Arthur et al

Evaluating mesoscale model predictions of diurnal speedup events in the Altamont Pass Wind Resource Area of California

Manuscript wes-2024-137

<u>Introduction</u>

This manuscript presents a comparison of season-long (~75 days) simulations of atmospheric flow at a wind power production site located in hilly terrain in central California. Simulations are performed with the WRF model at 1-km horizontal grid spacing and with two alternative parameterization schemes for planetary-boundary layer turbulence. The first PBL parameterization is the well-known and widely adopted MYNN scheme, which models only vertical mixing and additionally relies on Smagorinsky-type horizontal diffusion. The second PBL parameterization is a recently developed three-dimensional scheme that represents horizontal mixing in a physically consistent manner. Given the complex-terrain nature of the site, it is expected that horizontal heterogeneity of the turbulence could be misrepresented in a purely one-dimensional parameterization. Therefore, it is expected that the 3D PBL scheme has greater forecast skill.
This wind farm features prominent diurnal variability of the wind field, with speed-ups in the late afternoon/evening driven by differential heating of the atmosphere. Average diurnal cycles of the vertical profiles of wind speed, wind direction, temperature, vertical velocity, turbulent kinetic energy at three measurement sites are derived from the simulations and compared with observations from two wind lidars and a meteorological tower. The analysis deals in depth with data from one lidar site and from the tower. The primary purpose of the analysis is to evaluate if the new 3D PBL parameterization compares more favourably with observations. A secondary purpose is to verify that coupling the new 3D PBL scheme with an existing wind farm parameterization provides reasonable results.

<u>Recommendation</u>

The manuscript is well written and contains clear and well-designed figures. The research is clearly relevant to the scope of the journal. The focus on boundary-layer modelling over complex terrain makes this work interesting in a broader context beyond wind energy, e.g., for mountain meteorologists. In fact, the introductory chapter is an excellent concise crash course on the challenges of boundary-layer modelling over mountains.
The analysis is generally solid. The only aspect I find questionable is a degree of confusion between the systematic and random components of model error (see comments 1-2-3 below). The results are probably less compelling than one might expect, in that the impact of the new turbulence parameterization is rather weak (see comments 4-5). In fact, the most relevant systematic errors (persistent wind speed underestimation near the ground, and overestimation across the rotor diameter during the evening speed-up) are corrected only marginally by the new scheme. However, the small difference in the skill of wind forecasts translates into a somewhat more marked improvement of the power production.
The manuscript does not break new scientific ground, but the results will likely be of interest for those who use WRF simulations for wind resource forecasting and assessment. Acceptance is recommended, conditional to minor revisions.

<u>General comments</u>

1. Equation 1 defines the “model bias”. An instantaneous, local deviation between model and observation is NOT bias. At a given time and place, model error has both a systematic and a random component. Strictly speaking, bias is the *systematic* component of the deviation between model and observations. Thus, it is the *average* difference evaluated over a sample. I would recommend adding an averaging operator to the RHS of the equation. Averaging can consider multiple dimensions (e.g. time only, or time and height), and therefore I would recommend that the authors accurately define the averaging operations.

2. Line 278-279. Here I see another example of confusion between systematic and random error. “The expected maximum error is smaller than the standard deviation of the diurnal composite”. From the preceding text (around line 115), I understood that the estimate by Bingöl et al refers to the systematic (mean) component of the model error (accuracy). In this figure, it seems to be used as a standard deviation (precision). I’m not sure this is appopriate, because the standard deviation represents deviations from the mean; thus the random component of the model error, not the systematic one.
   A side note: There are two shades of gray around the observed profiles in Figure 4. The potential 10% error in the observations is likely the narrowest shaded area (plus-minus 1 m/s). The broadest one (plus-minus 3 m/s) likely represents the standard deviation. Please clarify.
3. Figure 11 and line 392. Here I see yet another example of misinterpretation of model error statistics. The mean absolute error of the wind speed is regarded as a measure of bias. It is not! The mean absolute error conveys essentially the same information as the root mean square error, but using a different norm: the absolute value norm instead of the Euclidean norm. A sensible measure of bias would be the mean error.
   Furthermore, I’m not sure Fig. 11 provides useful information. The authors argue that differences in “bias” among the simulations with the two schemes explain errors in the modelled capacity factors. I would buy the argument if the dots (one dot every 10 minutes in a full diurnal cycle) were well aligned on the diagonal. However, quadrants I and III together contain roughly the same number of dots as quadrants II and IV together.
   All considered, I would recommend removing this figure.
4. A general remark on Fig. 4, 5, 7, 8: the differences between MYNN and 3D PBL simulations are always very small in comparison to the respective deviation from the observations. This applies to the mean wind speed profiles (Fig. 4), temperature profiles (Fig. 5), areal distribution of the wind speed bias (Fig. 7), diurnal cycle of the wind speed bias (Fig. 8). This fact could be interpreted in two ways: either (i) the PBL scheme is a minor contributor to total model error; or (ii) neither of the two PBL schemes is able to successfully reduce model error. None of these conclusions is particularly encouraging. Can the authors comment on that?
5. The conclusions (line 450-454) state that “several notable differences were found between PBL treatments” and that “the 3D PBL scheme showed evidence of a more pronounced near-surface jet and reduced wind speeds aloft”. There are differences, indeed, but honestly they seem rather minor (see comment 4). It would be fair acknowledge this fact, and to discuss openly the possible reasons for the limited impact of the new PBL scheme. Also, it could be useful to point out that, because wind power is proportional to the cube of wind speed, small relative improvements in modelled wind speed translate in noticeable improvements in modelled power production (or modelled capacity factor).

<u>Specific comments</u>

6. Line 119: I guess the “dynamic conversion factors” are a way to quantify systematic errors in wind speed simulations. Could the authors clarify, and maybe use one or two sentences to explain how these corrections are computed?
7. Lines 158-159: My understanding of the boundary-layer approximation, which usually applies to horizontally homogeneous boundary layers, is that all horizontal gradients of mean quantities (and the turbulent fluxes that are assumed to be proportional to them) are neglected. You probably mean something more subtle here. To help the reader, could you please spend a few more words to clarify?
8. Lines 168-169: “positive-definite 6th order diffusion”. It would be good to specify that this (pseudo-)horizontal diffusion is computational, not physical. I guess it is used to help dissipate 2dx noise and maintain numerical stability.
9. Line 215 explains that bias is computed also for the wind direction. It is a delicate operation, because of the cyclic nature of the variable. Consider the case of an observed timeseries of wind directions such as 0, 1, 359; and a modelled timeseries such as 358, 359, 1. The

model is actually very accurate (the error is 2 degrees at most), but computing the mean error without accounting for the periodicity yields values around 180 degrees. Could the authors explain how they circumvent this problem?

10. Figure 3: In looking at this figure I wondered if the slopes of the modelled and true terrain are similar or not. The text says (line 88) that the WOP site is on a ridgeline, but the modelled orography in Fig. 1 looks like an eastward facing slope. Low-level wind vectors are subject to a parallel flow condition, and if the modelled slope geometry is somewhat inaccurate, biases in the u and w components will inevitably follow. This aspect might be worth commenting.

    Furthermore, the authors state that “while the model captures some negative vertical velocities at the study site during the speedup events, its vertical velocities are too weak and thus do not translate to near-surface accelerations of the magnitude seen in the observations.” I understand that this reasoning is based on incompressible mass continuity arguments. However, the vertical convergence/horizontal divergence concept is seemigly only relevant to the evening and night hours. Fig. 3b instead shows that the horizontal wind speed at low levels is underestimated also at daytime (when vertical velocity is predominantly upward). Is there a different explanation for the daytime bias?
11. Line 255: “The vertical velocity error is not normalized because w has both positive and negative values”. I'm not sure I understand. You mean that, by normalizing, you would likely run into divisions by zero in Eq 2 and 3?

<u>Technical corrections</u>

12. Line 87: “Diablo Range”. This geographical name is unexplained, and is unlikely to be broadly known. Please label the site in Figure 1, or describe its position in the figure caption.
13. Figure 1: It may be a pet peeve of mine, but I'm allergic to terrain colormaps with deep blue shades in inland areas. It looks like the figures were plotted with python/matplotlib. If so, it is fairly straightforward to truncate colormaps; e.g. using only the upper 80% of the range (green and brown shades).
14. Figure 2, caption: “Capacity factors”, first mentioned here. It might be worth to explain the term. It is a basic concept, but casual readers might not know it.
15. Line 128: Reference to “(Synoptic, 2023)”. This is likely a broken reference. See also line 323.
16. Table 1: Symbols (Mfr-PR, H, D) have rather obvious meanings, but it would be good to explain them in the caption.