Response to Zhao et al., ACPD, 2026

RC1

In "Inefficient Consumption of Natural Gas Drives Methane Emissions from a Megacity," Zhao et al. utilize multiple measurement platforms to investigate the sources of methane emissions in the New York City Metropolitan Area. The observations presented in this work represent a dense dataset from which much can be learned about the largest metropolitan area in the United States. Understanding the sources of methane emissions in cities is key to successful mitigation strategies. Accordingly, I find the study timely and of interest to a broad audience.

The study is very thoroughly presented with clear figures and method descriptions. However, I believe there are some areas that should be addressed prior to publication.

1. Section 2.6. This section lacks citations for the transport models and their application in this work. Another option could be to reference the later section, 2.6.3.

   We thank the reviewer for this comment. The citations were added in the main text:

   "We calculated hourly simulated methane enhancements (ΔCH_4, ppb) by convolving a series of global, national, and regional methane emission inventories (units: nmol m−2 s−1) with surface influence footprints (ppb (nmol m−2 s−1)−1) derived from a Lagrangian atmospheric transport model described in Section 2.6.3 (Fasoli et al., 2018; Lin et al., 2003)."

2. Line 302: The term "suburban outflow" is first introduced here, but it wasn't clear to me what this means. It appears to reference the difference between this work and that of Schiferl et al. 2025, but it wasn't clear how they are different. I assume it has to do with the footprint of the Mineola Tower versus that of what is used in Schiferl at al.

We thank the reviewer for this valuable comment. To clarify our use of the term "suburban outflow": Schiferl et al. (2025) employ a rooftop observation site located within Manhattan's urban core (Upper Harlem), whereas our Mineola Tower site is located in a suburban setting on Long Island. Under westerly wind conditions, Mineola Tower is capable of detecting urban emission signals from New York City (see Figure 1 in the main text for the site footprint). Notably, the footprint of our study overlaps with that of Schiferl et al. (2025) (See the Figure 1 in the main text and Figure 4a in Schiferl et al., 2025), providing a basis for comparison while also highlighting the contrasting perspectives offered by a near-source urban core site versus a downwind suburban receptor site. In the main text, we revised the text to be clearer in line 310:

"The seasonal cycle of NYCMA methane emissions quantified ~~for the suburban outflow~~ using suburban Mineola Tower measurements is consistent with the ~~city core methane~~ emissions from Schiferl et al. (2025) using city core observations, which saw peak methane emissions in January and February, followed by a decline into spring {Schiferl et al.,2025}.

The 75% percentile footprint at Mineola Tower encompasses the majority of the urban areas within the NYCMA (Figure 1), including northeastern New Jersey, Manhattan, the Bronx, and western Long Island. This footprint largely overlaps with the 75% percentile footprint reported by Schiferl et al. (2025) for their urban core site, providing a robust spatial basis for comparison between the two studies {Schiferl et al.,2025}."

3. There is clearly significant added value from concurrent measurements of CH_4, CO, and C_2H_6. However, is it fair to attribute all the variability in observed CH_4:CO and CH_4:C_2H_6 to source behavior? If I understand correctly, "plumes" in this work correspond to relatively short time periods (as suggested by Figure 5b) and each undergo the slope analysis. If that is the case, is it possible they have experienced different levels of atmospheric mixing, such that, depending on the sources, that could influence the derived CH_4:CO and C_2H_6:CH_4 slopes? So, is it fair to completely attribute all the observed variability to combustion efficiency in the statement below:

Line 332: "Thermogenic plumes with moderate R2 ΔCO:ΔCH_4 (between 0.1 and 0.8) are observed more frequently during months with building heating loads (Figure S12), suggesting greater variability in combustion efficiency, likely driven by various natural gas combustion processes for heating purposes."

We thank the reviewer for this valuable comment and apologize for the lack of clarity in the original manuscript. To clarify our methodology: we use the linear correlation (R^2) of methane and ethane enhancements, rather than the slopes, to infer different sources. We also use the correlation between trace gas enhancements calculated at 1 Hz after removing the background (defined as the 5th percentile of the 1 Hz dry mole fractions over a centered 48-hour running window) to reduce the influence of regional background changes, which we assume vary on timescales longer than 5 minutes. Furthermore, when discussing the CO-to-ethane relationship, where we also only discuss the correlation rather than slope, we only use thermogenic plumes identified by RMA regression with $R^2_{\Delta C2H6:\Delta CH4} \geq 0.95$, which indicates air masses with limited atmospheric mixing. This two-step approach minimizes the influence of background variability and dilution when quantifying thermogenic air masses.

We clarified the method in lines 151 to 155:

"*To identify thermogenic methane plumes, we calculated the Reduced Major Axis (RMA) ~~fit~~ correlation between 1 Hz ethane enhancements (ΔC_2H_6) and methane enhancements (ΔCH_4) for 5-minute tower measurement intervals ($R^2_{\Delta C2H6:\Delta CH4}$), which considers variability in both x and y axes. To identify times of incomplete combustion, we calculated the ~~fit~~ RMA correlation between 1Hz CO enhancements (ΔCO) and ΔCH_4 ($R^2_{\Delta CO:\Delta CH4}$).*"

Line 336: “In contrast, only 7% (summer) and 6% (winter) of the identified thermogenic plumes were observed to have no combustion signature (R2 ΔCO:ΔCH4 < 0.1), showing that natural gas emissions from pipeline leaks were less frequently observed at the Mineola tower compared to combustion-related natural gas sources.” What was the reasoning for using 0.1 as the cutoff here? For lower correlations between CO and CH4, one might expect noise/uncertainty to have a larger effect on the calculated R2 value. Would the conclusions change if you used a larger threshold for these conclusions?

We thank the reviewer for the insightful comment regarding the $R^2_{\Delta CO:\Delta CH4}$ threshold, which was also mentioned by Reviewer 2. The $R^2_{\Delta CO:\Delta CH4}$ threshold was selected to be consistent with our thermogenic plume identification criterion of $R^2_{\Delta C2H6:\Delta CH4} > 0.95$. To evaluate the robustness of this threshold, we examined the distribution of $R^2_{\Delta CO:\Delta CH4}$ for all 5-minute thermogenic plumes. As shown in the attached histogram and in Figure S12, the distribution peaks near $R^2_{\Delta CO:\Delta CH4}=1$, indicating that many thermogenic plumes exhibit strong $R^2_{CO:CH4}$ correlations and therefore likely include a combustion influence. A smaller peak is observed near $R^2=0$ in summer; however, when the cutoff for identifying plumes without a combustion signature is increased above 0.1, the overall conclusion does not change: most thermogenic plumes remain associated with a combustion-related ΔCO:ΔCH4 relationship, while a smaller fraction shows no clear combustion signal.

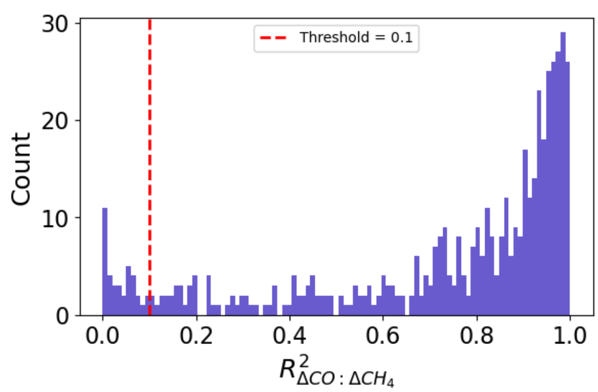

Figure 1. Histogram of $R^2_{CO:CH4}$ between 1 Hz ΔCO:ΔCH4 within each 5 minute interval for summer 2023.

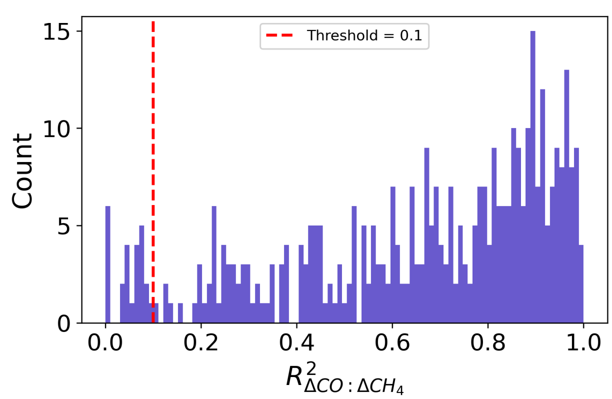

Figure 2. Histogram of $R^2_{CO:CH4}$ between 1 Hz ΔCO:ΔCH4 within each 5-minute interval for winter 2024.

In supplemental section 3, the authors examine street-level data of two pipeline leaks, with high correlation between CH4 and C2H6, but not CO. Were there other instances of strong correlations between CH4, C2H6, and CO in the street-level data? This would indicate incomplete combustion of natural gas and support the attribution conclusions since the transport time would be relatively short. In the current manuscript, it wasn't clear to me how the street-level observations were incorporated into the larger study.

We thank the reviewer for this comment. The mobile measurements were conducted to serve three specific purposes:

1. **Quantifying the local pipeline ethane-to-methane ratio.** The mobile drive allowed us to directly measure the ethane:methane ratio of local natural gas pipeline leaks, providing a local tracer signature for comparison with tower observations.
2. **Assessing the presence of natural gas combustion signals at street level.** Exhaust from natural gas combustion is typically emitted from elevated heights, such as building chimneys, rather than at street level. Mobile sampling was conducted during daytime conditions when the boundary layer is relatively well-mixed, meaning hot combustion plumes are unlikely to descend and be sampled at street level. Consistent with this expectation, we measured natural gas leaks but no natural gas related combustion signals were detected in our street-level mobile dataset.
3. **Evaluating the method that distinguishes between combustion and natural gas leak sources using the correlation between trace gases.** Street-level sampling is more sensitive to near-surface sources, including natural gas pipeline leaks and CO emissions from traffic. Traffic-related CO is not co-emitted with methane, but different sources can be mixed during mobile sampling. The presence of nearby CO sources and methane enhancements from different sources does not usually produce strong correlations at high temporal resolution, indicating that even in environments where nearby CO sources and methane enhancements from different sources are common, **we do not observe strong methane and CO correlations at 1 Hz level observations.** This further supports our method of using 1 Hz tracer–tracer correlations to inform potential incomplete combustion of natural gas.

We clarified and revised our manuscript about mobile sampling in the main text Line 105:

"*To complement our tower-based observations, we conducted street-level sampling of 1 Hz methane, ethane, CO, and N2O dry mole fractions around the Mineola tower using the New York Atmospheric Composition and Air Quality (NYAAQ) Mobile Laboratory on August 16th, 2023. More details are described in Section S3 and Figure S2. Trace gas correlations were used to characterize methane sources in the area around the tower. We used mobile sampling to assess if natural gas combustion signals could be detected at* *street level and to evaluate whether the 1 Hz tracer–tracer correlation method (Section 2.5) can distinguish natural gas signals from nearby sources, such as traffic-related CO (detailed information can be found in supporting information Section S3).*"

And in the Supporting Information Line 55:

"*During the mobile sampling drive (Figure \ref{Fig:S2}), we observed two pipeline leaks with strong correlations between methane and ethane, with no correlations between methane and CO despite nearby sources such as traffic. We did not see consistent street-level co-peaks of CO and natural gas (methane and ethane). Since natural gas combustion exhaust is typically emitted at heights far above street level (e.g., from building chimneys), daytime street-level mobile sampling rarely interceptes these elevated plumes, and we therefore did not observe methane plumes strongly correlated with CO. Instead, our measurements primarily captured near-surface sources, and the absence of strong 1 Hz* $R^2_{\Delta CO:\Delta CH4}$*, despite CO being a near-surface traffic tracer, supports the use of high-frequency tracer–tracer correlations to distinguish natural gas leaks from incomplete natural gas combustion.*"

RC2

The authors used ~2 years of tower-based measurements (1 urban site, 3 rural) of CO_2 and CH_4 mole fraction in the NYCMA, plus tower measurements from a targeted campaign measuring CH_4, C_2H_6, CO, and other constituents. They used measurements to develop scaling factors (SF) for enhancements simulated using a footprint model and inventory estimates for each month, then multiplied the inventory estimate by the scaling factor to produce an observation-informed emissions estimate. They also used $\Delta C_2H_6:\Delta CH_4$ and $\Delta CO:\Delta CH_4$ ratios to characterize CH_4 sources in the NYCMA. They identified a seasonal trend in CH_4 emissions, that thermogenic sources dominate year-round, and that incomplete combustion of natural gas is a major source of CH_4 emissions.

This manuscript describes a thorough analysis of a unique dataset that, in my opinion, makes a valuable contribution to the field. It provides useful insight on urban CH_4 sources through characterization of CH_4 sources in the largest city in the U.S. I recommend this manuscript for publication following some revisions described below.

Major comments:

1. In the methods, the authors explain that they used three inventories to compute simulated enhancements. Then, they compute the SFs for an inventory and use that to compute an observation-informed emissions estimate. But it is unclear if just one inventory's simulated enhancements were used for the SFs and emissions estimates, and if so, which one, or whether all three were used and somehow incorporated into the uncertainty. I suggest the authors clarify this important detail.

   We thank the reviewer for this comment, and we are sorry this was not clear in the original manuscript. We are using all the regional, national and global methane inventories to quantify the observation-informed emissions. We revised the methods section as below.

   "We calculated observation-informed methane emissions by applying a simulated scaling factor (SF) to the corresponding regional, national and global inventories over the NYCMA. The SF allows us to capture biases in inventory methane emission fluxes while minimizing the impact of meteorological variability, and is widely applied in other studies over urban areas \citep{schiferl_missing_2025,schiferl_multi-year_2024,sargent_majority_2021}. "

   Regarding the uncertainties, please see Section 2.7:

   "*We quantified monthly observation-informed methane emissions for the afternoon only (12:00 to 18:00 EDT) and calculated the mean and 95% CI for each month by bootstrapping from the hourly observation-informed methane emissions derived from all inventories during the afternoon (12 - 18 h), The 95% CI accounts for* *uncertainties in the ensemble of various methane inventories, background calculations, and variations in hourly observed and simulated ∆CH4."*

2. The overall language about incomplete combustion being a major source is not very quantitative (i.e., what fraction of emissions are associated with incomplete combustion), but incomplete combustion being the driver of emissions is presented as the major finding of this study. I am referring particularly at the paragraph at lines 328-343, but this is also a major theme of the paper. I certainly agree that your results suggest that incomplete combustion is an important source (and that itself is a significant finding worthy of publication). I would say a dominant source is >50%. It would help if the authors would be more quantitative about this – for example, what fraction of thermogenic methane plumes have an $R^2_{\Delta CO:\ \Delta CH4}$ above a threshold that definitively points to incomplete combustion? If the authors don't have that kind of metric yet (I believe using ΔCO: ΔCH_4 ratios for this purpose is a relatively new practice), I suggest being more careful about the wording.
    1. In particular, lines 336-337: $R^2 < 0.1$ for ΔCO: ΔCH4 seems low to illustrate the point being made in this sentence. Even in the biogenic example in Fig. 3, the ΔCO: ΔCH_4 R^2 is 0.22. I understand maybe there is a range above 0.1 that is more ambiguous on whether there's a combustion signature or not, but this comparison implies everything above 0.1 has a combustion signature. Maybe focusing on R^2 values above a more definitive threshold would be less ambiguous.

    We thank the reviewer for the comments. In this manuscript, we found that the median of the $R^2_{CO:CH4}$ was~ 0.7 (0.72 in February and 0.73 in May) during the heating season and ~0.9 (0.86 in June, 0.87 in July, and 0.93 in August) during the cooling season, which indicates that the majority (larger than 50%) of the thermogenic plumes (for which the $R^2_{C2H6:CH4}$ are already larger than 0.95) had a strong combustion signal.

    We revised the manuscript as follows, changing "the majority of $R^2_{\Delta CO:\Delta CH4}$ exceeding 0.9 during the cooling season" to "exceeding 0.8." **The underlying values are unchanged; only the phrasing was revised.** Previously, we reported the cooling-season median $R^2_{\Delta CO:\Delta CH4} \approx 0.9$ (0.86 in June, 0.87 in July, and 0.93 in August, shown in Figure S12). Since a "majority exceed" statement requires the threshold to lie below the median, the cooling-season threshold is stated as 0.8, consistent with a median of ≈0.9. The heating-season threshold ($R^2_{\Delta CO:\Delta CH4} \approx 0.7$) follows the same convention (median $R^2_{\Delta CO:\Delta CH4}$ was 0.72 in February and 0.73 in May shown in Figure S12):

    "The majority of $R^2_{\Delta CO:\Delta CH4}$ values exceed 0.7 during months with building heating demand ~~and exceed 0.9~~ (median of thermogenic $R^2_{\Delta CO:\Delta CH4}$ >0.7

during heating season), and exceed 0.8 during months with building cooling demand (median of thermogenic $R^2_{\Delta CO:\Delta CH4}$ > 0.8 during cooling season) (Figure S12).”

We thank the reviewer for the insightful comment regarding the $R^2_{CO:CH4}$ threshold, which was also mentioned by Reviewer 1. The $R^2_{\Delta CO:\Delta CH4}$ threshold was selected to be consistent with our thermogenic plume identification criterion of $R^2_{\Delta C2H6:\Delta CH4}$ larger than 0.95. To evaluate the robustness of this threshold, we examined the distribution of $R^2_{\Delta CO:\Delta CH4}$ for all 5-minute thermogenic plumes. As shown in the attached histogram and in Figure S12, the distribution peaks near $R^2_{\Delta CO:\Delta CH4}$=1, indicating that many thermogenic plumes exhibit strong $R^2_{CO:CH4}$ correlations and therefore likely include a combustion influence. A smaller peak is observed near R^2=0 in summer; however, when the cutoff for identifying plumes without a combustion signature is increased above 0.1, the overall conclusion does not change: most thermogenic plumes remain associated with a combustion-related ΔCO:ΔCH4 relationship, while a smaller fraction shows no clear combustion signal.

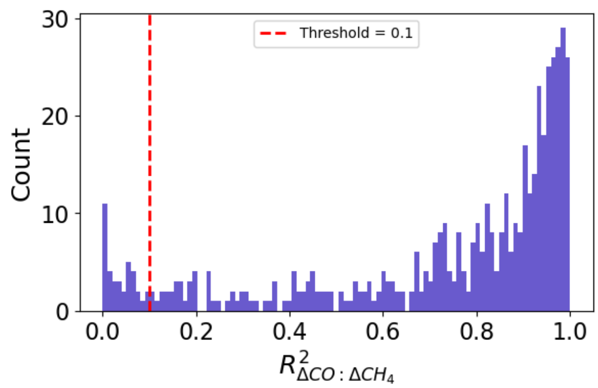

Figure 1. Histogram of $R^2_{CO:CH4}$ between 1 Hz ΔCO:ΔCH4 within each 5 minute interval for summer 2023.

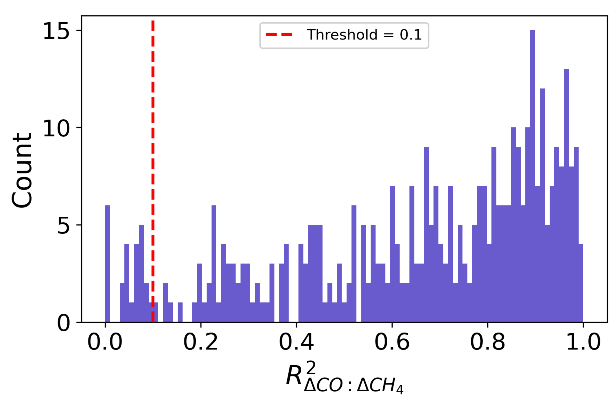

Figure 2. Histogram of $R^2_{CO:CH4}$ between 1 Hz ΔCO:ΔCH4 within each 5-minute interval for winter 2024.

3. It is unclear how the data from the Aug 16 mobile measurement campaign contribute to the study. I suggest leaving this out of the paper.

   We thank the reviewer for this comment, which was also mentioned by Reviewer 1 (above). The mobile measurements were conducted to serve three specific purposes:

   1. **Quantifying the local pipeline ethane-to-methane ratio.** The mobile drive allowed us to directly measure the ethane:methane ratio of local natural gas pipeline leaks, providing a local tracer signature for comparison with tower observations.
   2. **Assessing the presence of natural gas combustion signals at street level.** Exhaust from natural gas combustion is typically emitted from elevated heights, such as building chimneys, rather than at street level. Mobile sampling was conducted during daytime conditions when the boundary layer is relatively well-mixed, meaning hot combustion plumes are unlikely to descend and be sampled at street level. Consistent with this expectation, we measured natural gas leaks but no natural gas related combustion signals were detected in our street-level mobile dataset.
   3. **Evaluating the method that distinguishes between combustion and natural gas leak sources using the correlation between trace gases.** Street-level sampling is more sensitive to near-surface sources, including natural gas pipeline leaks and CO emissions from traffic. Traffic-related CO is not co-emitted with methane, but different sources can be mixed during mobile sampling. The presence of nearby CO sources and methane enhancements from different sources does not usually produce strong correlations at high temporal resolution, indicating that even in environments where nearby CO sources and methane enhancements from different sources are common, we do not observe strong methane and CO correlations at 1 Hz level observations. This further supports our method of using 1 Hz tracer–tracer correlations to inform potential incomplete combustion of natural gas.

   We clarified and revised our manuscript about mobile sampling in the main text Line 105:

   "*To complement our tower-based observations, we conducted street-level sampling of 1 Hz methane, ethane, CO, and N2O dry mole fractions around the Mineola tower using the New York Atmospheric Composition and Air Quality (NYAAQ) Mobile*

Laboratory on August 16th, 2023. More details are described in Section S3 and Figure S2. Trace gas correlations were used to characterize methane sources in the area around the tower. We used mobile sampling to assess whether natural gas combustion signals could be detected at street level and to evaluate whether the 1 Hz tracer–tracer correlation method (Section 2.5) can distinguish natural gas signals from nearby sources, such as traffic-related CO (detailed information can be found in supporting information Section S3)."

And in the Supporting Information Line 55:

"*During the mobile sampling drive (Figure \ref{Fig:S2}), we observed two pipeline leaks with strong correlations between methane and ethane, with no correlations between methane and CO despite nearby sources such as traffic. We did not see consistent street-level co-peaks of CO and natural gas (methane and ethane). Since natural gas combustion exhaust is typically emitted above street level (e.g., from building chimneys), daytime street-level mobile sampling rarely intercepted these elevated plumes, and we therefore did not observe methane plumes strongly correlated with CO. Instead, our measurements primarily captured near-surface sources, and the absence of strong 1 Hz $R^2_{CO:CH4}$, despite CO being a near-surface traffic tracer, supports the use of high-frequency tracer–tracer correlations to distinguish natural gas leaks from incomplete natural gas combustion.*"

4. I strongly suggest the authors add some discussion and a figure describing the footprint of their estimates (footprint size and location, what lies in the footprint, etc.). This would be useful generally, but especially helpful to understand the comparison to Schiferl et al. in lines 302-312.

   We thank the reviewer for the valuable comments. We revised the manuscript as follows:

   Line 303 to 312:" The seasonal cycle of NYCMA methane emissions quantified ~~for the suburban outflow~~ using suburban Mineola Tower measurements is consistent with the ~~city core methane~~ emissions from Schiferl et al. (2025) using city core observations, which saw peak methane emissions in January and February, followed by a decline into spring\citep{schiferl_missing_2025}. The 75% percentile footprint at Mineola Tower encompasses the majority of the urban areas within the NYCMA (Figure 1), including northeastern New Jersey, Manhattan, the Bronx, and western Long Island. This footprint largely overlaps with the 75% percentile footprint reported by Schiferl et al. (2025) for their urban core site, providing a robust spatial basis for comparison between the two studies {Schiferl et al., 2025}."

5. Section 2.4, Figure 1: Could the authors elaborate on why they chose this bootstrapping method to determine the background? It seems like, with multiple background sites, it would be appropriate to use a transport model to determine which background site to use. On a similar note, what is the predominant wind direction? It seems like W/SW winds may not be represented by these background sites. Of course, it's not realistic to have a background site in every direction, but I suggest the authors add some discussion about this to the manuscript.

   We thank the reviewer for this comment. During the course of this study, we evaluated several approaches for estimating background concentrations. One approach is to use HRRR-STILT transport model footprints to select upwind background sites on an hourly basis as the reviewer suggests; however, this method presented several challenges. First, it is difficult to quantify the background uncertainties associated with the transport model. Second, we do not have background measurement sites in all directions. As the reviewer mentions, this task is physically unrealistic. Having limited background sites limits our ability to constrain background conditions across varying wind regimes. Third, the HRRR-STILT model is designed to simulate hourly footprints at the top of each hour; sharp transitions in wind directions cause the background site selection to jump from one background site to the other, causing unrealistic sharp changes in background methane concentrations, introducing additional representation errors that are difficult to characterize and quantify. We therefore adopted a bootstrap-based statistical approach because, unlike the transport-model-based approach, it does not introduce uncertainties associated with transport model simulations, while still statistically accounting for temporal variability and inter-site variability due to changing wind conditions in the background uncertainty.

6. Lines 243-247: Why compare the monthly mean diel cycle? Based on the text, it seems like the authors are using this to evaluate and compare the inventories. However, it seems like comparing afternoon hours only would be more appropriate given the higher model uncertainty at other times of day. (Minor note: this paragraph is also missing the reference to the figure number)

   We thank the reviewer for this comment and apologize for the lack of clarity in the original manuscript. The comparison between observed and simulated diel cycles was conducted primarily to evaluate the performance of the HRRR-STILT transport model across different hours of the day, and the goal of minimizing the sensitivity of our final results to model configuration choices. To isolate the effect of transport model configuration, these comparisons were performed using a single emissions inventory while systematically varying the model configuration. Through this analysis, we identified the hours during which model performance was relatively insensitive to the choice of configuration, which we found to be between noon and 18:00 local time. These hours and their corresponding configurations were subsequently used to calculate emissions scaling factors against regional, national, and global inventories over the NYCMA.

We revised our manuscript Line 230:

“It is difficult to directly quantify the uncertainties in the surface influence footprint calculated by the Lagrangian transport model. Instead, we conducted a sensitivity study using alternative configurations of the atmospheric transport model and convolved the resulting surface influence footprints with the regional methane inventory to simulate a range of methane enhancements (Eqn \ref{eq:eqSim}). The observed and simulated monthly mean diel cycles were compared solely to evaluate the sensitivity of the HRRR-STILT transport model to different model configurations and to identify the hours with the most robust model performance for subsequent emissions scaling.”

Minor comments:

1. Title: did the authors mean to say inefficient "combustion" instead of "consumption"?

   We did not. As the reviewers noted, we can’t attribute all the methane emissions to incomplete combustion when much of the methane is also related to “slip” and planned emissions. The use of inefficient consumption includes all the mechanisms that allow methane to be released into the atmosphere in NYC.

2. Lines 24-25: It does not seem accurate to refer to a city as an individual source. I suggest rewording or removing this part of the sentence.

   “The New York City Metropolitan Area (NYCMA) has been identified as one of the world's top 100 largest persistent area sources of methane and is the largest area source of methane in the Northeastern United States (U.S.).”

3. Line 89: Please specify the frequency when you say “high-frequency” – although it is specified elsewhere in the paper, I suggest this to avoid confusion as this section is the primary description of that dataset and some readers may not consider 1Hz to be high-frequency.

   “We used ~~high-frequency~~ 1 Hz atmospheric composition data from two intensive measurement campaigns to inform our analysis:”

4. Lines 94-96: I am unsure what the authors mean by “discrete” 15- and 5-minute windows. Are the measurements not continuous? If not, how frequent are these discrete sampling periods?

   “We sampled from 60 m AGL in ~~discrete~~ 15-minute block windows (e.g., noon to 1 pm 1 Hz sampling time at 60 m AGL would be: 12:00–12:15 and 12:30–12:45 ~~etc~~) from May 19th to July 21^s^{t}, 2023 and in 5-minute windows from July 21^s^{t} to August 10th, 2023, and in 15-minute windows during February 2024. “

5. Line 167: This table is missing.

   Thank you for catching that. We evaluated bottom-up methane emission estimates available at global, national and regional spatial scales (Table S2).

6. Lines 210-211: Are the authors referring to the results of this study? If so, this sentence should be in the results, not the methods.

   We agree. Text has been moved to later. "There are no active landfills in the five boroughs of New York City, and only two active landfills in northern New Jersey within the 75\% surface influence footprints for the Mineola Tower. These landfill methane emissions are included in the inventories, and together are about 1\% of the total methane emissions estimated for February 2024 \citep{epa_inventory_2024}. ~~Previous studies have pointed out that methane emissions from landfills may be higher in winter \citep{gillespie_estimating_2024} and should be reassessed in the inventories. Our results show the microbial sources are unlikely to dominate methane emissions from the NYCMA.~~"

7. Line 230: Do the authors mean "variables" or "parameters" instead of "configurations"?

   We thank the reviewer for the comments. We use the term *configuration* to refer to the method choices governing the PBL parameterization scheme, not numeric values.

8. Line 238: I suggest replacing "indicating" with "suggesting".

   Agreed.

   "We also found that the simulated diel variability of ΔCH$_4$ closely matches the observed diel pattern, ~~indicating~~ suggesting that the Lagrangian transport model HRRR-STILT captured the overall atmospheric dynamics for the NYCMA (Figure S10)."

9. Line 259: I suggest rewording "Emission Flux" to avoid redundancy.

   We have changed the wording to just emissions and removed the word flux.

10. Line 296: The authors state there is a "distinct" seasonal variation, but the months seem generally within each other's uncertainty bars.

    "We identified ~~distinct~~ seasonal variations in methane emissions from NYCMA from January 2023 to December 2024 (Figure 4), with larger uncertainties during summertime possibly driven by background variation and wetland activities."

11. Lines 315-317: The authors use the slope/correlation method with $\Delta C_2H_6:\Delta CH_4$ to determine whether there is a thermogenic contribution, but I got lost on how this turns into a percent thermogenic contribution. Is that just the fraction of the data points that have a thermogenic contribution? Please clarify.

We thank the reviewer for the comments. There are two steps to calculate the thermogenic fractions: step one applied the correlation, and step 2 applied the slopes:

1. We used the 5-minute ΔC2H6:ΔCH4 correlation to identify the natural gas (thermogenic) plumes (purple ones in Figure S5).
2. We used the natural gas (thermogenic) plumes to calculate the natural gas ΔC2H6:ΔCH4 ratio, and compared the calculated natural gas ΔC2H6:ΔCH4 ratio to all observed ΔC2H6:ΔCH4 ratios to quantify the thermogenic attribution each month:

$$Thermogenic\ fraction\ (\%) = \frac{Observed\ \Delta C2H6{:}\Delta CH4\ ratio\ (black\ slope)}{natural\ gas\ \Delta C2H6{:}\Delta CH4\ ratio\ (purple\ slope)} \times 100\%$$

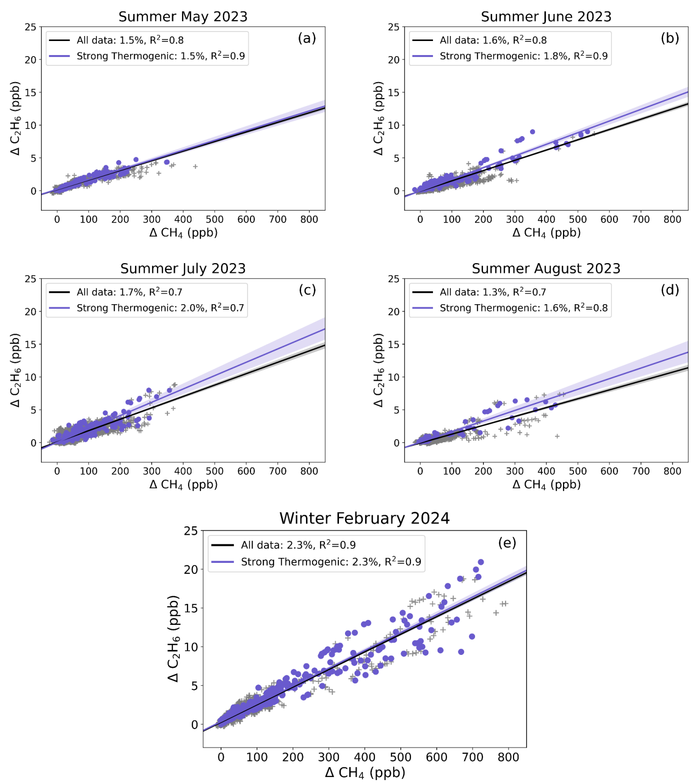

Figure S5. Scatter plot of 5-minute average ΔC2H6 vs ΔCH4 for (a) May, (b) June, (c) July, (d) August, 2023, and (e) February, 2024. In each plot, all observed 5-minute data are plotted as gray crosses and the RMA linear regression is represented as black lines with 95% CI shown in gray shading. Thermogenic plumes with R2ΔC2H6:ΔCH4 ≥ 0.95 are shown in purple circles and linear regression of thermogenic ΔC2H6:ΔCH4 are shown in purple lines, with purple shading representing 95% CI.

12. Line 330: I suggest specifying the “larger fraction.” The figures are informative, but it’s hard to tell what fraction quantitatively because all the crosses and dots are on top of one another.

    We thank the reviewer for the comments. We have included a further description in our response to other comments (see major comment 2). We also included the histogram of $R^2_{\Delta CO:\Delta CH_4}$ in Figures S12 and Secstion S12 in the manuscript.

13. Figure 5: Why do the colors change between the top two and bottom three pie charts, and why is there a gap between them? If no reason, I suggest removing the gap and keeping the color scheme consistent.

    We thank the reviewer for the comments. Here we use different color regimes and the gap to distinguish the heating and cooling seasons.