 and 80 m vertical levels. The Physics Site was a heavily instrumented area that was a key aspect of the WFIP2 observational strategy.

Figure 2. Extents of the three levels of WRF nested domains drawn over contours of terrain height, indicating for each whether they are run in mesoscale or large-eddy simulations (LES) modeling mode and their horizontal grid spacing.
Further details are provided in Wilczak et al. (2019). All domains also included the Wasco, OR airport site a few kilometers away. As described in Section 6, output from this second location is also used in the parameter sensitivity analysis. We note that the goal of our study is not to calibrate LES model parameters relative to observations; rather we are focused on determining which parameters among those we test are most influential and should be retained in future analyses. In the context of our study, comparisons with observed values are primarily used as a check that a relevant range of parameter values have been sampled in the PPE, as this range is not well established for many of the model parameters a priori.

Additionally, a common vertical grid with 129 levels is used for all nests. The levels are chosen to provide a gradual increase from 10 to 25 m vertical spacing over the lowest 1 km of the domain. Sensitivity to vertical resolution is not evaluated within our study, largely due to computational constraints, although in general simulated boundary layer winds and turbulence can be sensitive to the choice of vertical grid spacing. WRF’s fifth (third) order scheme is used for horizontal (vertical) advection of momentum and scalars. The mesoscale domain time step is 3 s, with time step ratios of 10 and 5 applied to the successive nests. To help account for these differences in temporal resolution, all simulation data is either interpolated or downsampled to 1 Hz frequency (as are the observational data) prior to analysis.

In addition to the description of the key model settings provided here, sample WRF namelist input files for the simulations can be obtained from the Atmosphere to Electrons Data Archive and Portal (https://a2e.energy.gov/about/dap).

5. Simulation Results

In this section, we present some key results of the PPE and examine parameter sensitivity on a qualitative basis, preparatory to creating machine learning models of the data to facilitate additional sensitivity analysis.

Figure 3 shows the 10-minute-averaged horizontal wind speed, \( U_{10\text{m}} \), at an 80 m elevation from the Physics Site-12 (PS-12) location. Measurements are shown as black lines and mesoscale domain results as gray lines. Blue (orange) solid line is the mean of perturbed parameter ensembles (PPE) at 50 m (150 m) resolution and blue (orange) shading is the range of the PPE.

**Figure 3.** Observed and simulated 10-minute averaged wind speeds, \( U_{10\text{m}} \), at 80 m vertical level from the Physics Site-12 (PS-12) location. Measurements are shown as black lines and mesoscale domain results as gray lines. Blue (orange) solid line is the mean of perturbed parameter ensembles (PPE) at 50 m (150 m) resolution and blue (orange) shading is the range of the PPE.
coherent structures in the flow. These structures can also be inferred from observations by examining streamwise and cross-stream turbulence timescales. Particular parameter values tend to enhance those structures so that shifts in their position produce a strong signal in the wind speed time series at a fixed location. Such structures are not as significant for the July case, indicating the importance of day-to-day variability even for cases associated with nominally similar meteorological phenomena. This change in the coherent structures in the flow is generally consistent with the higher wind speeds and lower surface heat fluxes for the August case relative to the July case, although understanding of the transitional behavior between roll and cellular organization of the convective boundary layer remains imperfect (e.g., Salesky et al., 2017) and is further complicated by terrain effects (Rai et al., 2017).

Surface sensible heat flux $H$ at the PS-12 location (Figure 4) is in excess of $300 \text{ W m}^{-2}$ throughout the LES simulations. The relative range of $U_{10_{\text{max}}}$ at a height of $80 \text{ m}$ among the ensemble members is three to six times greater than the relative range of $H$. This is suggestive that the spread in wind speed is attributable to factors in addition to variations in surface heat fluxes among the ensemble members (note that surface latent heat fluxes vary more weakly than sensible heat fluxes). Later results will show that wind speed predictions are not very sensitive to the surface roughness factor, which also supports this conclusion.

While predictions of winds and turbulence at hub-height have clear importance for wind energy applications, it is also important that models be able to capture the overall boundary layer structure well, as a prerequisite for simulating interactions within and between wind farms and for assessing the effects of wind farms on their environment. More generally, prediction of boundary layer height is an important component of modeling how the boundary layer interacts with the overlying layer of the atmosphere, entraining air with different characteristics of momentum and scalar concentrations. Figure 5 shows the variation of boundary layer height $h$ over the 2-hr windows used for detailed analysis of the LES. This quantity is diagnosed by computing bulk Richardson numbers from 10-minute averaged profile data and using a critical value $Ri_{bc} = 0$, similar to the method for determining boundary layer height used in WRF's Yonsei University PBL scheme (Hong et al., 2006). For the July case, $h$ is in general agreement among all three nest levels, although it is consistently lower on the coarser resolution LES domain compared to the higher resolution one. However, $h$ predicted by the LES ensembles (at both resolutions) is markedly shallower than $h$ obtained on the mesoscale nest for the August case, which is actually quite similar in value to $h$ predicted on the mesoscale domain for July. This result again highlights the differences in the finer-scale structures of the flow on the two case days, which can only be fully revealed through higher resolution simulations.

The results presented in Figures 3–5 indicate the significant spread among ensemble members for a selection of several important quantities. Now we turn to relating the differences among the ensemble members to differences
in the parameter values used within the simulation, continuing to focus on the 80 m vertical level at the PS-12 location. While the details of the results depend on the specific height and site, the key conclusions to be drawn remain consistent. All model outputs discussed in this section are computed over 2-hr windows (20:20–22:20 UTC for July 22 and 20:00–22:00 UTC for August 21).

For example, Figure 6 relates the range in $U_{10min}$ (averaged over the 2-hr windows defined above) to the six perturbed parameters, considering both the influence of the specific case study and grid resolution. Consistent with Figure 3, results of the July case show a low bias relative to the observations. For approximately $c_k > 0.2$, there appears to be a decreasing trend in wind speed with increasing eddy viscosity coefficient for the 150 m resolution results. Otherwise, no clear relationships with other parameters can be detected by visual inspection alone. In contrast, there is a marked trend of decreasing $U_{10min}$ with increasing $c_k$ at both resolutions and across the full range of $c_k$ values for the August case. As for the July case, relationships with the other parameters are not readily discernible.

The strength of turbulent fluctuations in the horizontal winds is an important quantity to be gleaned from LES, and Figure 7 considers the effect of parameter values on such predictions (here, specifically, the resolved TKE associated with timescales less than 10 min which is denoted as $E_{<10min}$). Only for the eddy viscosity parameter is a trend in values apparent, with $E_{<10min}$ declining for high $c_k$ until at some value, which depends on the resolution and case study, the flow is essentially non-turbulent. Note, however, that at smaller $c_k$ values (again, with this range of $c_k$ values being case and resolution dependent) the sensitivity of $E_{<10min}$ is quite weak. For both cases and resolutions, agreement with observed $E_{<10min}$ can be obtained using $c_k$ at or below approximately 0.1, which is close to the estimate of Lilly (1967) and equal to the value recommended by Deardorff (1980) and used in many LES codes. WRF’s default $c_k = 0.15$ seems to be an acceptable choice for the July case but too large for the August case.

While $E_{<10min}$ indicates the intensity of turbulence, it does not characterize its structure. For this purpose, we compute an integral timescale of the wind speed fluctuations. Although the magnitude of $E_{<10min}$ computed from observations was similar for both cases (black lines in Figure 7), the results in Figure 8 show the differences in how these fluctuations are organized: an integral timescale of just less than 90 s for the July case compared to nearly 400 s for the August case. The simulation results show a large range of predicted values. Repeating a by now familiar pattern, the overall sensitivity to $c_k$ appears much larger than to any other parameters, albeit with the details of the sensitivity to $c_k$ depending on whether $c_k$ falls into the range of low, intermediate, or high values for the specific case and resolution (as for $E_{<10min}$).

Although it might be expected that the tripling of horizontal grid spacing moving from the coarser, 150 m LES domain to the finer, 50 m LES domain would significantly improve the ability of the simulations to match observed characteristics of wind speed fluctuations, this does not seem to borne out in practice for these case studies. Depending on the choice of parameter values, model predictions at either resolution can approximate

---

**Figure 5.** Boundary layer height, diagnosed using critical bulk Richardson number $Ri_{rb} = 0$ at Physics Site-12 (PS-12). Line styles and colors as in Figure 3. Mesoscale domain results are shown as gray lines. Blue (orange) solid line is the mean of perturbed parameter ensembles (PPE) at 50 m (150 m) resolution and blue (orange) shading is the range of the PPE.
Figure 6. Dependence of the means of 10-minute averaged wind speeds $U_{10\text{min}}$ on perturbed parameter values for (a) the July 22, 2016 case study and (b) the August 21, 2016 case study. Results of the finer, 50-m (coarser, 150-m) resolution domain are shown by blue (orange) markers. The observed values are indicated by black horizontal lines.
Figure 7. Dependence of the turbulent kinetic energy at sub 10-minute timescales \( E_{<10\text{min}} \) on perturbed parameter values for (a) the July 22, 2016 case study and (b) the August 21, 2016 case study. Plotting conventions as in Figure 6: results of the finer, 50-m (coarser, 150-m) resolution domain are shown by blue (orange) markers. The observed values are indicated by black horizontal lines.
Figure 8. Dependence of the turbulence integral timescale on perturbed parameter values for (a) the July 22, 2016 case study and (b) the August 21, 2016 case study. Plotting conventions as in Figure 6: results of the finer, 50-m (coarser, 150-m) resolution domain are shown by blue (orange) markers. The observed values are indicated by black horizontal lines.
the observed values. However, the value of increased horizontal resolution is seen in predictions of wind shear (Figure 9), evaluated between 50 and 80 m vertical levels and averaged over the 2 hr periods defined above. Only at the higher resolution are some ensemble members able to match the observed values, whereas shear is over-predicted at the coarser resolution. At 50 m horizontal resolution, best results are obtained with an eddy viscosity coefficient of around 0.1. Shear increases roughly linearly with $c_k$ until values plateau for approximately $c_k > 0.2$. Sensitivity to other parameters appears quite limited.

Because the strong dependence of the model’s predictions on $c_k$ might impede detection of more subtle relationships with other parameters, and because simulation predictions were clearly degraded at large values of $c_k$, we created a subset of half the original ensemble members that were deemed to show better performance. The simulation quality was evaluated on the basis of subjectively chosen, but quantitative, criteria described below. Rather than directly selecting specific ensemble members for inclusion in the halved sub-set, we ranked the results of the finest grid spacing (50 m) simulations on the basis of the sum of the relative errors in their predictions of the three response variables shown in Figures 7–9, considering both the July and August cases, and selected the ensemble members with ranks 1 through 32. Although no explicit criterion on the value of $c_k$ was applied, this approach led to the selection of ensemble members with $c_k < 0.2$, roughly halving the original range in $c_k$ values (see Table 1). The sampling range of the remaining parameters is preserved, as shown in Figure 10.

This subsetting differs from the splitting of the data set that is inherent in the construction of the RF models that are discussed in the next section, as each RF model considers a single response variable. In contrast, the subsetting of the ensemble that we make considers a combination of response variables that characterize different aspects of the simulated winds.

6. Sensitivity Analysis Using RF

Machine learning tools can be used to create meta-models of how the LES results respond to input parameters (referred to as features in this context). Various model types can be used, offering different strengths and weaknesses. For example, Yang et al. (2017), Yang et al. (2019) and Berg et al. (2019) use a Generalized Linear Model. Here we focus on sensitivity analysis using RF regression (Breiman, 2001) as the meta-modeling approach. An advantage of the RF approach is its suitability for identifying parameters which affect the model output non-linearly and its ability to handle interactions between variables. In addition, as an ensemble machine learning approach, RF can address the issue of over-fitting. The application of RF models for variable selection and sensitivity analysis tasks is discussed by Genuer et al. (2010) and Antoniadis et al. (2021). We use the RF implementation made available through the scikit-learn library (Pedregosa et al., 2011).

Briefly, our methodology begins with the construction of a set of data with vectors of explanatory variables, or features, and vectors of responses. Here, the features are defined to include the values of the closure parameters ($c_v, c_n, C_n, R_i, n_p, z_p$) as well as discrete categorical features: the day/date (July 22, 2016 or August 21, 2016); site (PS-12 or Wasco), resolution (50 m or 150 m); and height level (50 m, 80 m, 120 m, and 190 m). The responses are various quantities computed from the WRF-LES simulation data, many of which have already been presented in Section 5.

The method proceeds by growing regression trees using a subset of the data referred to as the training data. The other data is reserved as testing data to allow the accuracy of the RF model to be assessed. Both the training and testing accuracy are expected to be satisfactory (e.g., greater than 90%) before we finalize a model. Once the RF model has been constructed, it can be used to assess the importance of any feature. RF FI is an impurity-based measure of importance that looks at the relative depth at which different features are used as decision nodes in the trees composing the RF. FI might be biased as it can overestimate the importance of continuous or numeric features relative to discrete features with only a few possible values (Strobl et al., 2007). The RF FI technique can be complemented by an alternative, more general (i.e., non-specific to RF models) procedure called permutation importance (PI) ranking, where feature values are randomly permuted and the corresponding change in the model score is computed. The greater the decrease in the model’s score, the more important the permuted parameter is deemed to be. We present FI and PI results and find them to be largely consistent. Additionally, we tested the alternative but related meta-modeling approach called gradient tree boosting (Friedman, 2001), as implemented via the XGBoost software library (Chen & Guestrin, 2016), as a verification of the RF analyses. Although these results are not presented here, they support the main conclusions of the RF analyses.
Figure 9. Dependence of the wind shear (evaluated between 50 and 80 m) on perturbed parameter values for (a) the July 22, 2016 case study and (b) the August 21, 2016 case study. Plotting conventions as in Figure 6: results of the finer, 50-m (coarser, 150-m) resolution domain are shown by blue (orange) markers. The observed values are indicated by black horizontal lines.
To verify that differences in FI between the full and halved ensembles were not simply due to changing the number of samples, we examined FI scores of $c_k$ obtained from 40 randomly selected sub-ensembles of 32 parameter combinations each. For all response variables examined here but one, the FI of $c_k$ for the full ensemble is within the interquartile range of FI for the random sub-ensembles. The exception is the turbulence timescale, for which the full ensemble’s $c_k$ FI is within the range of FI for the random sub-ensembles. However, the FI of $c_k$ for our selected sub-ensemble lies beyond or at the edge of the range of variability of FI scores of the random sub-ensembles for all response variables except the mean wind veer, and it is outside the interquartile range for all response variables.

Sensitivities in prediction of $U_{10\text{min}}$ are examined in terms of the first and second moments of the wind speed distribution. Results are shown in Figure 11, where FI results (summing to unity) are shown by blue and orange bars, respectively, for the full and halved (i.e., the parameter combinations shown in Figure 10) ensembles and feature rankings using PI are indicated by the numerical annotations of the bars. Note that the testing accuracy of the RF model for wind speed variance of the halved ensemble was lower (77%) than for other response variables. Predictions of the mean of $U_{10\text{min}}$ is (unsurprisingly) most sensitive to the case date. For the full ensemble, the importance of $c_k$ is similar to that of the site or height. The FI of $c_k$ is diminished in the sub-ensemble, but the eddy viscosity coefficient remains the most important of the parametric features. PI rankings are consistent with the FI values. Turning to the variance of $U_{10\text{min}}$ shown in the lower panel of Figure 11, $c_k$ has the greatest FI ranking among all features in the full ensemble (followed by the categorical features) and remains important in the sub-ensemble. PI ranks $c_k$ slightly lower among the most important features (consistent with the known biases of FI), but nonetheless clearly confirms that sensitivity to $c_k$ is much greater than for any other parametric features. This is also consistent with Figure 6.

Figure 12 displays FI and PI results for the turbulence quantities computed from 1 Hz timeseries. The results for $E_{<10\text{min}}$ (top panel) and the integral timescale (bottom panel) are qualitatively similar to each other and also follow a similar pattern to the $U_{10\text{min}}$ variance results. The key aspect to note is that $c_k$ shows high importance, even in the sub-ensemble with reduced range of $c_k$, that is comparable to or greater than the importance of the categorical features. Sensitivity to the remaining parameters is low, confirming the earlier qualitative analysis (i.e., Figures 7 and 8).
Sensitivities in the prediction of vertical derivatives in wind speed (shear) and direction (veer) are presented in Figure 13. The “height” feature for shear and veer has a different interpretation than for the other response variables, as shear and veer are computed as the differences in wind speed and direction between a pair of heights. We use three such pairings: 50 and 80 m (allowing comparison to the data of the sonic anemometers at the PS-12 site), 50 and 120 m, and 120 and 190 m. The latter two height pairings are selected to represent the extremes of shear and veer across a turbine blade during its rotation, given that modern, large land-based wind turbines have hub heights around 120 m and blade lengths around 70 m. Not surprisingly, the height feature is uniformly

![Figure 11.](image)

Figure 11. Evaluation of variable importance for (a) mean and (b) variance of 10-minute averaged wind speed values. Blue (orange) bars show random forest feature importance values for the full (halved) ensemble. Numerical annotations above each bar show the variable's ranking measured by permutation importance, with 1 being most important and 10, least.

![Figure 12.](image)

Figure 12. Evaluation of variable importance for (a) sub-10 minute turbulent kinetic energy and (b) turbulence integral timescale. Plotting conventions as in Figure 11: blue (orange) bars show random forest feature importance values for the full (halved) ensemble. Numerical annotations above each bar show the variable's ranking measured by permutation importance, with 1 being most important and 10, least.
identified as the most important feature for prediction of shear for both full and halved ensembles using either FI or PI. Veer, however, is quite strongly influenced by the site, which indicates an importance of local topographic features. Continuing the established pattern, $c_k$ dominates the importance of the parametric features for both shear and veer.

Ranking of variables in terms of FI and PI scores is informative regarding which variables should be considered influential, but is not conclusive. In other words, features that show lower relative importance may still be important in an absolute sense. There is also the possibility that an importance score is not a reliable indicator of the feature's real influence (e.g., due to insufficient sampling of values in the simulation ensemble). To address these potential concerns, two additional analyses were performed using the full ensemble data.

The goal of the first analysis is to determine a threshold for PI scores to indicate a greater-than-random influence on the response variables. The testing procedure consists of replacing an input feature with a dummy feature with randomly assigned values (here, draws from a normal distribution) and computing the PI of this dummy feature as a reference. As Table 2 summarizes, PI values of the “real” input features date, site, resolution, and $c_k$ exceeded the dummy variable’s PI for all the response variables discussed here. Additionally, the height input feature has greater PI than the dummy variable for mean and variance of the 10-min wind speed and for the wind shear. The Prandtl number exponential factor $n_P$ has PI greater than the dummy for variance of the 10-min wind speed, sub-10-min TKE, and turbulence integral timescale.

The second analysis is aimed at gauging which FI scores show statistical significance, such that comparison of values among the input variables to ascribe relative importance is soundly based. This analysis is motivated by the fact (as detailed, e.g., by Wang et al., 2016) that variable importance measures can suffer from instability due to the intrinsic randomness in the construction of an RF model, even once variations due to data selection and

Figure 13. Evaluation of variable importance for wind (a) shear and (b) veer. Plotting conventions as in Figure 11: blue (orange) bars show random forest feature importance values for the full (halved) ensemble. Numerical annotations above each bar show the variable's ranking measured by permutation importance, with 1 being most important and 10, least.
RF model parameter choices have been set aside. Here, we generate 100 sample data sets for each input feature and response variable combination, in which the values of one input feature have been randomly permuted while leaving the other input features as in the original data set. An RF model is trained for each of the 100 sample data sets and FI scores are computed. We then compare the 95th percentile FI of the permuted feature to the FI of that feature in the original data set (the “original” FI). If an input feature’s original FI is greater than the 95th percentile FI from the randomized samples, the input feature’s FI score is deemed to be significant for that response variable. We note the 95th percentile threshold is not theoretically based, but rather reflects a heuristic choice. If an input feature’s original FI is less than the threshold, this does not necessarily imply that the feature is unimportant. Rather, it indicates that the FI score should be interpreted with caution and that additional data (such as the PI scores discussed above) should be used to supplement the assessment of that feature’s importance. The results of this analysis are summarized in Table 3, and are largely consistent with those shown in Table 2. Date, site, height, resolution, and \( c_k \) importance scores show significance for most of the response variables. In contrast to the analysis of PI scores, FI scores for \( c_{\epsilon N} \) and \( R_i \) show significance for the sub-10 min TKE and turbulence integral time. For the latter quantity, the FI of \( z_f \) also shows significance.

We interpret these results as confirming that \( c_k \) should be prioritized among the model parametric input features for sensitivity testing, but that other parametric features should be retained for consideration whenever it is feasible to do so. This is especially true for applications that depend on higher order turbulence statistics or when considering atmospheric conditions that differ from those considered here, especially different atmospheric stability conditions. In particular, the influence of the turbulent Prandtl number specification cannot be totally dismissed in our results. Furthermore, unless only specific dates, measurement sites, model resolutions, or vertical levels need to be considered for an application, these factors should also be prioritized in simulation sensitivity analyses.

### Table 2

| Response variable | \( c_{\epsilon N} \) | \( R_i \) | \( c_k \) | \( c_n \) | \( z_f \) | \( n_p \) | Date | Site | Resolution | Height |
|-------------------|----------------------|--------------|------------|------------|--------|-----------|------|-----|-----------|--------|
| \( U_{10\min} \) mean | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| \( U_{10\min} \) variance | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| Mean wind shear | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| Mean wind veer | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| \( E_{<10\min} \) | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |

*Note: A filled square (■) indicates that the input feature's permutation importance is greater than that of the dummy variable; an unfilled square (□) indicates that the permutation importance of the dummy variable is greater.*

### Table 3

| Response variable | \( c_{\epsilon N} \) | \( R_i \) | \( c_k \) | \( c_n \) | \( z_f \) | \( n_p \) | Date | Site | Resolution | Height |
|-------------------|----------------------|--------------|------------|------------|--------|-----------|------|-----|-----------|--------|
| \( U_{10\min} \) mean | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| \( U_{10\min} \) variance | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| Mean wind shear | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| Mean wind veer | □ | □ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| \( E_{<10\min} \) | ■ | ■ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |
| Timescale | ■ | ■ | ■ | □ | □ | □ | ■ | ■ | ■ | ■ |

*Note: A filled square (■) indicates that the input feature's feature importance in the original data set exceeds the 95th percentile of FI from the permuted samples; an unfilled square (□) indicates the opposite, and suggests that input feature's FI score for the given response variable should be interpreted with caution.*
7. Conclusions

We created a collection of 128 mesoscale-to-microscale coupled model runs using the grid nesting capability of the WRF framework. The simulation ensembles are comprised of 64 combinations of perturbed parameters related to the 1.5-order SGS turbulence closure originated by Deardorff (1980) that are used to simulate two case studies from the WFIP2 campaign. Due to the large-scale weather pattern in the Columbia Basin region on these two dates associated with marine intrusion events, both case studies exhibit strong westerly winds in excess of 10 m s\(^{-1}\) and strong surface fluxes. However, spatial organization of turbulence and other detailed features differed between the 2 days in both observations and simulations. From the simulation data, we computed a number of quantities relevant to wind energy associated with wind speed, direction, turbulence characteristics, and vertical variations of the wind profile.

Qualitative analysis of the results as well as quantitative assessments of FI using methods based on RF meta-modeling were performed. Both these approaches highlighted that, among all the tested parameters, the eddy viscosity parameter \(c_k\) stands out in importance. Even when the original 64 parameter combinations were subset to 32 (roughly having the range of \(c_k\)), \(c_k\) remains more influential than any other parameter, and it is often of comparable or greater importance than categorical features such as horizontal grid spacing of the nested domain or height level at which the quantity is computed. The only feature that consistently exceeds \(c_k\) in importance is the case study date. This finding reflects the important role of \(c_k\) within the turbulence modeling approach, where the magnitude of the eddy viscosity affects both the dissipation of resolved-scale turbulent fluctuations and the closure of shear and buoyant production of SGS TKE. Our findings also indicate WRF’s default value of this parameter, 0.15, is larger than ideal for these cases and a value around 0.1 (as originally used by Deardorff (1980) and commonly recommended for other LES models) yields better results.

In addition to ranking input features according to their PI and FI scores obtained from RF meta-modeling, further tests were performed to verify the significance of these importance scores relative to randomized inputs. The combined implications of the PI and FI analyses confirm the significance of \(c_k\)’s importance scores for all response variables considered here, and also indicate that a few or all of the categorical features have significant importance scores for each of the response variables. The FI and PI of some other SGS closure parameters (in particular, the parameter related to the turbulent Prandtl number) show significance for second-order flow statistics, such as the wind speed variance, but do not meet our significance criteria for mean flow statistics.

Lest it be concluded that the utility of LES is degraded by this sensitivity to the eddy viscosity coefficient; we note that our results indicate that the range of \(c_k\) values which is likely to yield satisfactory results is actually rather narrow and is consistent with recommendations in the literature for idealized LES, even though these are highly non-idealized simulations. This suggests that a useful suite of sensitivity tests can be conducted with a limited number of ensemble members. Furthermore, by assessing the sensitivity of our simulations using RF FI and PI metrics, we are able to provide context to the significance of parametric sensitivities with information on the importance of categorical factors such as the case study date, specific measurement site, and horizontal grid spacing (resolution). In particular, the former two factors show high importance for most of the response variables we evaluate, which argues there is a high value to accounting for mesoscale variability and details of terrain by performing realistic, mesoscale-to-microscale coupled simulations.

Finally, further work is needed to examine how these parameter sensitivities and FI rankings change under a wider array of case studies, and especially considering conditions of stable stratification such as might be encountered nocturnally over land or offshore. We also look toward combined analyses of mesoscale and LES sensitivities to more comprehensively understand the uncertainties of mesoscale-to-microscale coupled modeling, and especially those sensitivities that are most important for wind energy applications.

**Data Availability Statement**

Observational and simulation data used in the preparation of this article are available from the Atmosphere to Electrons Data Archive and Portal (https://a2e.energy.gov/about/dap). The sampled values of the perturbed parameters and example namelist files for the WRF simulations are also provided as supplementary information to the simulation data sets. We thank the DAP team for their assistance in archiving the simulation data.
Acknowledgments
This research was supported by the Wind Energy Technologies Office within the Office of Energy Efficiency and Renewable Energy of the U.S. Department of Energy. The Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830. Computational resources used to perform this work were sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory. We thank the WFIP2 event loggers (Justin Sharp, Jaynes Kenyon, Bob Banta, Eric Grimit, Kyle Wade, Mark Stoelinga, Qing Yang, Eric James, Kyle Wade, Clara St. Martin) for their effort in cataloging events. The WFIP2 study was also made possible by a large number of individuals involved in the deployment of the measurement equipment. We also thank our colleagues on the Mesoscale-Microscale Coupling project for many helpful interactions in the course of this study, as well as the anonymous reviewers of this work whose comments led us to make valuable improvements.

References
Antoniadis, A., Lambert-Lacroix, S., & Poggi, J.-M. (2021). Random forests for global sensitivity analysis: A selective review. Reliability Engineering & System Safety, 206, 107312. https://doi.org/10.1016/j.ress.2020.107312
Berg, L. K., Liu, Y., Yang, B., Qian, Y., Olson, J., Ma, P.-L., & Hou, Z. (2019). Sensitivity of turbine-height wind speeds to parameters in the planetary boundary-layer parameterization used in the weather research and forecasting model: Extension to wintertime conditions. Boundary-Layer Meteorology, 170, 507–518. https://doi.org/10.1007/s10546-018-0406-y
Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324
Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785–794). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
de Rooide, S. R., Jonker, H. J. J., van de Vel, B. J. H., Vertregt, V., & Perrin, V. (2017). A diagnosis of excessive mixing in Smagorinsky subfilter-scale turbulent kinetic energy models. Journal of the Atmospheric Sciences, 74, 1495–1511. https://doi.org/10.1175/JAS-D-16-0212.1
Deardorff, J. W. (1980). Stratocumulus-cumulus mixed layers derived from a three-dimensional model. Boundary-Layer Meteorology, 18, 495–527. https://doi.org/10.1002/qb19502
Dyer, A. J., & Hicks, B. B. (1970). Flux-gradient relationships in the constant flux layer. Quarterly Journal of the Royal Meteorological Society, 96, 715–721. https://doi.org/10.1002/qj.49709641012
Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29, 1189–1232. https://doi.org/10.1214/ aoas1033451
García-Sánchez, C., & Gorlì, C. (2018). Uncertainty quantification for microscale CFD simulations based on input from mesoscale codes. Journal of Wind Engineering and Industrial Aerodynamics, 176, 87–97. https://doi.org/10.1016/j.jewea.2018.03.011
Genner, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. Pattern Recognition Letters, 31, 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014
Gibbs, J. A., & Fedorovich, E. (2016). Sensitivity of turbulence statistics in the lower portion of a numerically simulated stable boundary layer to parameters of the Deardorff subgrid turbulence model. Quarterly Journal of the Royal Meteorological Society, 142, 2205–2213. https://doi.org/10.1002/qj.2818
Gorlì, C., García-Sánchez, C., & Iaccarino, G. (2015). Quantifying inflow and RANS turbulence model form uncertainties for wind engineering flows. Journal of Wind Engineering and Industrial Aerodynamics, 144, 202–212. https://doi.org/10.1016/j.jewia.2015.03.025
Grachev, A. A., Andreas, E. L., Fairall, C. W., Guest, P. S., & Persson, P. O. G. (2013). The critical Richardson number and limits of applicability of local similarity theory in the stable boundary layer. Boundary-Layer Meteorology, 147, 51–82. https://doi.org/10.1007/s10546-012-9771-0
Haupt, S. E., Kosovic, B., Shaw, W., Berg, L. K., Churchfield, M., Cline, J., et al. (2019). On bridging a modeling scale gap: Mesoscale to microscale coupling for wind energy. Bulletin of the American Meteorological Society, 100(12), 2533–2550. https://doi.org/10.1175/BAMS-D-18-0033.1
Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety, 81, 23–49. https://doi.org/10.1016/S0951-8320(03)00054-9
Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. Monthly Weather Review, 134, 2318–2341. https://doi.org/10.1175/MWR3199.1
Hornshøj-Møller, S. D., Nielsen, P. D., Forooghi, P., & Abkar, M. (2021). Quantifying structural uncertainties in Reynolds-averaged Navier–Stokes simulations of wind turbine wakes. Renewable Energy, 164, 1550–1558. https://doi.org/10.1016/j.renene.2020.10.148
Januzzi, J. A. (1993). The onshore push of marine air into the Pacific Northwest. Monthly Weather Review, 121, 194–202. https://doi.org/10.1175/1520-0493(1993)008<0194:TOPOMA>2.0.CO;2
Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-Bustamante, E. (2012). A revised scheme for the WRF surface layer formulation. Monthly Weather Review, 140, 898–918. https://doi.org/10.1175/MWR-D-11-00056.1
Kadavreisky, R., Sharma, A., Matić, C., & Binzawile, R. (2019). High resolution urban air quality modeling by coupling CFD and mesoscale models: A review. Asia-Pacific Journal of Atmospheric Sciences, 55, 539–556. https://doi.org/10.1007/s13143-019-00110-3
Lilly, D. K. (1967). The representation of small-scale turbulence in numerical experiments. In Proc. IBM scientific computing symposium on environmental sciences, Yorktown Heights, New York (pp. 195–210). Thomas J. Watson Research Center.
Lucor, D., Meyers, J., & Sagaut, P. (2007). Sensitivity analysis of large-eddy simulations to subgrid-scale-model parametric uncertainty using polynomial chaos. Journal of Fluid Mechanics, 585, 255–279. https://doi.org/10.1017/S0022112006076751
Mass, C. F., Albright, M. D., & Brees, D. J. (1986). The onshore surge of marine air into the Pacific Northwest: A regional aspect of complex terrain. Monthly Weather Review, 114, 2602–2627. https://doi.org/10.1175/1520-0493(1986)114<2602:osoma>2.0.co;2
Moeng, C.-H., & Wyngaard, J. C. (1988). Spectral analysis of large-eddy simulations of the convective boundary layer. Journal of the Atmospheric Sciences, 45, 3573–3587. https://doi.org/10.1175/1520-0469(1988)045<3573:sslaodc>2.0.co;2
Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada Level-3 model: Its numerical stability and application to a regional prediction of advection fog. Boundary-Layer Meteorology, 119, 397–407. https://doi.org/10.1007/s10546-005-9030-8
Nakanishi, M., & Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer. Journal of the Meteorological Society of Japan, 87, 895–912. https://doi.org/10.2151/jmsj.87.895
Nieuwstadt, F. T. M. (1990). The critical Richardson number and limits of applicability of local similarity theory in the stable boundary layer. Boundary-Layer Meteorology, 165, 69–89. https://doi.org/10.1007/s10546-016-0217-y
Rodrigo, J. S., Arroyo, R. A. C., Wirth, B., Dörenkämper, M., Gottschall, J., Avila, M., et al. (2020). The new European wind atlas model chain. Journal of Physics: Conference Series, 1452, 012087. https://doi.org/10.1088/1742-6596/1452/1/012087
Safa, C., Bloylock, M., Templeton, J., Domino, S., Sargsyan, K., & Njum, H. (2017). Uncertainty quantification in LES of channel flow. International Journal for Numerical Methods in Fluids, 83, 376–401. https://doi.org/10.1002/fld.4272
Salesky, S. T., Chamecki, M., & Bou-Zeid, E. (2017). On the nature of the transition between roll and cellular organization in the convective boundary layer. *Boundary-Layer Meteorology, 163*, 41–68. https://doi.org/10.1007/s10546-016-0220-3

Schumann, U. (1991). Subgrid length-scales for large-eddy simulation of stratified turbulence. *Theoretical and Computational Fluid Dynamics, 2*, 279–290. https://doi.org/10.1007/BF00271468

Shaw, W. J., Berg, L. K., Cline, J., Draxl, C., Djalalova, I., Grimst, E. P., et al. (2019). The Second Wind Forecast Improvement Project (WFIP2): General overview. *Bulletin of the American Meteorological Society, 100*(9), 1687–1699. https://doi.org/10.1175/BAMS-D-18-0036.1

Silveira, C., Ferreira, J., & Miranda, A. I. (2019). The challenges of air quality modeling when crossing multiple spatial scales. *Air Quality, Atmosphere Health, 12*, 1003–1017. https://doi.org/10.1007/s11869-019-00733-5

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). A description of the Advanced Research WRF version 4. NCAR Technical Note NCAR/TN-556+STR (Tech. Rep.). https://doi.org/10.5065/1dfh-bp97

Smith, N. H. (2018). A sensitivity study on wind ramp events in the Columbia River Basin (Unpublished doctoral dissertation). Lubbock, TX: Texas Tech University.

Smith, N. H., & Ancell, B. C. (2019). Variations in parametric sensitivity for wind ramp events in the Columbia River Basin. *Monthly Weather Review, 147*, 4633–4651. https://doi.org/10.1175/MWR-D-19-0019.1

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics, 29*, 143–151. https://doi.org/10.1080/00401706.1987.10488205

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources, and a solution. *BMC Bioinformatics, 8*, 25. https://doi.org/10.1186/1471-2105-8-25

Wang, H., Yang, F., & Luo, Z. (2016). An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics, 17*(60). https://doi.org/10.1186/s12859-016-0900-5

Wilczak, J. M., Stoelinga, M., Berg, L. K., Sharp, J., Draxl, C., McCaffrey, K., et al. (2019). The Second Wind Forecast Improvement Project: (WFIP2): Observational field campaign. *Bulletin of the American Meteorological Society, 100*, 1701–1723. https://doi.org/10.1175/BAMS-D-18-0035.1

Xiao, H., & Cinnella, P. (2019). Quantification of model uncertainty in RANS simulations: A review. *Progress in Aerospace Sciences, 108*, 1–31. https://doi.org/10.1016/j.paerosci.2018.10.001

Yang, B., Berg, L. K., Qian, Y., Wang, C., Hou, Z., Liu, Y., et al. (2019). Parametric and structural sensitivities of turbine-height wind speeds in the boundary layer parameterizations in the Weather Research and Forecasting model. *Journal of Geophysical Research: Atmospheres, 124*, 5951–5960. https://doi.org/10.1029/2018JD029691

Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., et al. (2017). Sensitivity of turbine-height wind speeds to parameters in planetary boundary-layer and surface-layer schemes in the Weather Research and Forecasting model. *Boundary-Layer Meteorology, 162*, 117–142. https://doi.org/10.1007/s10546-016-0185-2