We appreciate Reviewer 2 for giving insightful comments. We address the reviewer's comments below. Reviewer's comments are shown in normal fonts, while our response to each comment is shown in **bold characters**. Specific changes are highlighted in the track-changes file submitted with the revised manuscript, and these changes and their corresponding line numbers are also provided below.

Reviewer: The study analyzes the O_3 production regimes, OH reactivity, and atmospheric oxidative capacity (AOC) for a high-ozone episode in Houston using a large-eddy WRF simulation coupled with chemistry, which is named WRF-LES-Chem. The usage of WRF-LES-Chem is a highlight of the study, but all the primary analyses related to ozone formation and AOC are based on WRF-LES-Chem outputs, which are verbose and have been done in previous studies. I didn't find the necessity to use a large-eddy simulation based on the current results, especially when the results from the large-eddy simulation are similar to previous studies using mesoscale simulations. A possible way to improve the study is adding the O_3 formation and AOC analysis using the mesoscale simulation and comparing them with the results from the large-eddy simulation against observations and/or previous studies. The authors compared the large-eddy simulation and the mesoscale simulation against observed meteorology and concentrations of some chemical species, but those comparisons can't convince the community that we need a large-eddy simulation to investigate O_3 formation and AOC.

Response: We appreciate the reviewer's feedback and suggestions for improving the manuscript. The primary motivation for this paper is to understand the ongoing issue of high ozone (O_3) levels in Houston. Despite continuous emission controlling efforts, O_3 remains a significant problem in the region. Therefore, it is important to study the current O_3 formation regime more accurately. Previous studies have relied on mesoscale simulations for chemistry, whereas LES provides the advantage of better resolving turbulence and mixing. We use a WRF-LES-Chem approach, aiming to providing a more robust understanding of O_3 formation and AOC under better resolved boundary layer dynamics in Houston. Although at a compromised resolution (300 m), we find LES results, especially simulated surface concentrations of key chemicals, show better agreement with observations. Therefore, using LES results makes the O_3 formation analysis more accurate, which is why we do not use the mesoscale results for O_3 formation analysis and AOC calculations.

Reviewer: My second concern is the scaling of NEI2017 NOx emissions by 0.2 – 0.3 based on limited observations. Even if I can accept the scaling, why are the O_3 concentrations still much larger than the observations in Figure 6? Doesn't it mean that your scaling has problems? Notably, NOx emission has apparent spatial heterogeneity. The NEI2017 you used only has a resolution of 12 km, much coarser than the WRF-LES-Chem simulation at a resolution of 300 m. Therefore, much finer information about the NOx emission spatial distribution is missed. Without a compatible high-resolution emission inventory, WRF-LES-Chem can only provide better meteorological conditions, which is helpful but cannot justify the running of the chemical model at such a high resolution. A region with spatially homogeneous NOx emissions would be better to justify the usage of WRF-LES-Chem.

Response: We appreciate the reviewer's comments on the spatial heterogeneity of NOx emissions and the simulated higher O_3 concentrations compared to observations. The spatial extent of our innermost domain, which focuses on Houston, is substantially large for an LES. Therefore, using homogeneous NO_x emissions across the region may significantly deviate from reality due to the diverse emission sources unique to different subregions within the city. We attempt to reduce simulation biases by scaling NOx emissions, as these are key precursors in ozone chemistry. Ensuring that these emissions are at reasonable levels is essential for achieving a more accurate understanding of ozone chemistry. However, biases may still arise due to differences in grid-cell emission values between the mesoscale and LES, which are generated at different resolutions. The CESM2-WACCM model, used for the chemical IC & BC to drive the

simulation, is provided at a spatial resolution of 0.9° x 1.25°. This coarse resolution may introduce additional source of bias.

Indeed, the NEI is available at a coarser resolution and a compatible high-resolution emission inventory would be more suited for the finer resolution simulation to capture microscale emission sources. We acknowledge future work that implements high-resolution emissions inventories will be needed to improve the accuracy of O_3 simulations and minimize biases. We add this statement on lines 724-725 "*Using compatible high-resolution emissions inventories would better suit fine-resolution simulations, as they would more effectively capture microscale emission sources.*"

The application of LES with homogenous emissions has been well validated in previous studies, such as by Ouwersloot et al. (2011), Kasr et al. (2015), Li et al., 2016. However, the goal of this WRF-LES-Chem study is to more accurately simulate the spatial distribution of surface concentrations of ozone and its precursors in the complex Houson atmosphere. By leveraging better-resolved turbulence, this approach aims to improve the accuracy of simulated mixing and chemistry of key species under more realistic surface and boundary conditions.

Ouwersloot, H.G., Vilà-Guerau de Arellano, J., Van Heerwaarden, C.C., Ganzeveld, L.N., Krol, M.C. and Lelieveld, J., 2011. On the segregation of chemical species in a clear boundary layer over heterogeneous land surfaces. Atmospheric chemistry and physics, 11(20), pp.10681-10704.

Kaser, L., Karl, T., Yuan, B., Mauldin III, R.L., Cantrell, C.A., Guenther, A.B., Patton, E.G., Weinheimer, A.J., Knote, C., Orlando, J. and Emmons, L., 2015. Chemistry-turbulence interactions and mesoscale variability influence the cleansing efficiency of the atmosphere. Geophysical Research Letters, 42(24), pp.10-894.

Li, Y., Barth, M.C., Chen, G., Patton, E.G., Kim, S.W., Wisthaler, A., Mikoviny, T., Fried, A., Clark, R. and Steiner, A.L., 2016. Large-eddy simulation of biogenic VOC chemistry during the DISCOVER-AQ 2011 campaign. Journal of Geophysical Research: Atmospheres, 121(13), pp.8083-8105.

Reviewer: My third concern is the uncertainties of those observations used in the study. I suggest adding more details about those observations, e.g., how the data were measured, their accuracy and precision, and the assumptions used to derive the data.

Response: We thank the reviewer for the comments on observational uncertainties. Indeed, field observations (meteorology and chemical variables) are not without biases, which may arise from instrument calibration, the influence of external environmental conditions, sampling errors, and possible interference from other pollutants, etc. Although we do not focus on observational data analysis in this study since these datasets are publicly available and their full descriptions and uncertainties are thoroughly detailed in other publications, we recognize the importance of highlighting concerns regarding observational uncertainties. Therefore, we now provide additional details on the uncertainties in the observational data used in the Methods section in lines 217-224:

"*Additional in-situ observations for specific VOC species are measured by the Proton Transfer Reaction – Mass Spectrometry (PTR-MS) at the San Jacinto Battleground State Historical site (29.753 °N, 95.091 °W). As with any instrument, there are inherent uncertainties in these measurements. For example, NASA TROPOZ and LMOL O_3 lidars have about ±15 % error margin, and uncertainties may arise from data processing retrieval algorithms and DIAL wavelengths (Wang et al., 2017). Conversely, uncertainties in sonde retrievals are linked with radiosonde pressure offsets and retrieval algorithm calculations (Witte et al., 2018). The uncertainties in VOC concentrations measured by the PTR-MS range from 7.2 – 12.5%, with possible sources of error including instrument noise, sensitivity errors, and calibration standards (Shrestha et al., 2022).*"

Minor comments:

Reviewer: Line 89: What do you mean by this sensitivity regime? The 120 ppb O_3 concentration? If you referred to the general NOx-VOC-O_3 sensitivity regime, the papers you cited right before this sentence have shown some results. Why did you say that it is poorly understood in the Houston area? Please consider rewriting this sentence.

Response: We thank the reviewer for the observation. What we mean by the "sensitivity regime" is the variation of O_3 photochemistry to NO_X and VOC atmospheric concentrations, which is also referred to as the general NOx-VOC-O_3 sensitivity regime. Despite evidence from previous studies cited in the manuscript and continuous emission controlling efforts, elevated O_3 concentrations persist in the Houston area. For example, during the TRACER-AQ campaign, observed O_3 levels reached as high as 120 ppb. We rewrite the sentences on Lines 89-93 as "*Despite evidence from the TexAQS-2006 campaign that reducing NO_X emissions from power plants lowered O_3 concentrations (Cowling et al., 2007; Parrish et al., 2009), elevated O_3 concentrations persist in the Houston area, reaching as high as 120 ppb during the TRACER-AQ campaign in 2021. We further investigate the complex O_3 photochemistry that may have contributed to these elevated O_3 concentrations during this period, adding to existing evidence in the region*"

Reviewer: Lines 136-137: Do you mean that you ran two separate simulations, one with D01 and D02 and the other with D03 and D04?

Response: Yes, we perform two simulations separately. The first simulation with D01 and D02, while the second simulation is performed with D03 and D04 using the output from D02. We clarify this in the manuscript on Line 141: "*We conduct two separate simulations, where a two-way nesting simulation between domains D01 and D02 is used to drive another two-way nesting simulation between domains D03 and D04.*"

Reviewer: Lines 169-171: The simulation period has been mentioned in Lines 151-152.

Response: We thank the reviewer for this observation. The initial simulation period on Line 151 is used to describe the sensitivity test for the meteorological parameters (NCEP and HRRR). For clarity, we include the information on line 170 to describe the simulation length for all domains in the final simulation. We modify the sentence slightly for clarity on Lines 176-180 as "*In this study, simulations for the mesoscale domains (D01 – D02) cover the period from September 6, 2021 (00:00 UTC) to September 9, 2021 (00:00 UTC), which corresponds to Houston's local time from September 5, 2021 (19:00 Central Daylight Time, CDT) to September 8, 2021 (19:00 CDT), with the first 20 hours as the spin-up time. The output from D02 is used to drive parallel simulations for D03 and the LES domain and cover the period from 11:00 to 21:00 UTC (06:00 – 16:00 CDT) on September 8, 2021.*"

Reviewer: Line 183: Why did you use D01 but not D02?

Response: D01 is the parent domain and is not used to drive the LES simulation. Using ndown, D02 is used to drive D03 and D04 as we define in lines 181-183 "*The output from D02 is used to drive parallel simulations for D03 and the LES domain, covering the period from 11:00 to 21:00 UTC (06:00 – 16:00 CDT) on September 8, 2021.*" The sentence in line 192 "*Initial conditions and boundary conditions for each nested domain are generated from the parent domains, which are produced by the 8.1 km simulation*" is now removed, to avoid duplicate information as stated in Lines 193-194.

Reviewer: Lines 193-194: The scaling is either 0.3 or 0.2. How did you get a value of 11.32%, even smaller than 0.2? Or do you mean the speciation of NO and NO₂ emissions? In addition, did you use the TCEQ and Pandora observations when determining the scaling factors?

Response: We thank the reviewer for the observation. The scaling is done separately for NO using a scale factor of 0.3 and NO₂ using a scale factor of 0.2 from the NEI inventory. During the model testing phase, we compared simulated NO and NO₂ surface concentrations with TCEQ station observations to determine the scaling factor. The value of 11.32 % is the ratio of NO2/(NO+NO2), not the scaling factor.

Reviewer: Line 292: In Line 183, you said you used output from D01 as IC & BC for the inner domains. Why did you show D02 here?

Response: We show the output from D02, as it represents a less coarse mesoscale domain. D01 covers a larger parent domain but at a much coarser resolution, making D02, with a less coarse resolution, a more suitable choice. Additionally, D02 is used to drive the D03 and D04 domains. We clarify this on Lines 181-183 as "*The output from D02 is used to drive parallel simulations for D03 and the LES domain, covering the period from 11:00 to 21:00 UTC (06:00 – 16:00 CDT) on September 8, 2021*." This information on line 193, which we address as duplicate information in the last comment is now removed.

Reviewer: Figures 3 and S2. The first rows of the two figures have the same contours. Are you sure the two figures represent different times? In addition, subplot (j) in both figures is unclear. Please consider a more straightforward way to show wind directions.

Response: We appreciate the reviewer for this observation. Figure S2 is updated accordingly, with the upper panel showing the meteorology spatial distribution at 9:00 CDT. In subplot (j), the wind direction plot previously resulted from overlaying wind arrows at each grid cell due to the fine resolution and large number of grid cells in the LES. We have replaced the wind direction plots for D04 with those from the intermediate domain D03, which, despite its coarser resolution, shows a similar wind pattern due to two-way nesting feedback.

Reviewer: Lines 306-307: I see what you mean, but I'm not entirely convinced. The LES just shows finer structures with more significant spatial heterogeneity.

Response: The LES wind speed maps in Figure 3 indicate higher turbulence kinetic energy (TKE) and stronger mixing of the chemicals horizontally and vertically, resulting in a more robust distribution of chemical species and their subsequent chemical reactions in each grid cell

Reviewer: Line 322: "diurnal"? You only showed 9:00 and 15:00 CDT.

Response: We use "diurnal" to describe changes between morning and afternoon hours. For clarity, we revise this to specify exact times on Lines 342-343: "*However, the model well captures the variations of these chemical species at both morning and afternoon hours compared to observations.*"

Reviewer: Line 325: Delete "slightly."

Response: "slightly" deleted (now Line 347).

Reviewer: Lines 341-343: Figures 4 and S3 show larger HCHO in the morning than in the afternoon!

Response: We thank the reviewer for the comments on the HCHO daytime changes. We also find the argument interesting since we expect a larger HCHO concentration in the afternoon. However, when comparing with other observations, such as Pandora measurements at La Porte (Fig. S4), we observe higher HCHO levels in the morning compared to the afternoon. In the revised Fig. S5, we present a time series plot of selected VOCs compared to PTR-MS measurements, where HCHO shows good model agreement with observations. Nevertheless, we note that D02 significantly overestimates HCHO in the early morning, which could impact the D04 domain, as D02 output is used to drive D04. While D04 provides a more accurate representation of HCHO, as shown in Fig. S4 and S5, a finer-resolution LES could yield even better HCHO observations.

Reviewer: Line 348: Are isoprene, MEK, and xylene concentrations better in D04 than in D02 compared to observations? Figure 5 doesn't show that!

Response: We thank the reviewer for the comment. We also provide a new Figure S5 that shows a time series of VOCs including isoprene and MEK. This is discussed in Lines 381-390 as "*Figure S5 compares the simulated diurnal variability of four VOCs with observations from the PTR-MS. Isoprene is underestimated in the LES domain D04 in the morning, but increasing concentrations are simulated in the afternoon, reducing biases and aligning more closely with observations. The simulated diurnal pattern of HCHO in the LES domain D04 shows better agreement with observations. MEK is overestimated in the morning across both the LES and mesoscale domains, but modeled concentrations become more aligned with observations in the afternoon, with both observations and modeled results indicating high levels in the morning and decreasing concentrations throughout the day. Acetaldehyde, however, is underestimated in the afternoon in both domains, suggesting missing secondary chemical sources from lumped alkanes (Read et al., 2012). It should be noted that the PTR-MS measurement accuracy may be constrained by uncertainties in calibration standards and compound identification based on mass information only (Ammann et al., 2004).*"

Reviewer: Lines 394-396: Firstly, the relative bias is larger for the LES simulation than the mesoscale simulation. Secondly, I can't believe that a time lag of less than 20 minutes can explain such a large model bias in ozone concentrations throughout the boundary layer.

Response: Although the LES bias for O_3 is higher than that of the mesoscale simulation in the profile comparison shown in Figure 6, other evaluations, including lidar comparison (Fig S6), surface comparisons (Fig 4 and S3), and time-series plots (Fig S4 and S5), demonstrate better performance of the LES. The time lag accounts for only part of the differences observed in the model-observation profile comparisons. In the afternoon, when convection is strong, the time for a single convective overturn (~15 mins) may intermittently influence pollutant concentrations due to the influence of resolved eddies. Most importantly, different measurement instruments used to collect observational data have varying uncertainties, which may also contribute to biases when compared with models.

We highlight some of these measurement uncertainties in lines 218-225 as "*Additional in-situ observations for specific VOC species are measured by the Proton Transfer Reaction – Mass Spectrometry (PTR-MS) at the San Jacinto Battleground State Historical site (29.753 °N, 95.091 °W). As with any instrument, there are inherent uncertainties in these measurements. For example, NASA TROPOZ and LMOL O_3 lidars have about ±15 % error margin, and potential uncertainties may arise from data processing retrieval algorithms and DIAL wavelengths (Wang et al., 2017). Conversely, uncertainties in sonde retrievals are linked with radiosonde pressure offsets and retrieval algorithm calculations (Witte et al., 2018). The uncertainties in VOC concentrations measured by the PTR-MS range from 7.2 – 12.5%, and with possible sources of error including instrument noise, sensitivity errors, and calibration standards (Shrestha et al., 2022)..*"

Also, we revise Lines 405-406 to clarify this bias and provide explanations as "*Although the LES performs better when compard with other measurements, such as lidar data and surface observations, the LES domain D04 overestimates surface O_3 mixing ratios by about 15 ppb compared to sonde measurments, while the mesoscale domain D02 underestimates by 5 ppb at both sites. This suggests different measurement instruments used to collect observational data have varying uncertainties, which may contribute to biases when compared with models.*"

Reviewer: Figure S5. Could you please show the model results and observations consistently to facilitate comparison? It is really hard to read the figure. Anyway, the figure indicates that LES better captures the observed evolution patterns compared to the mesoscale simulation. Do you know the uncertainties of the NASA Lidar O_3 measurements?

Response: We update Fig. S5 (now Fig. S6) to provide a clearer comparison between Lidar and model values. The NASA TOLNet Lidars for O_3 measurements are subject to uncertainties due to measurement and retrieval techniques. A study by Wang et al. (2017) indicates that NASA TROPOZ, LMOL, and NOAA TOPAZ Lidars have an accuracy of about +/- 15% within the troposphere. The uncertainties are mainly associated with retrieval algorithms and the selection of DIAL wavelengths, which affect sensitivity to interferences by other chemical species.

Wang, L., Newchurch, M. J., Alvarez II, R. J., Berkoff, T. A., Brown, S. S., Carrion, W., De Young, R. J., Johnson, B. J., Ganoe, R., Gronoff, G., Kirgis, G., Kuang, S., Langford, A. O., Leblanc, T., McDuffie, E. E., McGee, T. J., Pliutau, D., Senff, C. J., Sullivan, J. T., Sumnicht, G., Twigg, L. W., and Weinheimer, A. J.: Quantifying TOLNet ozone lidar accuracy during the 2014 DISCOVER-AQ and FRAPPÉ campaigns, Atmos. Meas. Tech., 10, 3865–3876, https://doi.org/10.5194/amt-10-3865-2017, 2017.

Reviewer: Lines 409-411: Do you have any explanations for that? This is out of my expectation. I think it is because Houston is heavily polluted. Photochemical production of O_3 in the boundary layer exceeds the default O_3 positive vertical gradient. But I don't understand why the mesoscale simulation can't reproduce it, considering NEI2017 has a resolution of 12 km, coarser than the mesoscale simulation.

Response: We appreciate the reviewer's concerns regarding the statement on Lines 409-411 of the manuscript, which refers to Fig. 7. Our explanation for the domain differences focuses on vertical mixing variations between D02 and D04. Both domains show a positive vertical O_3 gradient above the PBL, with these changes primarily driven by pollutant vertical mixing within the PBL, which also impacts O_3 photochemical production. The differences between LES and mesoscale simulations could be mainly due to turbulence-chemistry interactions. We clarify this in the manuscript in lines 450-455 "*In general, the differences between LES and mesoscale simulations may largely stem from turbulence-chemistry interactions. These interactions can disrupt chemical processes by influencing the production of secondary compounds, which rely not only on NO_X emission variations but also on meteorological factors such transport flow rate and the mixing of primary pollutants. Due to the limited availability of observational data, we do not include model evaluations for turbulence parameters. However, the LES wind speed maps in Fig. 3 suggest higher TKE and stronger mixing of chemicals both horizontally and vertically, resulting in a more robust distribution of chemical species and their subsequent chemical reactions in each grid cell*."

Reviewer: Figure 9. Please change the legends to make the plots more straightforward.

Response: We apologize as we are not sure what changes the reviewer suggested. However, we revise Fig. 9 and include the time 9:00 CDT and 15:00 CDT to make the figure easier to interpret. We added another plot here with some changes to the legend in subplot(c).
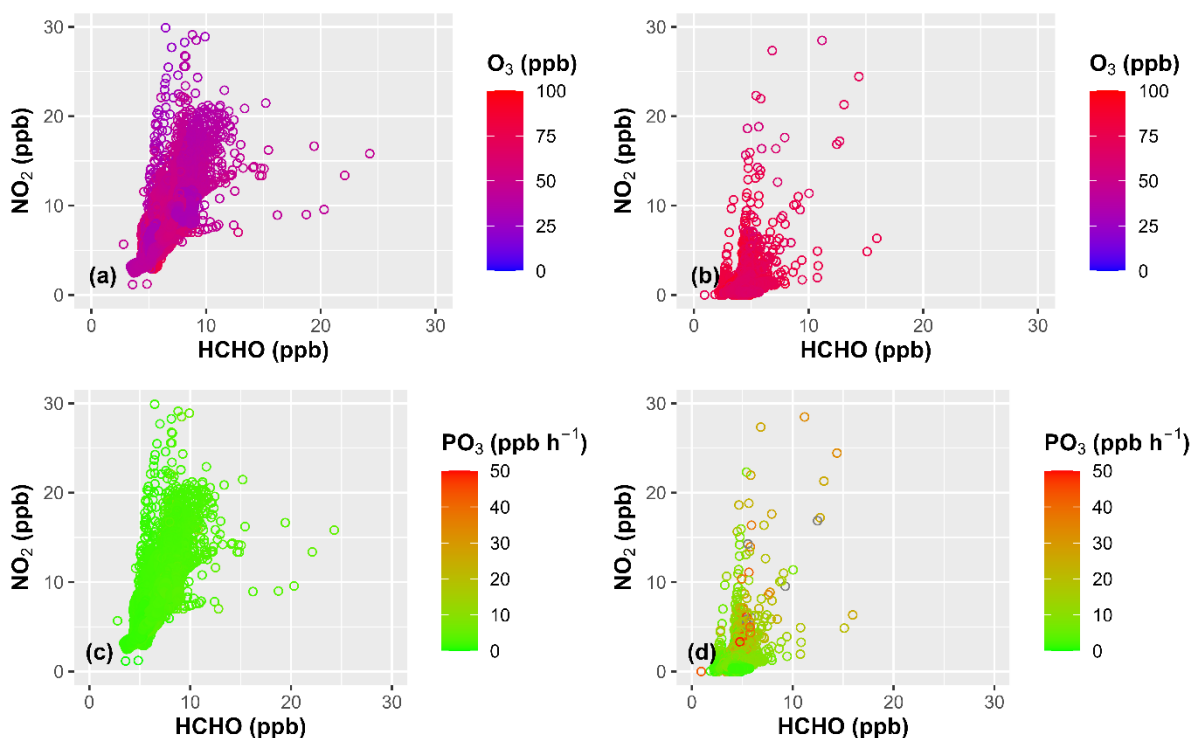
Reviewer: Lines 484-486: Please rewrite this sentence.

Response: The sentence is revised on Lines 527-528 as "*O_3 formation is mainly dominated by contributions from point and mobile emission sources in La Porte, downtown Houston, and Texas City near Galveston (Hossan et al., 2021).*"

Reviewer: Line 497: Photodissociation of NOx? Do you mean NO_2 -> NO + O? The primary loss of NOx in the daytime is NO_2 + OH -> HNO_3.

Response: We thank the reviewer for this reminder. Yes, the primary daytime loss of NO_X is via NO_2 + OH, and both are possible in the daytime. We revise the sentence on Lines 541-543 as "*Notably, NO_X levels are generally low from late morning into the afternoon, highlighting the diurnal variability of NO_X emission sources as well as the oxidation of NO_2 by OH and the fast dissociation of NO_X.*"

Reviewer: Line 514: Isn't it due to the limited vertical mixing of surface-emitted NOx into the upper boundary layer in the morning?

Response: We thank the reviewer for highlighting the role of the limited vertical mixing. The sentence on Lines 558 – 560 is now revised as "*We find that the production of O_3 is under a VOC-limited regime at the surface in the morning and transitions into a NO_X-limited regime in the upper atmosphere (> 200 m AGL) due to the limited vertical mixing of surface-emitted NO_X into the upper boundary layer in the morning.*"

Reviewer: Lines 574-580: I don't understand the logic. You cited two papers with different conclusions about the primary sources of HCHO. Then why did you choose the second paper as the basis of your analysis?

Response: We thank the reviewer for the observation. The conclusions from the two papers cited (Johansson et al., 2014 and Parrish et al., 2012) are not entirely different as both agree

that industrial emissions are the primary source of HCHO. Johansson et al. (2014) identified industrial emissions as the main source of primary HCHO, aligning with Parrish et al. (2012), which highlighted petrochemical facilities as key industrial sources of primary HCHO. However, the secondary production of HCHO, primarily from VOC oxidation, is the predominant source of total HCHO (including both primary and secondary sources), as highlighted in the second paper. To clarify this, we modify the sentences on Lines 622-625 as *"Industrial and mobile emissions are identified as the main sources of primary HCHO in the Houston area (Guven and Olaguer, 2011; Johansson et al., 2014), with an average emission rate of 8631.02 mol km^{-2} hr^{-1} estimated over the Houston region during the simulation period from the NEI inventory. Additionally, a comprehensive chemical analysis of airborne and ground measurements of HCHO and emissions data concluded that the predominant source of total HCHO in the Houston area is secondary production from VOC oxidation (Parrish et al., 2012)."*