Response to reviewer comments for "Assessing Ozone Dynamics During the 2023 Summer STAQS Field Campaign Using Synergistic Observations and Model Simulations"

We thank the referees for their insightful comments as they will undoubtably help the strength of the manuscript. The replies are written in black below each comment.

Referee 1 Comments:

This manuscript presents an analysis of O_3 and precursor variability during the 2023 STAQS campaign over the NYC/LIS region, using a combination of TEMPO L2 v3 NO_2 and HCHO retrievals, PANDORA observations, GCAS measurements, in situ aircraft data, TOLNet and Doppler lidar observations, and high-resolution WRF-Chem simulations. The topic is timely and relevant to ACP, particularly given the early scientific use of TEMPO observations in a complex urban-coastal environment where both chemistry and mesoscale meteorology play important roles. The manuscript also benefits from the unusually rich observational context available during STAQS. Overall, I find the study potentially suitable for publication after careful revision. The manuscript has clear strengths, but several of the central interpretations are currently stronger than the supporting evidence warrants. In particular, some conclusions regarding TEMPO performance across pollution regimes, the mechanistic attribution of the selected O_3 episodes, and the use of HCHO/NO_2 ratios in diagnosing chemical sensitivity require a more cautious and better substantiated presentation.

Several major issues must be addressed before the paper can be accepted for publication.

First, some important information of the method is missing, and some discussions require clarifications.

1. The manuscript draws broader conclusions about TEMPO performance than are currently supported by the available validation dataset.

The intercomparisons with PANDORA and GCAS are useful, but the dataset remains limited in ways that should be more explicitly reflected in the scope of the conclusions. In particular, the GCAS comparisons appear to rely on only two flight days, and the manuscript itself notes limited spatiotemporal coincidence, small sample sizes, and the possibility of sampling-related smoothing of gradients for short-lived species such as NO_2.

Given these limitations, some of the broader statements regarding regime-dependent TEMPO performance appear too strong. In particular, the conclusion that TEMPO captures precursor variability more accurately under cleaner conditions but is less capable in moderate or more complex environments, may be reasonable as a preliminary inference, but it is not yet demonstrated in a sufficiently systematic way. This manuscript would benefit from a clearer distinction between conclusions that are directly supported by the collocated comparisons and those that remain case-specific interpretations. At least, the authors should: (1) report sample sizes more consistently for each intercomparison and subset, (2) clarify which conclusions are robustly supported by the available coincidences, and (3) moderate language where the current evidence remains limited.

Author Response:

This is a valid concern, and the manuscript has been revised to reflect the following changes. We have revised the manuscript to more carefully to distinguish between conclusions directly supported by the collocated comparisons and those that represent preliminary, case-specific inferences.

1. We have added sample sizes to all relevant figures, tables, and in-text descriptions to improve transparency across each intercomparison and subset.
2. The revised text more explicitly acknowledges the limited spatiotemporal coverage of the GCAS dataset and the Pandora network and reframes the regime-dependent performance conclusions. Specifically, the conclusion that TEMPO better captures precursor variability under cleaner conditions has been reframed as a preliminary inference requiring confirmation from more extensive observations.
3. We have updated language throughout the manuscript to better reflect the results as inferential. We replaced any overstated phrasing (e.g., "reveals," "demonstrates") with appropriately substitutes (e.g., "suggests," "consistent with"). We also added explicit caveats noting that HCHO comparisons are based on hourly TEMPO snapshots rather than the multi-day averages typical of satellite HCHO validation, which limits the statistical robustness of those results.

The updates appear in the Abstract, Section 3.1.2, Section 3.3.3, and the Conclusion. The following lines can be found updated in the revised manuscript: Lines 20 - 21; 30 - 31; 246; 249 - 250; 262; 265; 267; 293; 297; 304 - 310; 445 - 480; 537 - 557; 570 - 596; 608; 609; 626 - 629; 653 - 666; 677; 679; 683.

Rather than listing each revised line individually, a representative example of the revised language is provided below:

Line 600 - 603: Overall, TEMPO appears to better capture both magnitude and temporal variability on the low pollution day, however, with only two flight days available, this result represents a preliminary inference and a more systematic evaluation across additional cases is needed before drawing broader conclusions about regime-dependent TEMPO retrieval performance.

Following figures and tables are updated:

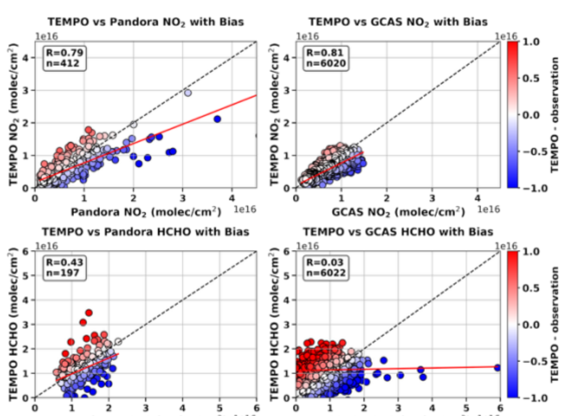

Figure 2: TEMPO comparison with Pandora (left column) and GCAS (right column) retrievals for NO_2 (top row) and HCHO (bottom row) (molec/cm2). Points are colored by the bias (TEMPO – Pandora or GCAS). The dashed black line indicates the 1:1 relationship and the solid red line shows the least-squares regression fit. Sample sizes are indicated in each panel (n).

Table 1. TEMPO vs. Pandora average campaign VCD statistics

	NO_2					HCHO				
Site	n	R	NMB (%)	RMSE (molec/cm^2)	Mean Pandora (molec/cm^2)	n	R	NMB (%)	RMSE (molec/cm^2)	Mean Pandora (molec/cm^2)
New Brunswick	44	0.80	-25.5	4.8e+15	5.6e+15	27	0.29	-1.6	6.2e+15	1.3e+16
Bayonne	52	0.65	-28.3	6.4e+15	1.4e+16	46	0.42	-8.4	5.1e+15	1.3e+16
Queens	49	0.75	-20.2	3.5e+15	1.0e+16	42	0.51	-13.3	5.5e+15	1.5e+16
Manhattan-CCNY	44	0.69	7.4	2.7e+15	7.4e+15	7	0.82	6.2	5.8e+15	1.3e+16
Bronx	49	0.72	11.1	2.4e+15	6.3e+15	--	--	--	--	--
Old Field	34	0.10	-7.6	1.2e+15	2.9e+15	--	--	--	--	--
Westport	51	0.25	62.0	1.7e+15	1.9e+15	34	0.24	-17.4	5.2e+15	1.3e+16
New Haven	46	0.31	-33.7	1.7e+15	3.8e+15	--	--	--	--	--
Madison	43	0.42	-1.2	9.9e+14	2.0e+15	41	0.14	-7.6	4.3e+15	1.0e+16

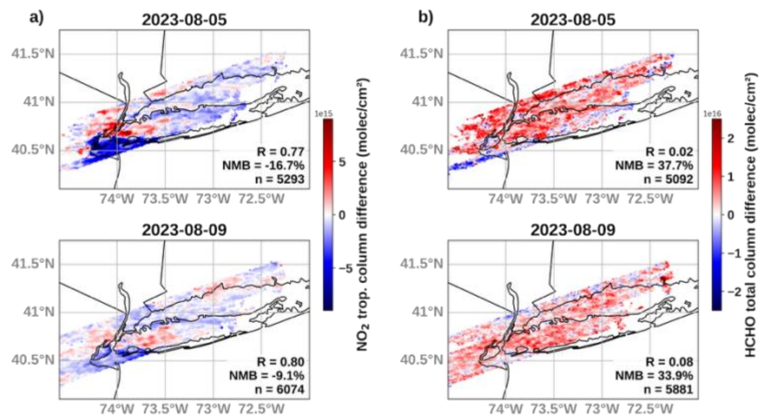

Figure 3: TEMPO - GCAS NO_2 (a) and HCHO (b) tropospheric VCD differences (molec/cm2) for two DC-8 flight days: August 5 and 9, 2023. The main statistics of these differences are displayed in the figure legend. Sample sizes are indicated in each panel (n).

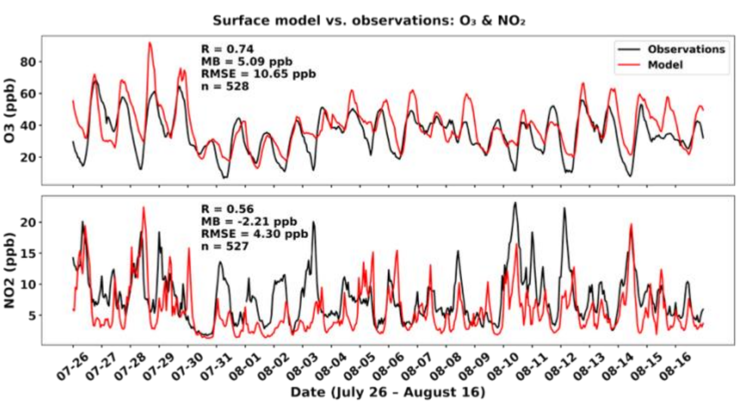

Figure S2: Hourly-averaged observed (black) and simulated (red) surface O_3 and NO_2 (ppb) timeseries during the STAQS campaign period (July 26 – August 16, 2023). The statistical comparison (correlation (R), mean bias (MB), and root mean squared error (RMSE) of the model and surface observations are provided in the inset of the figure. Sample sizes are indicated in each panel (n).

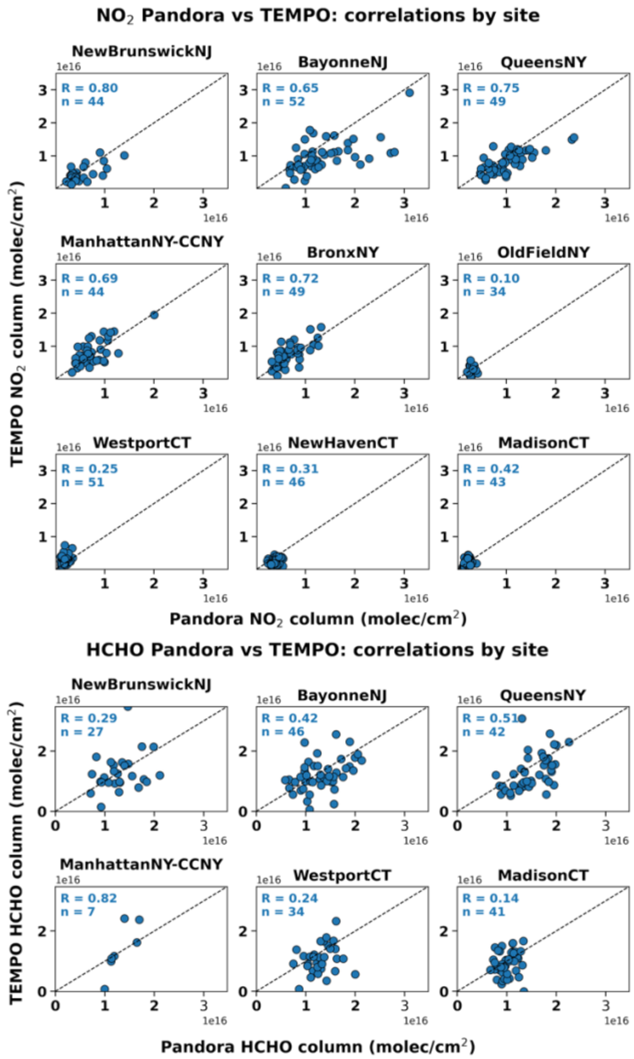

Figure S3: TEMPO vs. Pandora VCD comparison at each Pandora site in the NYC/LIS region for NO_2 (a) and HCHO (b) (molec/cm^2). The correlation of TEMPO and Pandora observations are provided in the insets of each panel. Sample sizes are indicated in each panel (n).

2. The mechanistic interpretation of the high-pollution cases, especially 28 July, remains more inferential than the manuscript currently acknowledges.
The interpretation of the 26 July case is comparatively well supported by the combination of precursor buildup, observed wind evolution, and the spatial O_3 pattern. The interpretation of the 28 July case is less direct. The manuscript attributes the event to overnight accumulation within the shallow marine boundary layer and/or residual layer, followed by daytime entrainment into the growing boundary layer. While this is a plausible interpretation, the manuscript itself indicates that the evidence is indirect and that additional observations from the previous evening would have strengthened this conclusion.
This distinction is important because the scientific narrative of the paper relies substantially on the contrast between these two high-ozone cases. At present, the manuscript sometimes presents the 28 July mechanism in a manner that reads as demonstrated rather than inferred. I suggest revising this section so that the evidentiary basis is more transparent. In particular, the discussion

should more clearly separate: direct observational support, model-assisted interpretation, and remaining uncertainty or plausible alternative explanations.

Author Response:
We agree and have revised the July 28, 2023, discussion to more transparently separate direct observational support, model interpretation, and remaining uncertainty. Specifically, we explicitly acknowledge that direct overnight observations are unavailable, and that the residual layer entrainment mechanism relies on WRF-Chem simulated O_3 curtains to bridge the observational gap between the evening of July 27 and the morning of July 28. To partially address this gap, we have added a new supplementary figure (Fig. S8) showing TOLNet observations from July 27 (up to 00 UTC), which provide direct observational evidence that an O_3 rich layer (75-105 ppb) was present aloft at TROPOZ and TOPAZ going into the overnight period. The July 28th TOLNet profiles beginning at 12 UTC remain as direct observational support for elevated O_3 just above the surface layer deepening through the afternoon, consistent with PBL entrainment. We have also added discussion of coastal convergence and recirculation within the LIS region as a plausible alternative or complementary mechanism (per Referee 2 suggestions), noting that surface wind observations show some evidence of convergence along the northern LIS shore but that the available dataset is insufficient to quantitatively distinguish between these pathways. The revised text explicitly frames the July 28 interpretation as a provisional hypothesis rather than a demonstrated mechanism. The updated text appears in Lines 408 - 428.

Lines 408 - 428: To assess whether these mechanisms may have contributed to the pollution events on July 28, we examined the simulated O_3 curtains as well as the observed wind lidar profiles spanning the overnight and early morning period (Fig. 6, S8). TOLNet observations on July 27 show elevated O_3 (75–105 ppb) persisting aloft at TROPOZ and TOPAZ through the late afternoon and into the evening (up to ~00 UTC July 28), providing observational evidence that an O_3 rich layer was present going into the overnight period (Fig. S8). Direct overnight observations are unavailable, however, WRF-Chem simulated O_3 curtains suggest elevated O_3 concentrations (60 – 75 ppb) may have persisted within the residual layer (approximately 1 – 2 km) through the overnight hours at these same sites. Consistent with this, the TOLNet observed profiles beginning near 12–13 UTC on July 28, 2023, capture elevated O_3 levels just above the surface layer that deepen vertically by mid-afternoon as the PBL develops (Fig. 6c), which is consistent with entrainment of residual layer air into the growing daytime PBL, though direct observational confirmation of the overnight transport itself is limited. There is also an overnight shallow marine boundary layer most notably present at the Yale Coastal site and slightly at the Stony Brook site (Fig. 6d), however, direct observational evidence of elevated O_3 within this layer is not available, and its contribution remains inferred from the surface timing and spatial pattern of the coastal enhancements. An alternative or complementary mechanism involves coastal convergence and recirculation within the LIS region, where interactions between the south-shore sea breeze and opposing coastal flows from Connecticut could favor O_3 accumulation; surface wind observations show some evidence of convergence along the northern LIS shore, though the signal is not strong enough to conclusively distinguish this from the residual layer entrainment pathway. These results are consistent with the early coastal O_3 enhancements on July 28, 2023, having been influenced by overnight pollutant accumulation in the residual layer and possibly shallow marine boundary layer, subsequently entrained into the daytime PBL during its initial growth that favored sustained coastal O_3 enhancements. However, given the reliance on model-simulated overnight structure and the plausibility of coastal convergence as a contributing factor, this interpretation should be regarded as a provisional hypothesis rather than a demonstrated mechanism.

3. The manuscript relies heavily on WRF-Chem for process attribution, but the model evaluation provided is relatively limited for that purpose.
The authors state that Sect. 3.1.1 is not intended as a detailed model evaluation, but rather as a demonstration that the model can help fill observational gaps. That is a reasonable objective. However, later sections use WRF-Chem extensively to support process-level interpretations, including the vertical and horizontal continuity of pollution layers, precursor transport pathways, and residual-layer influences.
This creates some tension between the stated role of the model and the interpretive weight placed on it. The model skill shown in the manuscript appears acceptable in a broad sense, but not sufficiently comprehensive to support all of the more detailed mechanistic conclusions without additional caution. Surface NO_2performance is only poor to moderate in several subregions, and some O_3 performance is weaker in coastal areas than in the urban core. The manuscript would therefore benefit from one of two approaches: (1) provide a more thorough case-oriented model evaluation, particularly for the vertical structure that underpins the key interpretations, or (2) moderate the strength of conclusions that depend primarily on model results. At present, the manuscript does not always clearly distinguish between model use for interpolation/gap-filling and model use for mechanistic attribution.

Author Response:
We have updated the manuscript following the Referee 1's suggestion (2): moderate the strength of conclusions that depend primarily on model results. Since the model evaluation focuses primarily on surface comparisons, we have reframed cases where we used WRF-Chem to support any detailed mechanistic or vertical structure interpretations, replacing language to reflect the model's supporting rather than a main role. The most significant revisions are addressed with the July 28th case, where the model is now explicitly identified as the only basis for the overnight period, and the entrainment interpretation is framed as a hypothesis consistent with observations but not demonstrated by any available observations (as detailed in the response to comment 4 above). No additional model heavy language requiring revision was identified elsewhere in the manuscript. The updated text appears in Sections 3.3.1.

4. The discussion of $HCHO/NO_2$ (FNRs) ratios and O_3 sensitivity requires substantially more caution.
The manuscript uses HCHO, NO_2, and particularly $HCHO/NO_2$ ratios to infer aspects of the photochemical environment and to discuss transitions toward more VOC-rich or NOx-limited conditions during selected episodes. However, the manuscript also shows that HCHO is the least robust component of the observational analysis. TEMPO HCHO exhibits only modest agreement with Pandora and essentially no correlation with GCAS, with substantial uncertainty and noise. Given this, ratio-based interpretation needs to be presented much more carefully. In its current form, the manuscript risks assigning more diagnostic weight to the $HCHO/NO_2$ ratio than the HCHO evaluation supports. If the authors wish to retain this discussion, they should explain why ratio-based inference remains informative despite the weak HCHO performance and avoid

wording that suggests a robust diagnosis of ozone sensitivity where the supporting evidence is uncertain. This point is central because the chemical interpretation of several cases depends on these ratio-based arguments.

Author Response:
We agree and have revised the manuscript to more carefully frame all FNR-based and HCHO-informed regime inferences. Given the limited HCHO retrieval performance demonstrated in the validation section, the FNR analyses are treated as qualitative rather than to diagnose definitive sensitivity regimes. Any specific language implying regime characterization from FNRs has been updated throughout the results and conclusion, and a consistent caveat noting the known uncertainties in HCHO retrievals has been added where FNRs are discussed. The individual NO_2 and HCHO column patterns, which are better constrained observationally, remain the primary basis for precursor characterization.

The updated lines are provided below and appears in Section 3.2.1, 3.3.2, 3.3.3, and the Conclusion. Rather than listing each revised line individually, representative examples of the revised language are provided below:

<u>Lines 451 - 453</u>: At the same time, GCAS captures some HCHO enhancements over the LIS and along both coastlines, suggesting possible enhanced VOC contributions to the O_3 photochemical production downwind/over water in a potentially more NO_x-limited region (VOC-rich) (Fig. 7c, Fig. S9a).

<u>Lines 476 - 480</u>: The GCAS column measurements provide additional context for distinguishing two different types of coastal O_3 cases where a NO_x-dominated urban plume and sea breeze recirculation appear to have driven the July 26, 2023, event versus a more photochemically processed pollution regime on July 28, 2023, potentially influenced by residual pollution from prior days. FNR-based O_3 sensitivity inferences are qualitative given known uncertainties in HCHO retrievals; conclusions reflect broad photochemical regime characterization rather than a quantitative sensitivity diagnosis.

5. The comparison between TEMPO NO_2 columns and ACES in situ NO_2 should be framed more explicitly as qualitative consistency, not validation.
The manuscript compares TEMPO tropospheric NO_2 VCDs with ACES in situ NO_2 mixing ratios and notes that the quantities differ in units and vertical representativeness. Such a comparison can still be useful as a qualitative illustration of spatial consistency, but it should not be interpreted too strongly. At present, the corresponding discussion appears to provide additional support for TEMPO performance in a manner that risks overstating what can be concluded from a column-versus-in-situ comparison. I suggest revising the text to make clear that this figure demonstrates qualitative coherence in spatial patterns, rather than constituting a formal validation.

Author Response:
We agree and have revised the text to explicitly frame the ACES comparison as a qualitative illustration of spatial pattern links rather than formal validation. Language implying TEMPO

performance assessment from this column vs. in-situ comparison has been removed. The updated text appears in Section 3.1.1 (Lines 317 - 327).

Lines 317 - 327: TEMPO tropospheric NO_2 VCDs were also compared with in situ NO_2 (ppb) measured by the ACES instrument aboard the DC-8 aircraft on August 9 and 16, 2023 (Fig. 4). While these products differ in units and vertical representation, the comparison serves as a qualitative illustration of spatial patterns and the connection between surface and column observations rather than a formal validation. On both days, ACES-measured NO_2enhancements generally align with the elevated TEMPO tropospheric VCDs, particularly over the NYC urban core, suggesting well-mixed plumes and spatial consistency in the urban-suburban gradients in NO_2. Some mismatches occur such as along the LI coastline on August 16, 2023, where ACES captures sharp local gradients that are not as elevated in the TEMPO VCDs, likely reflecting the difference in vertical representation between the shallow in situ layer and the total column measurement. Overall, this qualitative agreement across multiple independent measurement types provides additional spatial context for interpreting TEMPO column retrievals. Combined with high-resolution WRF-Chem output, these observations allow us to bridge gaps among surface monitors, GCAS, and Pandora, providing a comprehensive view of O_3 dynamics and precursor species characteristics during the STAQS campaign.

Second, the manuscript lacks scientific explanations in some places.

1. The manuscript would benefit from a clearer separation of its two main objectives: TEMPO evaluation and O_3 process diagnosis. At present, the narrative sometimes shifts between these aims in a way that weakens the overall structure.

   We appreciate this observation. The two objectives, TEMPO retrieval evaluation and O_3 dynamics, are introduced sequentially in the introduction and addressed in dedicated sections (Section 3.1 and Sections 3.2, 3.3, respectively). We have reviewed the transitions between these aims and made minor clarifications where the narrative connection was unclear, but we believe the current section structure adequately separates the two objectives without requiring broader reorganization.

2. The classification into high-, moderate-, and low-pollution days is useful, but the criteria should be presented more explicitly in the main text, ideally in a short summary table.

   The classification criteria, including meteorological basis, assigned days, and O_3 thresholds for each regime, are stated explicitly in Section 3.2. We have added a brief clarifying sentence noting the classification metric to ensure this is unambiguous. We therefore believe a summary table would be redundant given the definitions already provided.

   Lines 330 - 332: Classification is based primarily on observed peak hourly surface O_3 concentrations and the prevailing synoptic conditions, with thresholds informed by the distribution of O_3 levels across the campaign period rather than fixed regulatory standards.

3. The terminology around "moderate pollution" and "localized exceedances" should be used carefully and consistently. The manuscript notes that some events involved brief hourly concentrations above 70 ppb rather than exceedances of the MDA8 standard; this distinction should remain explicit throughout. The statement already in the manuscript supports this distinction in Lines 489 - 490: The conditions for the moderate pollution days resulted in unique, localized O_3 exceedances, defined here as brief hourly concentrations exceeding 70 ppb (not considered regulatory exceedances events based on the maximum 8-h O_3 average). But more changes have been made throughout Section 3.3.2., to better clarify this:

- Lines 509 - 511: Strong winds on this day transported polluted air far inland and north of the city (Fig. S7f), a pattern typical of downwind transport from major urban centers and often associated with localized hourly O_3 enhancements.
- Lines 526 - 529: Together, the co-located lidar observations on August 11, 2023, highlight the sharp mesoscale differences that can occur across the LIS where NY north shore experienced lower pollution and the CT coastline sustained hourly O_3 concentrations exceeding 70 ppb for multiple hours, though these do not constitute regulatory exceedances based on the MDA8 standard.
- Lines 531 - 533: Near the urban region (Bronx), winds shifted from early morning easterlies to southerlies by evening, throughout ~1 km (Fig. 9d), while at Suffern, near-surface winds were stagnant, coinciding with the evening localized hourly O_3 enhancement (78 ppb) at the correspondent AQS site north of NYC.
- Lines 581 - 583: For the moderate pollution case, there were two minor O_3 events: first the localized hourly enhancement on Staten Island and second the moderate increase along the CT coast. The NO_2 columns indicate high morning concentrations over Staten Island, Manhattan, and the western end of LI that are consistent with the emission buildup and upwind source of the O_3 enhancement later in the day.
- Lines 681 - 684: This analysis suggests that similar mesoscale transport mechanisms can result in distinct surface O_3 cases depending on recirculation, the vertical distribution of pollutants, and chemically favorable conditions, underscoring the critical role of a multi-dimensional perspective in regulating localized hourly O_3 enhancements.

4. Sample sizes (n) should be reported more systematically for all intercomparisons and subsets, including the overall TEMPO-Pandora and TEMPO-GCAS comparisons, the site-by-site Pandora analysis, and any regime-specific subsets. This is particularly important for evaluating the robustness of the reported correlations. Changes completed for the following Figures and Tables: Figure 2, Figure 3, Figure S2, Figure S3, Table 1, Table S2.
5. Line 123: Please spell out TUV at its first occurrence. Change completed.
6. Lines 195 and 441: Please format the wind speed unit correctly as m s^{-1}. Changes completed throughout manuscript.
7. Section 3.2: Please specify more clearly which O_3 metric is being used. For example, at Line 326, it is not clear what is meant by *moderate overall O_3"* and whether the quoted value of 60 ppb refers to a daily mean, daytime mean, MDA8, or another metric. Based on hourly observed and simulated O_3. Changes completed.
8. Line 438: This sentence appears to be incomplete and should be checked. Change completed.

Referee 2 Comments:

Overall, this is a well-written and scientifically valuable manuscript that combines multiple observational platforms, including TEMPO, GCAS, Pandora, TOLNet lidar observations, surface measurements, and WRF-Chem simulations, to investigate Long Island coastal ozone evolution and transport processes during the 2023 STAQS deployment periods. The multi-platform observational framework and integration of satellite, airborne, lidar, and modeling datasets provide a comprehensive perspective on ozone variability and coastal transport dynamics. The manuscript is generally well organized, and the observational analyses are detailed and informative. Before supporting publication in ACP, I have several comments as listed below:

(1) The manuscript relies heavily on simulated vertical ozone structure to support the residual-layer entrainment interpretation on July 28; however, model evaluation is primarily limited to surface ozone comparisons. Since the proposed mechanism depends strongly on elevated residual-layer ozone reservoirs and subsequent downward mixing, additional evaluation of the modeled vertical ozone distribution against the TOLNet ozone lidar observations would strengthen confidence in the interpretation.
(2) Because the interpretation depends strongly on sea-breeze-driven transport and coastal circulation evolution, additional evaluation of the simulated sea-breeze front against the available Doppler lidar wind observations would strengthen the manuscript. In particular, it would be helpful to assess whether the model accurately reproduces the timing, inland penetration, frontal position, and vertical structure of the observed sea-breeze circulation, since ozone transport pathways and recirculation patterns are highly sensitive to sea-breeze evolution for Long Island region.
(3) The interpretation of the July 28 ozone enhancement mechanism remains somewhat unconvincing to me. While the manuscript emphasizes residual-layer entrainment and overnight marine boundary-layer accumulation, the observational evidence for these mechanisms is not fully convincing. In particular, it is unclear whether elevated ozone concentrations were directly observed within the shallow marine boundary layer at the coastal sites. In addition, although the model simulations show elevated ozone aloft, I do not clearly see evidence of the subsequent downward transport or entrainment flux that would connect the elevated ozone reservoir to the later surface ozone enhancement.
Alternatively, the July 28 event may actually reflect coastal recirculation and convergence processes within the Long Island Sound region. The GCAS NO_2 and HCHO VCD observations suggest widespread photochemical processing across the coastal environment, while the spatial distribution of enhanced precursor levels appears consistent with convergence between the south shore sea breeze and opposing coastal/land-breeze circulations from Connecticut. Ozone produced in the urban downwind region could follow the south-shore sea-breeze and subsequently be transported toward the Long Island Sound region, where convergent coastal circulations could favor ozone accumulation and recirculation rather than primarily downward entrainment from elevated residual layers. If residual-layer entrainment is indeed an important contributor, it would be helpful if the model analysis could further quantify the relative contributions of residual-layer transport versus coastal convergence/recirculation processes to the observed July 28 ozone enhancement. Additional discussion of these potential convergence-driven ozone production mechanisms would strengthen the interpretation of the July 28 event.

Author Response:
We thank the reviewer for the thorough and constructive assessment of our manuscript. We agree that the July 28, 2023, interpretation was presented with more confidence than the available evidence supports, and we have revised the manuscript accordingly. We address both referee's comments together as they all pertain to the July 28 mechanism and the role of WRF-Chem in supporting it.

For the model evaluation of vertical O_3 structure (comment 1) and sea-breeze circulation (comment 2), we acknowledge these would strengthen confidence in the interpretation. However, given the scope of the current study and the campaign-period focus, we have opted to follow the suggestion from Referee 1 (comment 2) to revise the conclusions and language that rely primarily on WRF-Chem rather than add any new model evaluation or figures. The model is now explicitly presented as a 'gap-filling' and hypothesis role rather than to derive concrete conclusions.

Regarding the July 28th mechanism (comment 3), we have revised the discussion in Section 3.3.1. to better separate the direct observational evidence (from TOLNet), the WRF-Chem interpretation, and any remaining uncertainty. Specifically, TOLNet observations from July 27, 2023, (up to 00 UTC, new Fig. S8) provide direct observational evidence of elevated O_3 (75–105 ppb) aloft at the TROPOZ and TOPAZ lidars going into the overnight period. WRF-Chem curtains are now explicitly identified as the only evidence for the overnight period, with TOLNet observations starting up again at 12 UTC on July 28, 2023. We also agree that direct observational evidence of elevated O_3 within the shallow marine boundary layer is not available, and the marine boundary layer contribution is now explicitly framed as inferred. The coastal convergence and recirculation mechanism raised by Referee 2 has been incorporated as another reasonable alternative, but with a caveat that the available surface wind and Doppler lidar observations show only weak convergence signals insufficient to quantitatively distinguish between the leading mechanism. We would have to dedicate sensitivity simulations that are beyond the scope of the current study to really quantify any relative contributions of the residual layer entrainment vs. coastal convergence but we do note that this is as a direction for future work. The revised texts appear in Lines 408 - 428 and shown below.

Lines 408 - 428: To assess whether these mechanisms may have contributed to the pollution events on July 28, we examined the simulated O_3 curtains as well as the observed wind lidar profiles spanning the overnight and early morning period (Fig. 6, S8). TOLNet observations on July 27 show elevated O_3 (75–105 ppb) persisting aloft at TROPOZ and TOPAZ through the late afternoon and into the evening (up to ~00 UTC July 28), providing observational evidence that an O_3 rich layer was present going into the overnight period (Fig. S8). Direct overnight observations are unavailable, however, WRF-Chem simulated O_3 curtains suggest elevated O_3 concentrations (60 – 75 ppb) may have persisted within the residual layer (approximately 1 – 2 km) through the overnight hours at these same sites. Consistent with this, the TOLNet observed profiles beginning near 12–13 UTC on July 28, 2023, capture elevated O_3 levels just above the surface layer that deepen vertically by mid-afternoon as the PBL develops (Fig. 6c), which is consistent with entrainment of residual layer air into the growing daytime PBL, though direct observational confirmation of the overnight transport itself is limited. There is also an overnight shallow marine boundary layer most notably present at the Yale Coastal site and slightly at the Stony Brook site (Fig. 6d), however, direct observational evidence of elevated O_3 within this layer is not available, and its contribution remains inferred from the surface timing and spatial pattern of the coastal enhancements. An alternative or complementary mechanism involves coastal convergence and recirculation within the LIS region, where interactions between the south-shore sea breeze and opposing coastal flows from Connecticut could favor O_3 accumulation; surface wind observations show some evidence of convergence along the northern LIS shore, though the signal is not strong enough to conclusively distinguish this from the residual layer entrainment pathway.. These results are consistent with the early coastal O_3 enhancements on July 28, 2023, having been influenced by overnight pollutant accumulation in the residual layer and possibly shallow marine boundary layer, subsequently entrained into the daytime PBL during its initial growth that favored sustained coastal O_3 enhancements. However, given the reliance on model-simulated overnight structure and the plausibility of coastal convergence as a contributing factor, this interpretation should be regarded as a provisional hypothesis rather than a demonstrated mechanism.