Review comments

The paper is generally well written and structured. The figures, results, and comparisons between simulations and reference data are strong. I compliment the authors for the framework and the simulation effort. However, the title, the claims, and the arguments in the discussion and conclusions are not sufficiently supported by the presented results.

If the study were reframed to focus on a more specific aspect—such as *"Effect of the added TKE on wake recovery in the Fitch parameterization"* (to give an example)—I will immediately recommend acceptance. In its current form, the manuscript does not fully reflect what is stated in the title and abstract. Substantial content changes are needed to align objectives, title, and conclusions.

Below is a list of major and specific comments:

Major comments

1. Throughout the paper, mesoscale resolution is presented as the main cause of what is referred to as "under-resolved" gradients. This interpretation is problematic because coarse resolution effects are expected—you can not resolve something that is smaller than your grid. Mesoscale simulations are indeed coarser than LES, but they resolve gradients at their own grid scale. What is described as "under-resolved" actually corresponds to sub-grid-scale features, which should primarily be handled by the wind farm parameterization (WFP) and the local PBL diffusivity. Ideally, the applied force should account for the local wake expansion and turbine interactions, and then be consistently integrated over the grid cell in a way that accounts for grid-scale wake effects. In the current Fitch WFP, momentum is injected without these considerations. Therefore, the term "under-resolved" seems incorrectly applied to which is a flaw in the WFP.

2. In section 4.1, the argument about reduced eddy diffusivity, reduced shear, and less TKE is reasonable, but it would be much stronger if supported by figures. For example, showing eddy diffusivity and shear production (or other MYNN budget terms such as dissipation and advection) would help visualize the interplay you describe. I am not suggesting LES plots comparisons—those simulations are expensive—but mesoscale plots would add clarity and really sustain what you described. You can support the statement of *sustained TKE levels via shear production* for example.

3. Why only Fitch? Including other WFPs, such as Jensen (which claims to represent sub-grid wakes) or Abkar et al. (2015), would provide valuable insights. I am suggesting the latter, because the force applied in the "BASE" experiment is not the same as in the LES even if you are using a C_t constant in both cases. In Abkar you can control how much force (and consequently TKE) you apply. And if you apply the same force in the mesoscale, would that result in a better recovery? Limiting the study to Fitch sensitivity reduces the scope.

4. I commend the authors for using idealized comparisons and matching inflow conditions between mesoscale and LES simulations, including the removal of inertial oscillations.

5. The manuscript does not mention that in WRF, eddy diffusivity (K_m) is treated separately in the vertical and horizontal directions. Vertical mixing is handled by the PBL scheme, while horizontal mixing is typically represented by a 2D Smagorinsky scheme or constant K_m under ideal conditions. Since horizontal gradients are discussed, it would strengthen the paper to either show the influence of horizontal K_m or argue why it is negligible compared to vertical mixing. If K_m in the vertical is the only important thing then argue around the PBL scheme.

Specific comments

1. Some statements are vague. For example, line 346: *"this resemblance is misleading and fails to represent important physical processes"*—which important processes? Be specific. Similarly, line 288: *"The bias improves"* should be rephrased as *"The bias decreases"* for clarity.
2. Lines 316–320: The paragraph suggests that TKE dissipation inhibits accumulation in LES, yet TKE100 shows more TKE than LES. Please be specific.
3. Lines 280–282: What is meant by "momentum entrainment"? Is this a momentum sink or entrainment for wake recovery? Please clarify.
4. Treatment of "spikes": These are resolved features. The only valid comparison with mesoscale simulations is the coarsened LES, which looks very good in Fig. 8a,b.
5. The same for Fig. 9: Mesoscale results with WFP will naturally be worse than LES at the original resolution, except for inflow conditions, which only support the idea that the WFP is the problem.
6. Line 449: You state that slow wake recovery and fast TKE decay effects apply to other NWP and climate models using WFPs. However, in HARMONIE, Fitch performs well (see https://doi.org/10.1029/2021MS002947), possibly due to a different PBL scheme. You can comment in that line of the PBLs.