

Atmos. Chem. Phys. Discuss., referee comment RC1
<https://doi.org/10.5194/acp-2020-1291-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2020-1291

Anonymous Referee #1

Referee comment on "Improving predictability of high-ozone episodes through dynamic boundary conditions, emission refresh and chemical data assimilation during the Long Island Sound Tropospheric Ozone Study (LISTOS) field campaign" by Siqi Ma et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2020-1291-RC1>, 2021

This manuscript describes results from model calculations over the Long Island Sound (LIS) region and assesses improvements in predicted ozone distributions in the region in response to changes in model lateral boundary condition specification, model initialization strategies and representation of changes in emissions based on trajectories inferred from recent surface and space-based observations. Understanding the dynamical and chemical processes that lead to elevated O₃ in the LIS region is important as several monitors in the region, especially along the shoreline often measure O₃ values above the current standard. Consequently studies that help improve understanding of the O₃ formation and transport along this land-water interface and demonstrating improvements in modeling of these processes to improve accuracy of air quality forecasts and guide policy development and implementation for the region are of immense interest as also evidenced by the LISTOS field study, limited data from which the current analysis uses. While I am supportive of the theme of this manuscript, I feel there are aspects related to the presentation of the study motivation, the modeling experiment design, and interpretation of results that could be conveyed more effectively. The following suggestions are offered, addressing which may help improve the usefulness of the manuscript.

- At many places in the manuscript it is stated that a high-resolution forecast system has been developed for the LIS region. Is this an operational forecast system? Was the system operational for the entire LISTOS study? What relation does it have to the NAQFC system which is also often mentioned? Were the the base and additional simulations conducted in forecast or retrospective mode? Did the WRF simulation employ data assimilation – if so, was the assimilation strategy as one would use it in forecast mode? What was the extent of the WRF modeling domain? Was WRF deployed in a nested mode – is so what was the extent of the outer domain and how did it compare with the meteorology used for the NAQFC which provided the chemical LBC for the high-resolution domain?

- If the high-resolution forecast system has been deployed for an extended period, how does its general performance compare to that for the period of focus of this study, i.e., August 25-31, 2018? This aspect is specifically important to understand if there were conditions unique to the high ozone episode examined here or model attributes that led to the noted performance characteristics. On the other hand, if this is a limited modeling study that was conducted to examine this specific ozone episode it is okay to state that so the results can be viewed and assessed in the right context.
- It is not terribly surprising that for the limited geographic extent of the modeling domain considered here, the specification of chemical lateral boundary conditions is found to be influential on predicted ozone (and not NO_x but I suspect also for many other species such as CO, PM_{2.5} etc.) distributions. Given lifetimes of these species and the typical advective time-scales this should be an expected outcome. It is not readily apparent what constitutes the default static LBC profile – what were the typical values for O₃ and other species examined? Looking at the CMAQ documentation, it appears that the default LBC profile provided represents “clean” tropospheric conditions and is recommended for use along model boundaries that are typically over remote regions devoid of significant emission forcing. From the model documentation it also appears that even those are now often substituted with conditions derived from hemispheric versions of the model. It is thus a bit surprising that for a “new” high-resolution forecast system over a high emission density region such as the northeast corridor one would consider such a profile and not a nested configuration to capture the space and time varying chemical conditions of air masses advected to the LIS. In light of this, the statement on L265 “This suggests the default profiles provided by CMAQ represent a clean environment, such as marine air layer, and are not suitable for areas with active emissions and tropospheric O₃ production” is somewhat trivial. Clearly, such impacts are well recognized as the authors do attempt to account for such by using data from the NAQFC. A clearer description of the model set up and reasons for not using a consistent one-way nested configuration (with consistent treatment of meteorology, emissions, initialization for the outer domain) to better capture the LBCs would be useful.
- Perhaps one positive consequence of using the default LBC and comparisons with the run with the NAQFC LBC is demonstrating the expected influence of regional transport on O₃ levels in the LIS region, with a suggested enhancement of 10-20 ppb in peak hourly ozone on different days due to influences from outside the modeling domain, assuming that the discrepancies relative to the observations can be solely attributed to LBC and not other model processes or input? How representative is the limited set of conditions modeled here of the high ozone days in the LIS region? There were several days during summer 2018, outside the period examined here, when the LIS region witnessed high ozone levels.
- One interesting aspect depicted in Figure 2 is that the peak O₃ simulated in the control run is nearly the same on all days, suggesting that at least for the limited number of days examined in this study, emissions within the domain have about the same daily impact on average ozone. Assuming that the emissions within the domain are captured correctly and the WRF simulations represent the prevalent meteorology (neither of which are necessarily assessed), comparisons with the NAQFC and dynamic LBC case are suggestive of relatively large regional contributions on these days – how representative are these regional contributions of high O₃ in the LIS region? This also links back to my earlier comment on the skill of the new forecast system over an extended time-period.
- The description of the optimal interpolation application would benefit from additional clarity. If I understand the methodology, surface observations are used to adjust the model initial state for O₃ and perhaps NO_x (please state that explicitly if that is the case). However, it is not readily apparent if this is done every 24 hours? At what frequency is the OI applied? Also based on the strong forcing at the surface from emissions and deposition, it is to be expected that the influence of the OI fade away rapidly. It appears that such methods may be more useful to aloft data (e.g.,

<https://doi.org/10.1021/acs.est.8b02496>). Did the authors consider using aloft observations from the O₃ lidar in conjunction with the OI to explore possible improvements in short-term air quality forecasts? Improved initialization of the aloft conditions may also help better represent regional transport and modulate the inferred impact of LBC specification.

- The combined impacts of the LBC and emission adjustment simulations raise an interesting point on the representativeness of the NAQFC derived LBCs for the “forecast” year. Based on the arguments put forth, it appears that the “new” emission adjustments were applied only to the 3km resolution domain extent. By inference, the emissions utilized in the NAQFC are likely biased high since they did not benefit from these adjustments in the inferred emission trajectory since the NEI year. Conceivably, reducing regional NO_x emissions outside of the study domain will reduce the regional transported O₃ to the study region thereby possibly increasing the already low bias for the high O₃. What emissions were used in the NAQFC runs – did they benefit from similar “refresh” (reductions) relative to the NEI as the those within the LIS domain? Please clarify these aspects and any possible effects associated with a high bias in regional O₃ from NAQFC arising from possible high bias in emissions used in the NAQFC.
- Even though the study promotes high resolution modeling, much of the analysis focuses on aggregate metrics (averaged over model-station pairs). Thus, the relative advantages of using the 3km resolution are not readily apparent. Some discussion of the gains realized from higher resolution (say even relative to the 12km NAQFC) would be useful.
- L76-77: “complex urban areas” is a vague – please elaborate on the specific challenges for air quality forecasts.
- L82: perhaps should say six different sources or six different representations of lateral boundary conditions.
- L93: Is “modeling techniques” the correct terminology for the experiments that largely investigate different forms of input data to the CTM? I wouldn’t necessarily characterize changes in input data (initial conditions, boundary conditions, and emissions) as modeling techniques, especially since the methodology (at least the emission adjustment and OI) appear to be based on previously investigated methods.
- L 111: It’s not clear to me what does domain size has to do with fine-scale processes - I would have thought grid resolution would be more influential in that regard – perhaps this sentence would benefit from some restructuring.
- L181: it was not apparent to me what the 11x11 grid cell block refers to and what its relevance is? If it’s an area of influence in the OI, please provide some rationale for its choice?
- Was the emission “refresh” applied only to NO_x emissions or were emissions of other species also modulated relative to the base NEI. One would imagine that VOC emissions would have also changed between 2011-2018. Please also clarify why the 2011 NEI (L200) is used when conceivably updated (and closer to the forecast year) versions of the NEI (2014, 2017) may have been available?
- L150: Were model estimates of isoprene concentrations compared with observations from the LISTOS study? What may be the possible role of uncertainties in isoprene emissions within the LIS region on model predicted O₃ and its discrepancies relative to observations?
- Figs 2d and 3d - please state the time zone on the x-axis label - looks like UTC?
- L310-312: Please restate here and indicate in Table 2 caption the length over which these metrics are computed - are these for hourly paired model and observations.
- L323-324: This is somewhat of a misleading statement - the spike is just an indication of the model error at a specific time and location - not necessarily the impact of OI in large metropolitan areas. Please reword this sentence.
- L325-327: I find the suggestion that the magnitude of the OI adjustment is related to the emission strength/density to be speculative and not substantiated by any presented analysis. Restructuring the discussion would be useful.

- L327: The OI effect in large cities – this is to be expected since emissions at surface are the dominant forcing and thus one would expect the OI signal to get swamped out more rapidly in locations with higher emissions.
- L330: please clarify if these durations of the difference imply a corresponding improvement also in model skill?
- L347-348: Consider qualifying this statement for the specific use in a forecast system. The NEI's are updated for this very reason - to capture changes in emissions. Perhaps they are not available in time for the forecast application and thus the need to project the NEI's to the forecast year. As written the sentence is somewhat open to misinterpretation – note that a specific year is attached to each NEI to indicate the period of its representativeness.
- L350: Please elaborate why the uniform and spatially varying emission adjustments result in similar predictions? Do the likely differences get averaged out in the aggregate comparisons?
- L355-357: Please explain what "O3 production is NOx saturated" implies - is ozone essentially titrated at the monitors examined? If so, shouldn't there be a low bias in the case where emissions were not adjusted? What LBCs did the emission adjustment runs use?
- L374-375: Please state what time average values these metrics are computed for and over what time-period.
- L382: Please clarify what are the 3 adjusted runs? L370-374 indicate two adjusted runs? what does the reader associate runs #2,3,4 with?
- L384-385: It was not clear to me how one could infer concentration levels and over/underestimation from the Taylor plots in Fig 8 – please elaborate.
- L390: Please reword this sentence - ICs impact the initial state, BCs do not.
- Fig 10 and associated discussion: Presumably, the NAQFC which used older emissions overestimated the NOx emissions relative to the BOE - why then does it consistently predict lower NO2 relative to the BOE? Is it due to resolution or representativeness of emissions to the "forecast" year?
- L459-460: Are the meteorological drivers also not different between NAQFC and BOE simulations? Could this also not influence the comparisons of the chemical constituents?
- L466: If the peaks are underestimated, then should it not be expected that the false alarms will also not increase? Not sure what aspect of the BOE the authors are attempting to highlight here? Also, the sample size (number of days, and sites) does not appear large enough to make a conclusive statement on FAR.
- L470: Could the difference in the timing of the peak also result from differences in the meteorological fields?