Responses to Reviewer #1:
We appreciate the reviewer for taking time to review the revised manuscript and provide new comments. Please see below our point-to-point responses to these comments (*in Italic*).

1) L202-204: Please indicate at what model levels (altitudes) the ICs for O_3, NO, NO_2 were adjusted by the OI method – I believe it is only at the surface, but for completeness it should be explicitly stated.
Response: The ICs were adjusted for all layers within the PBL. We have explicitly stated it in Line 211:

"Next, as there is no information of vertical background profile in this method, the ratio between x^a and x^b at each surface layer grid point was used to scale the concentrations **for all layers within the PBL**, following Tang et al. (2015; 2017)."

2) The discussion related to the default BCs as a "control run", its representativeness for the conditions being simulated, and inferences drawn relative to the case with "dynamic" BCs still raise some questions. While it is okay to convey it as a control run, I feel some of the inferences drawn and the suggested importance of the use of the default BCs need to more carefully conveyed. If the intent of the "control" run is to infer the pollution burden solely associated with emissions within the LIS domain, wouldn't a case with no inflow (i.e., zero boundary conditions) be more appropriate? The static default LBCs appear to be so unrepresentative for the small LIS domain with surrounding regions of relatively high emissions, that anything with even slightly better representativeness would yield improved performance for the aggregate statistics by reducing the systematic biases induced by the default LBCs. I am thus a bit wary of some of the broad conclusions that are drawn on the suitability of one over the other for different seasons.
Response: Thank you. Following the suggestion, we have tuned down the BCs discussion and emphasized on the limitation in the conclusion drawn from this case study. See detailed changes below:
L326:
"For instance, the O_3 concentration over the Long Island Sound is lower than its surrounding areas during this episode."

L329:
"This indicates the high-resolution simulation can better reproduce the pollutant variability over this coastal urban area during this study period."

3) For instance, the discussion starting on L304 states that the "influence of BCs is less important during the cold season" – I would exercise some caution in making such broad sweeping statements based on the limited domain extent and conditions examined here. In fact, there are many studies documenting the importance of BC specification during times when "local" pollution production is slow (as in cool seasons). At the very least, the statement should be qualified to convey the specific and limited conditions it is based on. By the same token, I do not believe the importance of specifying space and time varying BC, especially for small geographic domains as the LIS region should be diminished just because deriving them from larger scale simulations is considered computationally burdensome for real-time applications (as suggested in the author response).

Response: Thank you. We made this statement (influence of BCs is less important during the cold season) based on the observation that there is a smaller difference between the upwind concentrations and the background concentrations used in the default BCs, compared to that during a hot season when the upwind photochemical production is more active (Fig. S2a, d). We agree that there could be cases in which long-range transport becomes even more important during winter. Therefore, we have modified the statement and discussed the aforementioned limitation (L308):

"In comparison, the influence of BCs is less important during the cold season. **There is a smaller difference between upwind concentrations and the background concentrations used in the default BCs, compared to that during a hot season when the upwind photochemical production is active,** resulting in better agreement between the prediction and observations (Fig. S2a, d). **Note that other studies have shown the influence of BCs becomes more prominent during the cold season when "local" pollution production is slow (e.g. Fiore et al., 2009). The magnitude of the actual influence is determined by several factors, such as the emission density, photochemical production and sink, and spatial range and gradients of the concerned species.**"

We agree that the importance of deriving the dynamic BCs should not be diminished, which is the motivation to test the NAQFC BCs, which are provided by NOAA's operational prediction and readily available for regional modelers to use.

Reference:
Fiore, A. M., Dentener, F. J., Wild, O., Cuvelier, C., Schultz, M. G., Hess, P., Textor, C., Schulz, M., Doherty, R. M. and Horowitz, L. W.: Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, J. Geophys. Res. Atmos., 114(D4), 2009.

4) I don't believe the conclusion on L311 that influence of dynamic BCs is "more significant during the high pollution time" is robust. Dynamic BCs could also be influential in conditions when local production may not be high. The dynamic BCs help better represent the changes in baseline pollution at a location based on changes in large scale forcing and thus should be important at all times. As stated, I think the statement can be easily misconstrued and should either be qualified for specific conditions, or further substantiated, or deleted. It appears that for the cases examined in this study the baseline levels in the LIS were enhanced by regional transport.
Response: We removed this statement.

5) L323: I think this sentence is somewhat trivial in that it should be expected that a finer resolution simulation will show "more detailed distributions" than a coarser resolution simulation – perhaps it would useful to point out specific features of interest that are better resolved and point the reader to the appropriate illustration supporting that. How does one assess if the more detailed distribution is better?
Response: We added in the revised manuscript the specific features of the spatial differences between the two simulations (L326):

"For instance, the O_3 concentration over the Long Island Sound is lower than its surrounding areas during this episode, which is better resolved by the 3 km simulation than the 12 km NAQFC (Fig.

2a-c). The O_3 distribution along the coastal area, such as the coasts of Connecticut and Rhode Island, also agrees better with the observations than the 12 km NAQFC prediction."

6) L324: please point to the observational evidence that suggests that "O_3 over the LIS is lower than its surroundings". Is this a persistent feature or only on specific days?
Response: We have specified that this is the case during the study period, as shown in Figure 2.

"the O_3 concentration over the Long Island Sound is lower than its surrounding areas **during this episode**"

7) L326: based on Figure 2 it is not readily apparent which specific feature of the O_3 distribution along the coast agrees better with the observations – it would helpful for the readers if that association is more explicitly pointed out.
Response: Thank you. We have added the following explanation in the revised manuscript (L328):

"The O_3 distribution along the coastal area, such as the coasts of Connecticut and Rhode Island, also agrees better with the observations than the 12 km NAQFC prediction."

8) L327: Given the many differences (in addition to model grid resolution) between these 3km resolution runs and the NAQFC configuration, can one emphatically claim that "This proves the high-resolution simulation can better reproduce the pollutant variability in coastal urban areas". While I do strongly believe that finer resolution simulations are needed for coastal urban areas, I do not think that the analysis presented necessarily proves that need in a robust manner – the statement appears to be a little too strong for the evidence presented.
Response: Thank you. Following this suggestion, we have tuned down this statement (L329):

"This indicates the high-resolution simulation can better reproduce the pollutant variability over this coastal urban area **during this study period**."

9) L462: "2014 NEIs" should be "2014 NEI"
Response: Revised to "2014 NEI". Thank you.

10) L581: How does one ascertain that the grid resolution solely resulted in the NO_2 spatial pattern differences between these runs and the NAQFC?
Response: The statement "the grid resolution solely resulted in the NO_2 spatial pattern differences between these runs and the NAQFC" cannot be found in L581 or anywhere in the revised manuscript.

11) L637: it would be useful to point the readers to the illustration that supports the statement that 3km resolution improves simulation of peak values and timing.
Response: We have added the illustration (Figure 12) into this sentence (L644):

"As the emission and meteorological inputs play an important role in determining the magnitude and timing of high peaks (Pan et al., 2017), high resolution emissions and meteorological data

contributed to the improved simulation of peak **O_3** value and **its** timing, especially over urban areas **(Fig. 12).**"

12) In response to my query on whether emissions of species other than NO_x were "refreshed", the authors response states that "regional O_3 production is more controlled by NO_x than VOC" – while this may be true for regional O_3 distributions, does it also hold for the NY metro area and the LIS region. I do acknowledge, that changing emissions over the past several decades has changed NO_x and VOC sensitive regimes in the area, and would think that capturing both changes in NO_x and VOC emissions would be important. It is also important to note that several recent studies suggest perhaps higher VOC emissions from volatile consumer products than represented in the inventories. Thus, it seems that updates to VOC emissions may also influence O_3 simulations in some urban areas. It may thus be worthwhile to at least acknowledge the need for updates similar to NO_x for other precursor species, while also acknowledging the challenges in doing so.
Response: Thank you. We have acknowledged the need for VOCs updates simiilar to NOx (L710):
"In addition, the importance of volatile consumer product VOCs has been identified in recent studies (McDonald et al., 2018), suggesting that updating other species than NO_x is also necessary. This may be challenging through a similar approach, due to limited measurements of VOCs from both ground and space instruments."

13) L402: the sentence "In large metropolitan areas, OI adjustments result in spikes in large metropolitan areas indicate the model errors at the time..." is awkwardly phrased and needs to be rewritten.
Response: Thank you. We have rephrased this sentence (L404):

"In large metropolitan areas, OI adjustments result in spikes ~~in large metropolitan areas~~ that indicate ~~the~~ larger model errors at the time of OI adjustment, with the mean errors ~~being~~ up to 14 ppbv in surface hourly O_3 concentrations over NYC and 16 ppbv over Philadelphia, respectively."

14) The additional tables 4 and 5 only re-enforce that the two adjustment techniques did not alter performance much, not necessarily explain likely reasons. I am not sure the various possible reasons outlined in the response, coherently suggest the reason either. If O_3 production is low in "NO_x saturated" regions, why compare O_3 there rather than in downwind regions where the production is enhanced? Is the suggestion that in downwind regions the urban emissions in the plume are "well-mixed" such that the spatial adjustments in emissions in the urban core have little effect? If so, it may be worthwhile speculating so.
Response: Tables 4 and 5 provided here cover regions with both high (four urban areas, NYC, PH, NHH and PP) and low (OTHR) emission density. Therefore, these tables show the influence of NO_x emission adjustments over both urban and well-mixed downwind regions.

Responses to the Comments by Reviewer #2:
We want to sincerely thank the reviewer for the helpful comments in the second round. We provide below the point-to-point responses to these comments, and corresponding changes have been made in the revised manuscript.

1. Line 189: "BCs that can be used" (insert "can")
Response: The word has been inserted:
"BCs that **can** be used to drive"

2. Line 228: "provided by the model" (insert "the")
Response: Inserted.

3. Lines 401-402: delete the second "in large metropolitan areas" and insert "that" after spikes
Response: This sentence has been modified as follows (L403),

"In large metropolitan areas, OI adjustments result in spikes ~~in large metropolitan areas~~ **that** indicate ~~the~~ **larger** model errors at the time of OI adjustment ~~at the monitor sites~~, with the mean errors ~~being~~ up to 14 ppbv in surface hourly O_3 concentrations over NYC and 16 ppbv over Philadelphia, respectively. "

4. Line 449: Elaborate what you mean by "Similarly, satellite observations are weighted more toward urban plumes."
Response: Changed to "Similarly, the retrievals of satellite observations are also more sensitive to urban plumes."

5. Line 486-487: "On average, the RMSEs of BOE (EmisAdj_avg) is slightly smaller." This is not true for ozone (Table 4),
Response: You are right that it is not true for ozone in Table 4. However, Table 4 describes the performance for the designed EmisAdj studies that are different from the final configuration (with BCON, OI and Emission adjustment or so called BOE). The referred statement was intended for the BOE case in Table 4&5, not in Table 4. The results are slightly different once all adjustment methods are used together from that when a single method was applied. We have clarified it in the revised manuscript.

"On average, the RMSE of the combined BOE setup is slightly smaller using the EmisAdj_avg method than that using the EmisAdj_sub method during the study period (Table S4, S5)**, which is different from that when a single adjustment method was used (see Tables 4 and 5).**"

6. Line 598: Rewrite the following phrase: "making it underestimates the hotspot but much closer to the VCD over the rest of the areas."
Response: Thank you. We have rewritten this sentence as follows,

"In addition, the NO_2 VCD from **the** simulation with **the** combined adjustments using **the** EmisAdj_sub method for emission refresh shows a similar spatial pattern ~~with~~**to** that ~~of~~ using BOE (Fig. S3). The NO_2 VCD level, however, is ~~over the NYC area~~ lower over the NYC area, ~~making~~

suggesting an ~~it~~ underestimate over the hotspot but much ~~closer~~ **better prediction** over the rest of the areas."

7. Lines 602-603: This may be another indication (in addition to surface ozone results) that EmisAdj_sub is indeed the better approach. Any comments?
Response: Agreed. We added the following statement to reflect this view of point:

"It indicates the advantage of adjusting emissions with a finer spatial resolution in simulating NO_2 vertical column in this study."

8. Line 637: "inputs play an important role" (insert "an")
Response: Inserted. Thank you very much!