Re-review of Venecek et al., (2018) “Predicted Ultrafine… In the Continental United States”

The authors of the manuscript have now completed a lengthy reply to the first round of reviews, and while successfully addressing some of the issues, the responses and changes do not fully address others and introduce some added concerns, leading me to recommend against publication in ACP.

Comment #1: First, the manuscript, in my opinion, does not comport with ACP policy “If the data are not publicly accessible, a detailed explanation of why this is the case is required.” Further, it states “Data do not comprise the only information which is important in the context of reproducibility. Therefore, Copernicus Publications encourages authors to also deposit software, algorithms, model code, video supplements, video abstracts, International Geo Sample Numbers, and other underlying material on suitable FAIR-aligned repositories/archives whenever possible. These materials should be referenced in the article and cited via a persistent identifier such as a DOI.” They, instead, state “The model source code and input data are available to collaborators through direct email request to the corresponding author.” Thus, it would appear, that the authors are unwilling to make the model and data available to others to assess if their results are reproducible, and as importantly, there are no errors in the data or the source code. On this latter point, it is important to note that it was via the initial review that an error that was identified: “An error in the model wind fields was corrected in the revised version of the manuscript. This error had caused the winds in each row to advance by one column, effectively moving the winds over the Pacific Ocean over land for coastal California cities such as Los Angeles. The same error was corrected in all domains, but the effects were less severe at inland locations where winds were more uniform. All of the model results throughout the revised paper now reflect correct wind fields (all simulations were rerun).” Without others being able to look over the inputs and the code, such errors can propagate forever. Indeed, one might expect that a similar error has been present in prior modeling using the same or predecessor codes to develop windfields. If this is the case, those journals must be contacted, and depending upon the journals discretion, erratum should be added to those prior publications. If research using fields where this is an issue have been used in health assessments and regulatory rule making (e.g., in California or by the US EPA), the agencies, themselves, should be contacted as to assess their further use. Given the potential importance of this issue, the authors need to reassure your journal and others, as well as the other users of this information, that a similar error was not present in the prior applications. Thus, in their response on-line, they would need to add an addendum to their current response stating that this issue was singular to this submission. While I realize that each journal has its own policies, I will note that the stated data policy by ACP is weaker than others like GMD, Science and Nature where the code would also have to be provided if requested to show that the results are reproducible and the code has no apparent errors.

Response: All emissions inputs, spatial surrogates, field measurements used to generate model inputs and evaluate model results were obtained directly from the EPA database / model clearing house. These data are available to anyone wishing to re-create the inputs. The size and composition distribution profiles used to generate emissions of ultrafine particulate matter are publicly available in peer-reviewed journal articles published over the past several decades. Once again, these data are available to anyone wishing to re-create the inputs. If fellow researchers do not wish to go to the trouble of accessing the publicly available information to assemble the model inputs for themselves, then the authors are also *willing to collaborate as stated in the manuscript. Please contact the corresponding author to discuss future collaborations.*

All model systems have the potential for errors. The process of continual model development combined with appropriate quality control checks identifies these errors so that they can be corrected. In the current manuscript, the error in the wind fields was associated with an updated program that formatted WRF output fields for the UCDF/CIT air quality model. This wind field error was not present in any previous manuscripts in ACP or in any other journal, since the prior version of the program did not contain the error. We invite fellow researchers to inspect our source code and rerun model simulations as a quality control check if they wish to collaborate on ultrafine particle simulation studies. Please contact the corresponding author to discuss future collaborations.

California and the US EPA do not base health assessments of regulatory rule making on a single model result. A weight of evidence approach is always used in these efforts. Independent measurements and/or model results would be needed to verify important findings before they are used in public policy decisions.

Comment #2: A second rather important error, particularly since it was pointed out in the prior review, is the issue about model evaluation. There are no EPA guidelines, and the cited reference Boylan and Russell(1), does not represent EPA policy. I do not believe either of those individuals represents EPA. To assert that the cited manuscript represents EPA guidelines is quite wrong. The authors would be better guided to look at the Simon et al., paper (2) as most of the authors of that paper are from EPA, though they don't seem to suggest that they are establishing EPA guidelines. This was pointed out in the first review.

Response: We acknowledge the point the Boylan and Russell manuscript does not represent official EPA policy even though these criteria are widely used for model evaluation studies. We have modified the manuscript to remove references to "EPA" when discussing performance criteria and now reference the Emery et al (2017) paper as discussed in comments below.

Comment #3: A further, extremely important issue is that the application of the Boylan et al. guidelines is also done incorrectly. That paper is only applicable to PM, not gaseous pollutants. To use the same suggested criteria for gaseous pollutants and PM is totally wrong. Further, they use this to suggest "excellent performance". This is where things get very concerning. They are misapplying suggested performance measures, to a subset of the model application (they keep saying 95% of the data: if you remove the worst 5% of the points, of course performance will improve: those metrics, I suspect, were not developed after removing the worst performing results: if so, the authors should so state), calling the guidelines "EPA", then saying "excellent" performance. In general, the authors tend to use very subjective superlatives rather casually and, I think, without justification.

Response: We apologize for not correctly updating the model performance evaluation in Figure 3 during the last revision. We have now updated Figure 3 and associated discussion based on the criteria suggested by Emery et al. (2017). As discussed in the response to the following comment, over 95% of the predicted daily maximum ozone concentrations meet the performance criteria of <25% NME.

Please note especially that the worst points were not "removed" from the analysis. We provide quantitative comparisons between model performance and published guidelines so that the readers can reach their own conclusion about the accuracy of the simulations.

Comment #4: A more realistic assessment of their model performance relative to others in North America is found from the Simon et al., article, which did not specify guidelines, or possibly a more recent article from researchers at Environ(2, 3) or the series from the AQMEII initiative, which has a more thorough set of approaches (4-11). The Simon/EPA article was a review of performance of ozone and PM modeling, again, not setting guidelines. Emery et al., reviewed the Simon et al., paper and others and provide the following suggested guidelines:

	NMB		NME		r	
Species	Goal	Criteria	Goal	Criteria	Goal	Criteria
1-hr or MDA8 Ozone	<±5%	<±15%	<15%	<25%	>0.75	>0.50
24-hr PM2.5, SO4, NH4	<±10%	<±30%	<35%	<50%	>0.70	>0.40
24-hr NO3	<±15%	<±65%	<65%	<115%	None	None
24-hr OC	<±15%	<±50%	<45%	<65%	None	None
24-hr EC	<±20%	<±40%	<50%	<75%	None	None

Response: The criteria listed in the first two lines of the table are the same criteria used in the updated manuscript.

Comment #5: First, note that the bias for ozone suggested here as a goal is 5%, not the 60% they are using for fractional error. While it is difficult to do a direct comparison of fractional error to normalized mean error, it would appear that almost all of their results are performing worse than suggested. The NME suggested of 15% is, again, much tighter than the 75% they use for fractional error. Fundamentally, the ozone, performance, at least, does not appear to be very good vis a vis past studies. Considering these updated metrics, the performance looks to be substandard (though it is difficult to directly compare fractional metrics they use versus the normalized metrics used by Emery et al.). Just because it is a more recent paper, and that neither it nor the Boylan paper represent EPA policy, it is likely more appropriate to use the Emery article, and remove all references to being EPA guidelines. It is also very important that they specify precisely what data (e.g., hourly, no cut-off, specified days) is being used in the evaluation. I will note, the mean FB from Simon is only 0.03 for ozone, not the 0.6 they use. The typical results from this paper appear much higher. The Typical FE is only 0.22: not the 075 they show. A similar result is found when looking at other results from Simon: The mean results typically are much better than are used in the Venecek Table 2. If the authors do pursue publication, their results should be tabulated to provide statistics tht can be compared directly to the Simon et al. and Emery et al., papers, and should also consider the evaluation approaches and results from the AQMEII initiative(4-11).

Response: We apologize for failing to correctly update the gas-phase model performance criteria in the last revision. Figure 3 and SI tables have been updated to reflect the criteria noted above (by Emery et al (2017)) for daily, 1-hr maximum ozone. Out of ~350 monitors, just over 95% met the performance criteria *mentioned above of <25% NME for max 1-hr O_3 values. In addition, all model predictions met the typical FE value of 0.22 also noted above. The authors believe the modeled O_3 values capture the peak photochemical episodes across the majority (95%) of monitor locations throughout the study domain. If the editor would like the author's to present more model performance statics we are happy to address these concerns in further revisions.*

Comment #6: There is also an issue in terms of what they show and how it is described. For example, the figure caption for Fig. 5 is "Figure 5 Predicted vs. Measured (a) Organic Carbon and (b) Elemental Carbon (µg m-3)" Are those the episode averages of EC and OC at each city? The daily values in each city? This is much, much different than showing the 24-hour values, for which performance is usually measured. In the manuscript and SI, they also need to be much more explicit as to what periods are being used for model performance.

Response: Figure 5 shows the predicted 24-hr average EC or OC at each monitoring site location on each day that had available measurements. This is stated on the figure and included in the supporting information – however the Reviewer may have received an outdated SI (based on reviewer #1 comments).

Comment #7: I note that the authors now stress that the results are only for a very short period during the summer. This makes their results of much less interest. How characteristic are UFP levels during a week in the summer vs. all year?

Response: All versions of the paper have clearly stated that the focus was on a peak photochemical episode. PM concentrations are generally studied over short time periods of 24hrs and longer time periods such as an annual average. The current study is focused on short time periods during peak photochemical events.

Comment #8: The OM to OC ratios used look low. A long-chain (say 15 C) alkane would have an OM to OC ratio of 1.17, so if you add even a little oxygen, it would be even higher. I seem to hear that secondary OM is more on the order of 2xOC. They should cite a source for the ratios used.

Response: The author's agree with the reviewer than the OM to OC ratio should be slightly higher and have recalculated OC to have a ratio of 1.2 (Russell 2003). Figure 5 on page 18 has been updated in the main manuscript and citation added to the references.

Comment #8: In summary, I would have to recommend against publication in its current form as detailed above. The lack of conformity with journal guidelines on data (and further, I would argue, not providing the data and underlying code would look bad for the journal unless a very good reason, i.e., a specific restriction on the data/code, in which case another code should likely be used for scientific studies), the misuse and characterization of model performance metric comparison, and using a really short period without providing a great reason for why such short periods are of scientific interest, are all important. In particular, I don't think ACP should be publishing a paper with major mistakes pointed out and that does not appear to comport to their guidelines on data and data and code availability of other leading journals.

Response: We respect the Reviewers opinion, but respectfully disagree with their conclusion.

References

1. Boylan JW & Russell AG (2006) PM and Light Extinction Model Performance Metrics, Goals, and Criteria for Three-Dimensional Air Quality Models. Atmos. Environ. 40(26):4946-4959.

2. Simon H, Baker KR, & Phillips S (2012) Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012. Atmos. Environ. 61:124-139.

3. Emery C, et al. (2017) Recommendations on statistics and benchmarks to assess photochemical model performance. J. Air Waste Manage. Assoc. 67(5):582-598.

4. Galmarini S, Rao ST, & Steyn DG (2012) AQMEII: An International Initiative for the Evaluation of Regional-Scale Air Quality Models - Phase 1 Preface. Atmos. Environ. 53:1-3.

5. Hogrefe C, et al. (2015) Annual application and evaluation of the online coupled WRF-CMAQ system over North America under AQMEII phase 2. Atmos. Environ. 115:683-694.

6. Koo B, et al. (2015) Chemical transport model consistency in simulating regulatory outcomes and the relationship to model performance. Atmos. Environ. 116:159-171.

7. Nopmongcol U, et al. (2012) Modeling Europe with CAMx for the Air Quality Model Evaluation International Initiative (AQMEII). Atmos. Environ. 53:177-185.

8. Rao ST, Galmarini S, & Puckett K (2011) Air Quality Model Evaluation International Initiative (AQMEII) Advancing the State of the Science in Regional Photochemical Modeling and Its Applications. Bulletin of the American Meteorological Society 92(1):23-30.

9. Rao ST, et al. (2014) Air Quality Model Evaluation International Initiative (AQMEII): A Two-Continent Effort for the Evaluation of Regional Air Quality Models. Air Pollution Modeling and Its Application Xxii, NATO Science for Peace and Security Series C-Environmental Security, eds Steyn DG, Builtjes PJH, & Timmermans RMA), pp 455-462.

10. Solazzo E, Galmarini S, Bianconi R, & Rao ST (2014) Model Evaluation for Surface Concentration of Particulate Matter in Europe and North America in the Context of AQMEII. Air Pollution Modeling and Its Application Xxii, NATO Science for Peace and Security Series C-Environmental Security, eds Steyn DG, Builtjes PJH, & Timmermans RMA), pp 375-379.

11. Wang K, et al. (2015) A multi-model assessment for the 2006 and 2010 simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2 over North America: Part II. Evaluation of column variable predictions using satellite data. Atmos. Environ. 115:587-603.