Review of: **Evaluation of a cloudy cold-air pool in the Columbia River Basin in different versions of the HRRR model**

This paper evaluates the performance of a suite of WRF-based simulations of cloud cold-air pools in the Columbia River basin. The study is motivated by a need to improve NWP for wind energy, especially during variations in stratification and windiness in the boundary layer associated with cold-air pools. Model performance, spanning variations in physics and resolution, are evaluated using WFIP2 observations. Changes in physics include variations if formulation of the mixing length, inclusion of mass-flux terms for non-local mixing, subgrid cloud processes, gravity wave drag, a wind farm parameterization, and differences in diffusion and filtering. Model results are examine for two forecast periods. Results indicate some improvement in developing low level clouds, yielding better agreement in near surface temperatures during the first simulation period, but also produced some errors in cloudiness and temperatures in the later parts of the event/simulations. This mixed result make the take aways from this paper somewhat unclear, though there is obviously a strong radiative impact of clouds on CAP structure and evolution that must be simulated correctly to get the correct forecast. The paper includes many modifications to the model between the different experiments, but not in a well-defined "sensitivity" study approach that would allow a detailed assessment of individual changes. This is understandable in that the authors are working with quasi-operational model codes, rather than a suite of idealized simulations. Nevertheless the well-constructed observation-to-model comparisons provide some physical and numerical insights into how to improve CAP simulations with respect to clouds, and offers some tentative insights into what future work is needed to further improve cloud process representation (e.g., entrainment mixing). I find the paper incrementally advances our understanding of and ability to simulate CAPs, and I do not find any fatal flaws in the approach or results. I therefore believe the paper should be published following addressing a handful of specific concerns detailed below.

Major points, the first substantive, the second editorial

(1) There needs to be some improved discussion of observation and retrieval uncertainties. These include: (a) potential biases in the TROPoe retrievals due to smoothing of the profile and (b) LWP estimates. This is important since there are some large temperature biases and the reader needs to have a basis for understanding the comparative magnitude of simulation biases and observational biases. Perhaps these points are addressed in some of the referenced literature, but some discussion of these points needs to be surfaced.
(2) Conciseness: I'd encourage the authors to try to trim as many words as possible from the manuscript. There is a lot decipher so trimming extra verbiage would be helpful. I have a few specific examples below.

Specific Comments (more major comments in **bold**):

Abstract. I had to read the abstract a number of times, having stumbled over the wording and meaning. A few examples:

a. The phase "also related to this" strikes me as odd.
b. "Differences between different model versions were in particular visible" -> could just be "Differences amongst model versions were apparent in simulated temperature and near-surface clouds"
c.

Line 49: "the challenges" -> "these challenges". More broadly this tough to read sentence, consider revising if possible.
Line 55: "Besides the conduction" -> "besides completing"? conduction implies a physical process (e.g., heat transfer)
Line 93: "adapted" -> do you mean adopted?
Line 148: "A wind farm parameterization" -> probably needs a citation here so we know the details of this. Also, I don't recall seeing any further discussion of this modification in the results, so is it really necessary to detail this? As it is I was expecting to see something about this later on.
Line 172-176: This is one location you could explain if there are expected biases or uncertainties in the observations/retrievals that would impact our interpretation of the model-obs biases discussed throughout.

Lines 183-202: This paragraph is long and could be cleaned up to make it easier to interpret. There are many opportunities to rephrase and reduce the word count in here. Just one minor example on line 196: "from the easterly direction" could just be "easterly".

Line 189-: "exhibited" rather than "experienced"?

**Lines 207-220**: The notable warm bias is evaluated against the observations, but there is no discussion here as to potential bias in the observations either here on lines 100-110, where TROPoe was discussed. How good are the TROPoe profiles compared against other data sources? Could some of the bias here be from issues with retrievals from the observations? From the observations themselves (e.g., RASS). Perhaps these are addressed in Adler et al. 2021… but I did not have time to dig into that. The TROPoe profiles look quite smooth… are they washing out sharp layers that then results in parts of the profile that are too warm while others are too cold?

Line 220: "could be related to an initialization error on this day"… these are the sorts of things that make it tricky to take clear messages away from this paper. There are a mish-mash of things causing the observed discrepancies. No fault of the authors, but diminishes the impact of these analyses.

Line 227-231: Redundant with my comment above, but do you have any radiosonde observations at this time to compare with the TROPoe profile and the model profiles. The TROPoe profile strikes me as getting some of the critical details but also likely smoothing the capping inversion structure… which could results in some of the warm bias in the model (though the model is obviously getting the low levels totally wrong)

Section 3.3: In general I really like this section, finding the analysis nicely done and the results relatively easy to interpret. One suggestion is to move some of the discourse on Lines 321-330 to the intro of this section so that the reader is better equipped to think about the expected differences between pseudo-profiles and free-air profiles (e.g., impact of near surface heating/cooling on station data as compared to free air data).

Figure 10. panels a,bc,d,e all have the same "Wasco" heading. This is somewhat confusing/not helpful. Maybe either give them a topical name a.) profiles, b.) biases, c) condensate profiles…. etc or just use a), b), c) with no descriptor.

**Line 368**: Again we are in need of knowing how well the radiometer observation measure the LWP. The model-observation analysis requires that we have at least some sense for how uncertain the observations are here. How does this LWP compare with radiosonde observations?

Line 446: strike "intense" and either remove the superlative or use a more physical descriptor (high cloud fraction, deep louds, large LWP, etc)

Line 453: "evaluated three different versions" -> "evaluated three versions"

**Line 465**: I'm again wondering if the "near isothermal" structure is an artifact of the TROPoe, whereas an actual sounding would show something close to moist adiabatic? Should we conclude that the subcloud layer is truly characterized by "near isothermal" conditions in the CAPs?

Line 491: "was equally well capture in v4fp2" and what? The control V2fp1? The other runs?

Line 500: "greatly improve" -> I'd remove this superlative here and just say "improved"

Line 502: Should "lagging" be "lacking"?

Line 502: Most "realistic"-> perhaps indicate how so? Better cloud fraction, better diurnal cycle, etc?

Line 503: I might make this point clearer in the introduction as well. This study is not a clear sensitivity study that specifically isolates one process at a time. Setting the readers expectation on this early on helps to frame the paper.

Lines 513-518. These are important points about the lack of entrainment mixing in EDMF for stratocumulus type cloud that you are making here. I wonder if you could introduce a few of these thoughts when discussing the model results for fp2 above.