Reviewer 1

This manuscript discusses oxygenated volatile organic compound emissions and sources in the Berkeley/Oakland area of California. The authors tied together both flux tower measurements and mobile monitoring in this study to investigate emissions and sources of ethanol, methanol, acetone, and acetaldehyde with significant depth. They assessed the role of cooking in contributing to ethanol and acetaldehyde emissions, and the contributions of VCPs for acetone and methanol. They concluded that the total oxygenated volatile organic compound contribution to total measured VOCs was 70% of molar flux, 69% of measured mixing ratios, and 27% of calculated OH reactivity, with ethanol standing out as having the highest mixing ratio and flux of the focus chemicals. They also performed some comparisons of their flux measurements with the GRA2PES inventory and showed good agreement for ethanol sources but not with emission factor estimates for ethanol. Emission factor estimates for acetaldehyde and methanol similarly had poor agreement with GRA2PES, with the inventory generally overestimating emissions. This was a strong paper and is a good fit for the journal. I have a few minor comments and suggestions, which are listed in detail below.

> We thank the reviewer for the compliments and thoughtful suggestions.

Methods: Minor clarifying questions.

Line 104-105: When you say roughly the same roads were sampled for each domain, do you mean roughly the same roads daily, or across some other unit of time?

> Roughly the same roads daily. Updated text is below.
>
> "Roughly, the same roads were sampled on each day for each domain, and measurements were randomized in terms of street sampling order and sampling time.

Line 218: It would help readers to provide slightly more detail about how the flux footprints were computed rather than relying too much on the cited literature, and also to potentially move this discussion closer to the occurrence of Figure 1 where the flux footprint first appears.

> Thank you for the suggestion.
>
> "Flux footprints, which are the geographic area impacting tower measurements, were calculated using the methodology described by Kljun et al., (2015). Briefly, the footprint depends on the tower height, wind direction, wind speed, turbulence, stability, and surface roughness length, which was determined in 30° wind quadrants using Lidar scans. These parameters are described in more detail in the preceding publication (Katz et al., 2025). Flux footprint modeling, especially in urban environments, can be uncertain, so the intention with this analysis was to show the average geographic area of influence instead of to pinpoint exact emission sources on short time scales."

Line 133: Can you clarify how many times per day was the grid cell sampled (as you mention that you report the median of daily means)?

> "Grid cells were visited one or two distinct times per day, and there were multiple data points per visit since data was collected at 1 Hz."

Line 139-140: Can you provide some additional justification, validation, or references for this definition of background concentration used?

> Thank you for requesting clarification. The background-subtracted concentrations were used solely to identify transient spikes in concentration above the local background and were not used for any quantitative analyses in the manuscript. Therefore, the choice in percentile may impact the visualization of individual spikes but does not impact other results in the manuscript.
>
> Our approach follows the methodology of Apte et al. (2017) (https://doi.org/10.1021/acs.est.7b00891) who estimated the local background using the 5th percentile concentration within a rolling 120 second window. We also used a rolling window to estimate the local background but used the 10th percentile based on inspection of the resulting background-subtracted time series. We found that the 10th percentile provided a stable representation of the local background while producing spike identification consistent with the 5th-percentile approach. Changing the percentile which defined the background produced negligible differences in the identified spikes. We clarified this methodology in Sect 2.1.3 and noted that background subtraction was only used for spike identification.
>
> "Spikes in VOC concentration were identified using background-subtracted data. The background concentration for each VOC was designated as the 10th percentile in 10-minute rolling windows. Background subtracted data were only used for spike identification and were not used to present quantitative metrics throughout the manuscript. We considered designating the background as the 5th percentile, similar to Apte et al., (2017), but the differences were negligible."

Generally: How were isobaric interferences accounted for in the PTR-MS measurements (e.g., for acetone measurements, did you make any assumptions about the contribution of propanal?)?

> We thank the reviewer for asking this important question. We could not quantify the contribution from propanal and assumed the signal at $C_3H_7O^+$ was predominantly from acetone. We updated the text in Sect. 2.2.2.
>
> "The influence of isobaric interferences for the compounds of interest could not be quantified. Fluxes of $C_3H_7O^+$ were assumed to be predominantly from acetone, though some influence from propanal was possible. Reports of propanal fluxes in urban areas are sparse. In Beijing, the emission ratio of propanal (relative to acetylene) was ~4 % of the emission ratio of acetone, while background concentrations of propanal were ~16 % of acetone concentrations (Yuan et al., 2012). Together, these comparisons indicate direct emissions of propanal in urban areas are likely minimal relative to acetone, while regional transport or photochemical production could influence the observed $C_3H_7O^+$ mixing ratios by up to 15 to 20 %."

Results and Discussion:

Line 234: Any speculation as to why there was no significant correlation between OVOC fluxes and temperature? Can you show this explicitly in a figure (perhaps in the SI)?

We think that the main driver of the OVOC fluxes is human activity. Episodic activities (e.g., cooking, VCP use) cause spikes in the flux. There is probably an underlying dependence on temperature and other meteorological factors, but the main driver of the variability we observe is activities. Correlation plots are below for each OVOC of interest. We added this to the SI. Each datapoint represents a one-day average (only daytime data is included). Updated text below.

"We looked for but did not observe a significant correlation between OVOC fluxes and temperature (Fig. SX), which may indicate that the trends in OVOC fluxes are predominantly driven by human activity."

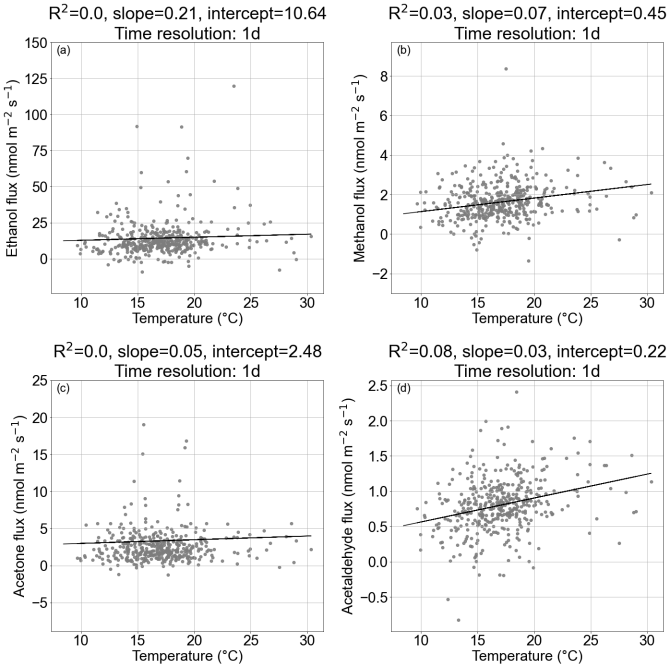
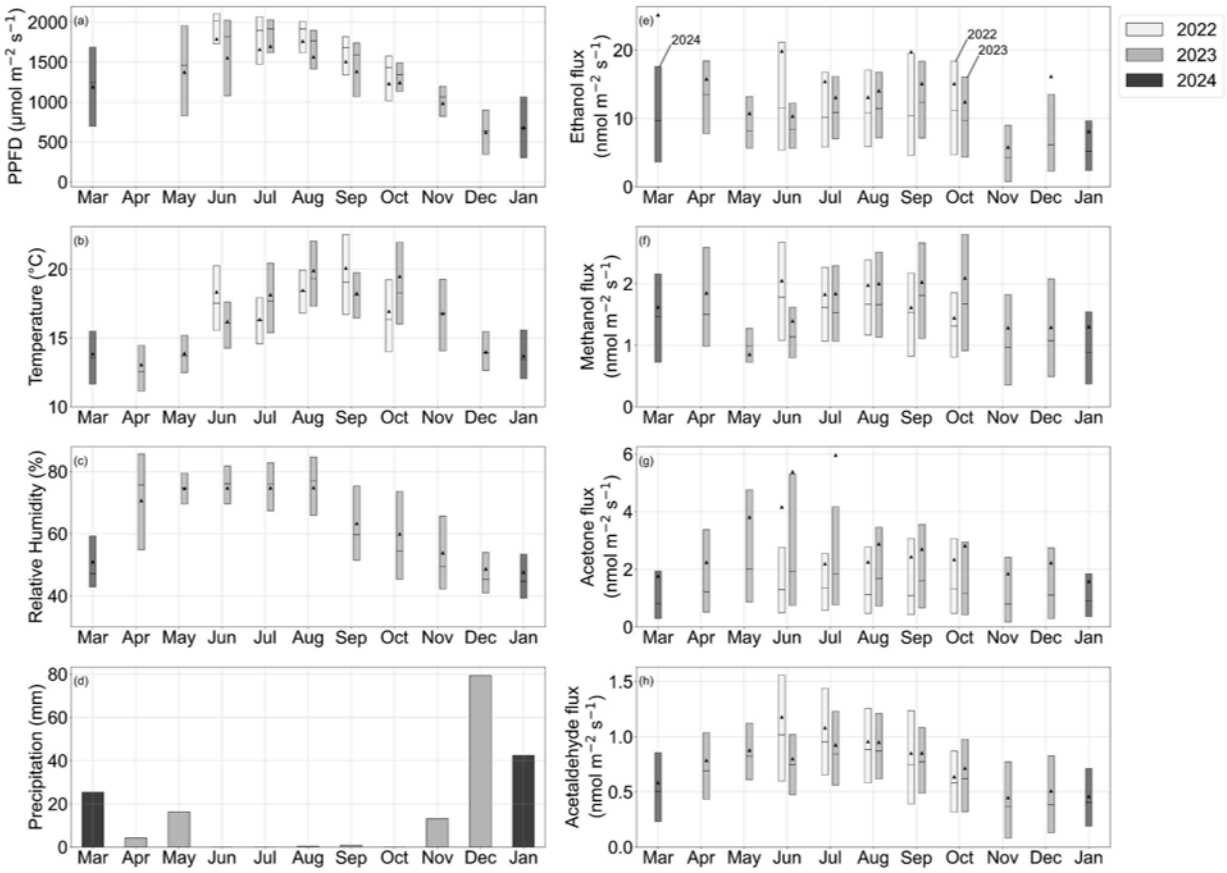


Figure 2: It would help to note the year on top of each bar for readability.

Thank you for the suggestion. This figure is already quite busy, so we decided not to add text to each bar. Instead, we added the annotation to one panel. (See comment below)

Figure 3: I wonder if this could be merged with Figure 4 so that the reader can directly compare the chemical fluxes in Figure 4 with the environmental parameters shown in Figure 3. This would also address my comment above with respect to line 234 (of showing temperature and fluxes explicitly together).

Thank you for the suggestion. The merged figure is shown below. We also adjusted PPFD from the original figure to only show the data during peak sunlight hours (10:00 to 14:00).



For the literature cited for ethanol to nonanal ratios in lines 219 to 323, two of the papers measure indoor mixing ratios and one measures outdoor. While each of these observations is valuable in constraining the possible range of ratios, what caveats need to be considered when interpreting the indoor values when considering outdoor measurements? I.e., other confounding sources of these chemicals either indoors or outdoors?

> "Indoors, other sources of nonanal could be present such as surface ozone interactions and skin oil oxidation (Kruza et al., 2017), though we expect their influence to be minor in these studies because the ratios were taken while cooking was actively occurring."

Generally: can the authors address, in a paragraph, any impacts of the flux tower and mobile measurements not being temporally aligned, and also any considerations regarding the only 12 days of mobile measurements included (and thus some commentary on how representative these are)? If possible, I wonder if some kind of figure similar to Figure S6 with the range of mobile measurements overlaid on the flux tower data would help establish the relationship between the two data sets (in terms of their relative mixing ratio ranges and relative levels between observed chemicals).

> We thank the reviewer for the suggestion. We commented on the representativeness of the mobile dataset in Section 2.1.1 and 2.1.2. We added a paragraph to Section 3.1.2 which discusses the range of mixing ratios observed in the two datasets and added the suggested figure to the SI.

"Mobile and tower measurements were not contemporaneous; mobile data collection occurred before the tower was established. Comparing the two datasets assumes that the major OVOC sources sampled were roughly the same between years. Figure S14 compares the range of mixing ratios observed with the mobile laboratory to the flux tower for overlapping months (November, December, January). Generally, there was good agreement between the two datasets for November and December, while tower data from January was lower. On average, mobile laboratory data was 1.1 to 3.6 times higher than tower observations. Since mobile measurements were closer to VOC sources and the sampled airmasses had undergone less dilution, we expected mobile laboratory mixing ratios would be higher than those measured at the flux tower 40 meters above the ground."

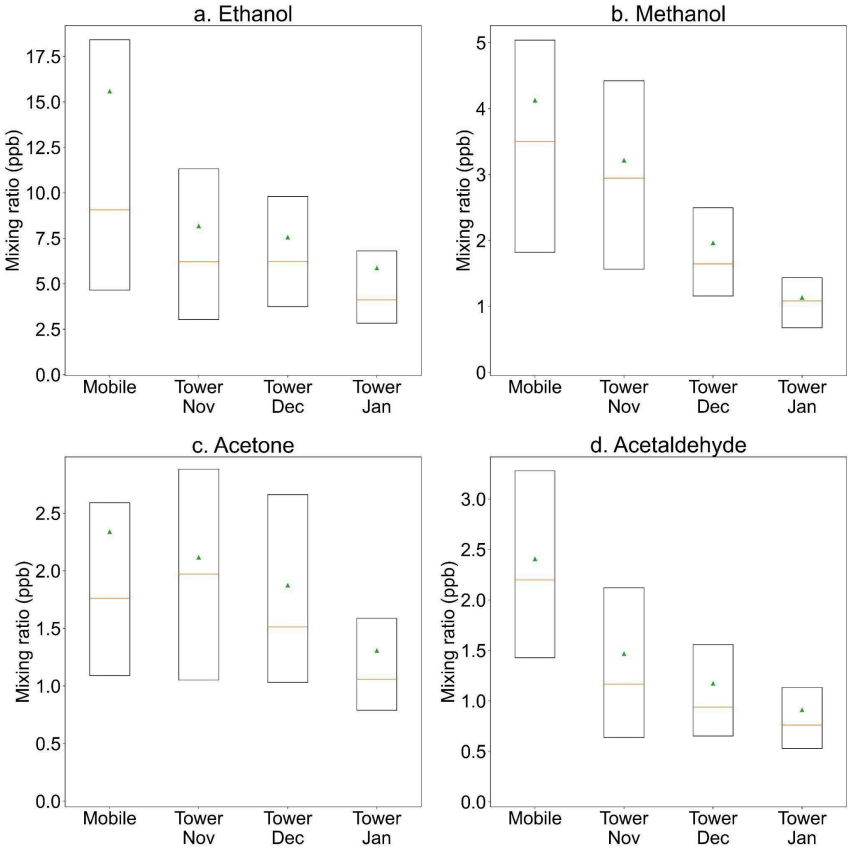
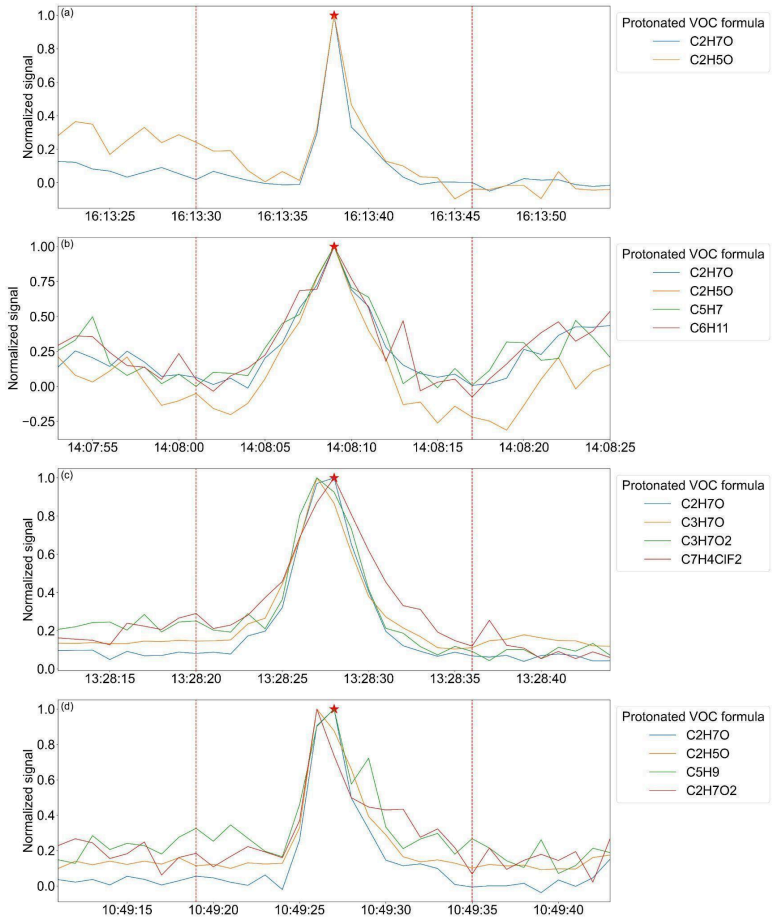


Figure 9: Can the authors provide a few more example time series to show spike correlations? Over what timescale is the co-occurrence quantified (i.e., do the peaks need to co-occur over the course of a few seconds like the SI example shows (Fig. S4)? Few minutes? And does the length of time remain consistent across spikes defined as "correlated"?)?

Co-occurrences were found by calculating VOC correlations within a ±8 second window around the spike peak (Sect. 2.1.3). The same length window was used for all spikes. So, peaks needed to co-occur over the course of a few seconds. This short window was chosen after visually examining concentration spikes in the dataset. We assumed these short spikes corresponded to

driving by a point source or group of point sources. We expanded Figure S4 to display a few more examples of spike correlations.



Line 679-680: This is a really important and big point that feels glossed over at the end (about the inventory/measurement comparison being more accurate for lower population density grid cells and potential associated issues with population density scaling). Can the authors elaborate slightly more here? This sounds like an excellent future work direction.

We thank the reviewer for this suggestion. The second reviewer suggested that this claim needed to be softened, since we did not make measurements in lower population grid cells. We softened the claim and added this sentence to the conclusion section.

"Future work evaluating emissions inventories should consider multiple regions with diverse population and restaurant densities to understand how population scaling impacts the comparisons."

Reviewer 2

General Comments

Katz et al. combine a very unique long-term eddy covariance (EC) VOC flux dataset with mobile observations to investigate the sources and spatial variability of ubiquitous OVOCs in Berkeley and Oakland. This is a valuable dataset, and the combination of tower and mobile measurements provides rare and much needed insights into urban VOC emissions and the representativeness of the tower footprint for those that are locally derived with the EC method. The large contribution of ethanol to the measured VOC flux is consistent with observations from other urban areas, and the flux-based source apportionment provides an important complement to the many published concentration-based apportionment studies.

My main concern is that some of the quantitative source apportionment results are presented with more confidence than the current analysis supports. In particular, many estimates rely on inventory- or observation-derived scaling factors and source ratios without fully evaluating the associated uncertainty. The comparison with GRA2PES is also useful, but some of the apparent agreement is not fully independent, and several conclusions about the causes of model-measurement differences seem stronger than the available evidence. Overall, I think the manuscript is suitable for publication after the authors provide additional sensitivity and uncertainty analyses and soften several of the source-specific and inventory-comparison conclusions. I hope the comments below are helpful in strengthening the analysis and interpretation.

> We thank the reviewer and appreciate these thoughtful suggestions, which we agree will strengthen the manuscript.

Specific Comments

Lines 98: Some more experimental methods could be helpful here. Can the authors please provide some additional information about the mobile platform and inlet, including the inlet length and location relative to the vehicle, whether the reported tubing diameter is the inner diameter, and the flow rate through the tubing.

> We thank the reviewer for the suggestion. We added the following details so Sect. 2.1.1.
>
> "The Vocus sampled ambient air through filtered (47 mm PTFE membrane filters with 2-micron hole size) 0.32 cm inner diameter Teflon tubing at ~2 liters per minute. The inlet tubing was mounted approximately half a meter above the van roof with the inlet facing away from the exhaust. Inlet tubing was ~3.8 meters long."

Lines 117-118: Some experimental details should also be included here. How were zeros produced (cylinder, heat platinum catalyst, etc.), and were calibrations performed using serial dilutions with multiple concentration steps?

The text says the calibration factors and zeros were stable, but it would be helpful to provide a quantitative value for what that stability threshold was such as typical variability or maximum deviation within a sampling day.

Thank you for the suggestions. We have revised the text to address these comments in Sect 2.1.2.

"Calibrations were performed approximately every 2-4 hours using two gravimetrically prepared cylinders (Apel Riemer). Zero air was generated using a platinum catalyst and was used to dilute air from the cylinders in multipoint calibrations (3-4 points). Calibration factors and zeros were stable (within 10 %) throughout the measurements each day, so one set of calibration factors and instrument zeros were used for each day."

Lines 120-121: How was the default kPTR value determined, and what compounds were used to derive it? Please also provide an uncertainty for the default sensitivity, like a range of potential values based on observed or representative species.

Thank you for the comment. The sensitivity of well-behaved (i.e., those with minimal to no fragmentation) VOCs in the PTR-MS is linearly related to the proton transfer rate constant (kPTR). Holzinger et al., (2019) (https://doi.org/10.5194/amt-12-6193-2019) report a plausible kPTR range of 1.85 to 3.39 cm^3 s^{-1} $molecule^{-1}$. Cappellin et al., (2012) (https://doi.org/10.1021/es203985t) calculated kPTR for an extensive list (n=180) of compounds under variable operating conditions. Depending on the operating conditions, the median kPTR ranges from 2.2 to 2.5 cm^3 s^{-1} $molecule^{-1}$, and we believe 2.25 cm^3 s^{-1} $molecule^{-1}$ is the most representative choice for our instrument. Similar approaches using a representative kPTR to estimate sensitivities of uncalibrated VOCs have been applied in PTR-MS measurements with uncertainties of ~ 50 % (e.g., Arata et al. 2021, https://doi.org/10.1111/ina.12906).

We agree that an uncertainty should be reported with the default sensitivity. Since the actual kPTR of an uncalibrated VOC can differ from this representative value, we are considering the possible range from Holzinger et al. (2019). We added the following text to Sect. 2.1.2.

"On average, the default sensitivity was 5500 (4600-8400) counts per second (cps) per ppb. The uncertainty range is based on the possible kPTR values for ambient VOCs, 1.85 to 3.39 cm^3 s^{-1} $molecule^{-1}$, as reported by Holzinger et al. (2019)."

Lines 133-134: How was the 200 x 200 m grid cell size selected, and was the sensitivity of the spatial patterns to this choice evaluated? In other words does this resolution average over important sub-grid variability? The grid size may not strongly affect comparisons using the overall climatological footprint, but it could matter if individual averaging period footprints are considered in this analysis.

On a related note, reporting the 50% relative standard deviation in daily means is good but I also think that there should be a more quantitative presentation of the relationship between high RSD and VOC emission hotspots.

We thank the reviewer for the question and interesting ideas. We're focusing on similar ideas (mobile laboratory methodology for pinpointing direct VOC sources in urban areas) in a separate manuscript. We chose 200 x 200 m grid cells because they are roughly the size of a city block and that level of resolution worked well for comparing the mobile laboratory data to the flux tower data and contextualizing the flux data within the broader urban region. Assessing VOC variability on smaller spatial scales is beyond the scope of this current manuscript, and we did not consider

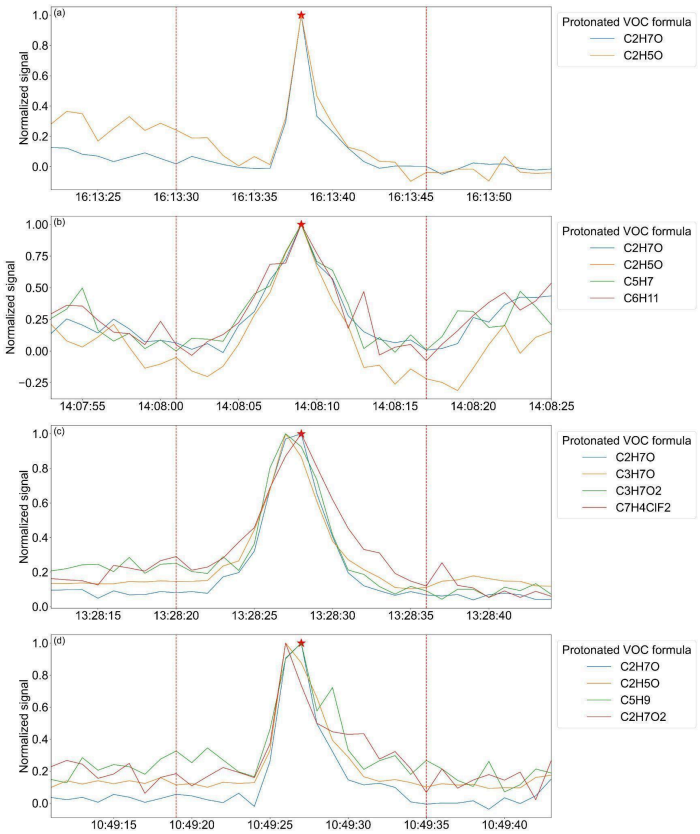
footprints on short time scales, and added the following sentence to indicate why we focused on the average footprint. “Flux footprint modeling, especially in urban environments, can be uncertain, so the intention with this analysis was to show the average geographic area of influence instead of to pinpoint exact emission sources on short time scales.”

Lines 142-144: The spike-correlation analysis is interesting, but correlated signals within a short time window may represent transport of a mixed source plume rather than necessarily direct co-emission, so I suggest using “correlated” or “coincident” spikes unless co-emission can be more directly established. Is there any example figure like Fig. S4 where there are some spikes for a group of compounds with a common source and no spike for a group or VOC from a different source?

Please also clarify whether ions known to be fragments of the same parent compound were excluded from the correlation analysis, since shared fragmentation could otherwise produce high correlations that do not represent separate co-emitted compounds. It seems like correlation analyses were only used for calibrated species known not to be fragments?

We changed the wording to refer to spikes as ‘co-occurrences’ or ‘coincident.’ We also added a few additional examples of spike correlations to Fig. S4 (see below). Known fragments and water clusters were excluded from the spike analysis (e.g., acetone water cluster and small OVOC fragments). Modified text in Section 2.1.3 is below.

“For a given VOC, spike co-occurrences were found by calculating VOC correlations within an 8 second window around each spike. VOCs with correlations (R2) above 0.85 in this window were considered coincident. Figure S4 demonstrates how VOC correlations were found around each aggregated spike data point. Known fragments and water clusters were omitted from the correlation analysis.”



Lines 152-153: While it is mentioned that additional information is available in the Katz et al. (2025) reference, some experimental details relevant to the EC measurements should be mentioned here. For example, the inlet length, inner diameter of the tubing, flow rate, and approximate inlet residence time should be provided here. Further, what was the calculated high frequency attenuation (Horst, 1997; DOI: 10.1023/A:1000229130034) and random uncertainty (Langford et al., 2015; https://doi.org/10.5194/amt-8-4197-2015) for select VOCs?

Thank you for the suggestion; we added more experimental details to Section 2.2.1.

"Instruments were housed in a small room on the roof, and ~36 meters of heated, filtered, Teflon inlet tubing (0.32 cm inner diameter) ran from the top of the tower to the instruments. The tubing flow rate was set to 10 liters per minute, which resulted in residence times under 3 seconds."

We calculated the flux noise root mean square error (RMSE), which is the RMSE of the covariance spectrum (w'c') at unreasonably high lag times. This is one of the outputs of InnFLUX (Striednig et al. 2020, https://doi.org/10.5194/amt-13-1447-2020) and is described as the most conservative estimate of flux LOD. To obtain the RMSE values listed below, we divided the flux RMSE by the measured flux for each 30-minute averaging interval. Median and interquartile range (IQR) values of the resulting random error are Ethanol: 16 % (IQR: 10-26 %), Methanol: 15 % (IQR: 11-24 %), Acetone: 14 % (IQR: 9-23 %), Acetaldehyde: 14 % (IQR: 9-22 %). We added the following text to Sect. 2.2.3.

"The flux noise root mean square error (RMSE), the most conservative estimate of flux noise, for the four OVOCs of interest was between 14 and 16 % of the measured flux (median value) (Langford et al., 2015; Striednig et al., 2020)."

To calculate high frequency attenuation, we followed the approach outlined by Striednig et al., (2020) in the InnFLUX manuscript (https://doi.org/10.5194/amt-13-1447-2020, see SI section S4). Briefly, in urban environments, the sources of sensible heat flux are distributed differently than VOC sources, which can cause dissimilar co-spectra. We have a sufficiently long dataset to filter co-spectra using strict criteria, and the geometry of our sampling set-up is defined, which allows for transfer functions of the main loss processes to be calculated. We created a model w'C' co-spectrum using strictly filtered toluene flux data and applied the transfer functions outlined in Section S4 of Striednig et al., (2020). The key transfer functions we used were tubing attenuation (based on inlet geometry) and sensor separation. When applied to our model cospectra, the resulting integral was less than 5 % different than the original, indicating most of the flux was attained. Similar results were described by Striednig et al., (2020).

Line 168: The manuscript assigns 50% uncertainty to compounds using the default sensitivity, but Katz et al. (2025) states that this value was assumed. How was the 50% value determined?

Holzinger et al., (2019) state that uncalibrated compounds can be quantified with 30 % (and up to 50 %) uncertainty. We have chosen the more conservative uncertainty estimate based on the range of possible kPTR values.

Line 171: How large was the isoprene fragmentation correction, either avg. magnitude or % correction during day and night?

The isoprene correction resulted in a 68% reduction in mixing ratio during the day and a 83% reduction at night. The correction resulted in a 64% reduction in flux during the day and a 89% reduction at night. This is also reported in our previous manuscript. We added the following text to Sect. 2.2.2.

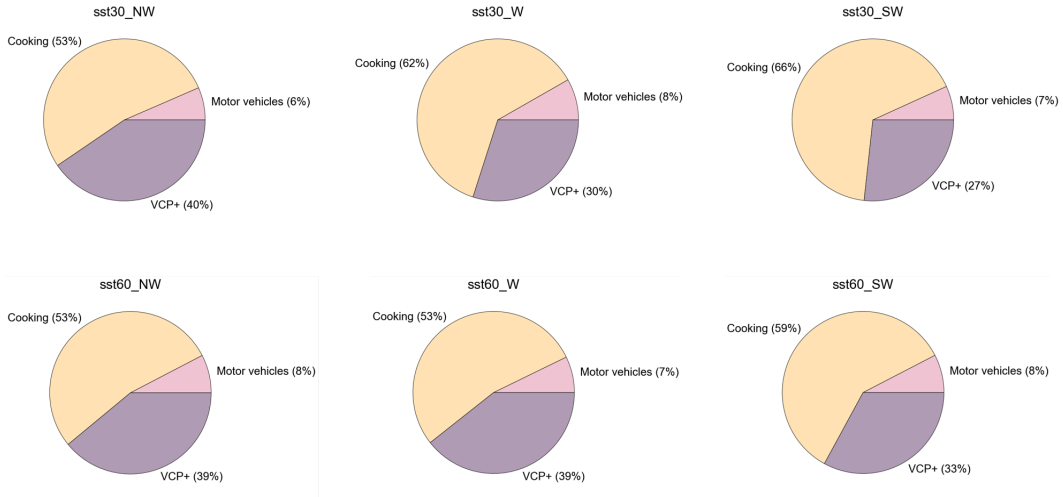
"On average, the isoprene correction resulted in a 68% reduction in mixing ratio during the day and a 83% reduction at night. For fluxes, the correction resulted in a 64% reduction during the day and a 89% reduction at night."

Also, was interference of ethanol fragments in the acetaldehyde signal quantified? The Coggon et al. (2024b) reference goes through this correction.

We looked for but did not observe the impact of ethanol fragmentation on acetaldehyde in our dataset.

Lines 189-190: Why was a 60% stationarity threshold selected rather than generally recommended 30%? If the covariance changes substantially within an averaging period, the calculated flux may combine different turbulent, footprint, or source conditions and may not represent a single steady-state period. This seems particularly relevant in a heterogeneous urban footprint with intermittent emissions. Are the main source-apportionment results sensitive to a stricter threshold?

The stationarity test resulted in removal of a large fraction of the dataset. The typical 30% stationarity threshold was relaxed to 60% to allow for more data to be included in the analysis. We relaxed this threshold following methods published by other urban eddy covariance studies, e.g., Velasco et al., 2009 (https://doi.org/10.5194/acp-9-7325-2009) and Acton et al., 2020 (https://doi.org/10.5194/acp-20-15101-2020). The main source apportionment results were not sensitive to the stricter threshold. Below are results from 2022 in different wind sectors with different stationarity tests applied (sst30 indicates the stationarity test had to be below 30%, and sst60 indicates below 60%).



Lines 192-193: What were the typical lag times and did this align with what was expected for a calculated inlet residence time? Were these calculated lag times still robust in the evening and early morning when covariances are typically low and approach the noise?

Lag times agreed with the inlet residence time (~3 s). The lag time analysis in Innflux allows for lag times determined during periods of high flux to be interpolated so they can be applied during times with noisy covariance spectra. Updated text is below.

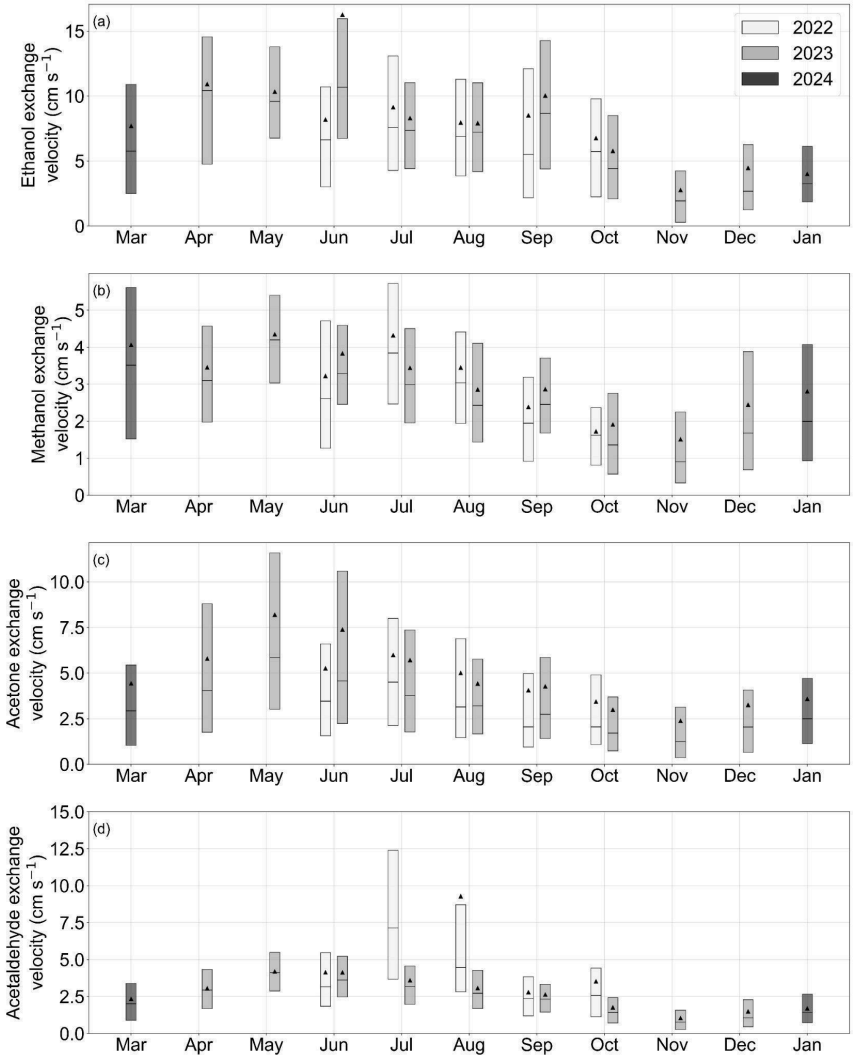
"Lag times were determined for each VOC separately using the maximum covariance between vertical windspeed and mixing ratio for averaging intervals with high signal-to-noise ratios and interpolated across the dataset following a procedure in the InnFLUX code (described by Striednig et al., (2020))."

Line 266: The authors suggest that lower winter ethanol fluxes may partly result from enhanced wet deposition. Do the authors mean enhanced dry deposition and/or uptake to wet surfaces rather than wet deposition? Could this proposed mechanism be evaluated using exchange velocity, dew point departure, surface wetness, or another available metric?

We thank the reviewer for the suggestion. We calculated exchange velocity for the OVOCs and found that it was generally lower in October through January. This aligns with time periods with

the highest mixing ratios (and lowest fluxes). During those months, the source strength is weaker relative to the atmospheric concentration, implying local emissions are lower. This contrast also implies that the concentrations during those months are decoupled from processes in the flux footprint, and are most likely impacted by processes like boundary layer height and atmospheric transport. We removed the suggestion that lower winter ethanol fluxes may partly result from enhanced wet deposition.

"Lower fluxes during these months were most likely caused by lower local emissions. The exchange velocity (Vex), the ratio of flux to concentration, was lowest in November through January (Figure S10), while mixing ratios were highest during these months. This implies that the source strength in the flux footprint was relatively weaker, and that high mixing ratios were largely decoupled from processes within the flux footprint."

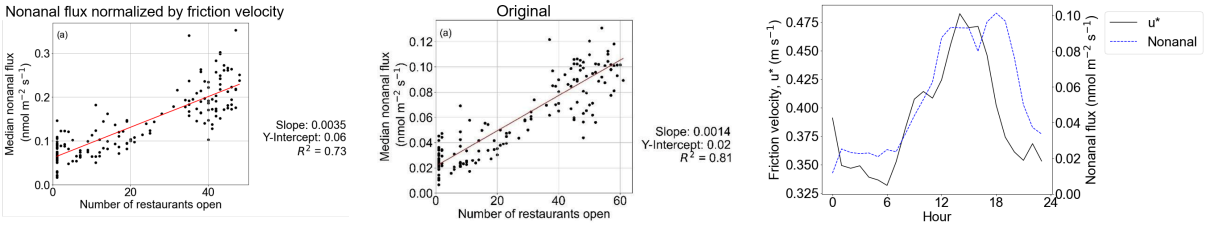


Lines 300-305: The relationship between nonanal flux and the number of open restaurants provides useful supporting evidence, but both quantities have strong diel patterns. I think the diel profile needs to be removed or accounted for before this correlation is interpreted. For example, the authors could compare anomalies at the same hour on different days or normalize the flux using sensible heat, friction velocity, or another measure of turbulent transport. Further, was the number of open restaurants determined dynamically within each half-hour footprint or within the overall climatological footprint? Those might provide different results and interpretations.

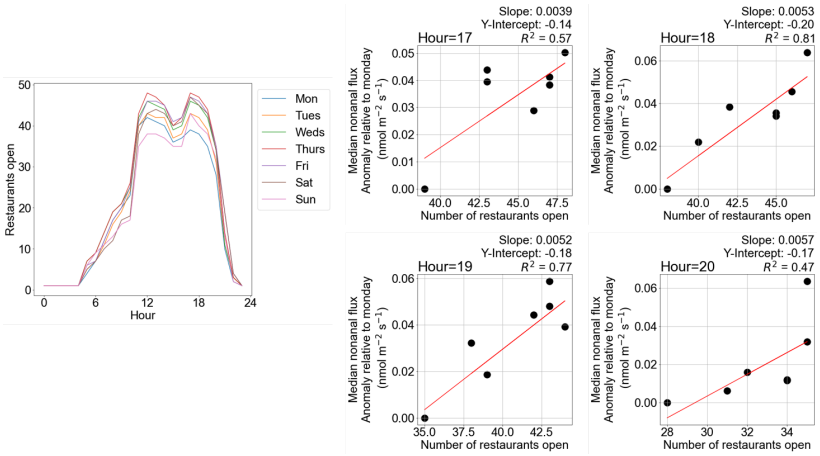
We thank the reviewer for the thoughtful comment. The goal of this plot is to show that there is a relationship between the number of open restaurants and nonanal flux, thus providing more evidence that nonanal is a robust tracer for cooking emissions in our flux footprint. The quantitative results of the fit are not used to determine any of the results reported in the manuscript, so a different fitting approach would not change the results and interpretations. We believe the normalization tests below show that the correlation between nonanal and restaurants is robust and can remain in its current format in the manuscript.

The number of open restaurants was determined for the overall average climatological footprint. We felt this provided a more representative estimate than half-hour averages given that flux footprint modeling can be uncertain, especially in urban areas. Further, most of the results in the manuscript consider averages or medians over longer periods of time which includes the entire footprint. Further, this approach also ensured that we were averaging over all of the different cooking styles present in the flux footprint, providing a more generalizable result.

We did not find evidence that the nonanal flux diel profile is only driven by turbulent transport. The median diurnal profiles of friction velocity and nonanal flux are shown below. Notably, the nonanal flux remains elevated after the friction velocity decreases, which corresponds with dinner hours. Also, the nonanal flux diurnal profile directly matches the trend of the number of open restaurants. We also normalized the nonanal flux by the friction velocity and did not see a change in the trend (see below). There was still a robust correlation between the normalized nonanal flux and the number of restaurants open.



We also examined nonanal flux anomalies at the same hour on different days per the reviewers request. We determined the nonanal flux anomaly relative to the Monday flux, which was the day with the lowest number of restaurants open for dinner hours. There is a clear correlation between the nonanal flux anomaly and the number of restaurants open (see plot below, each dot represents a different day of the week).



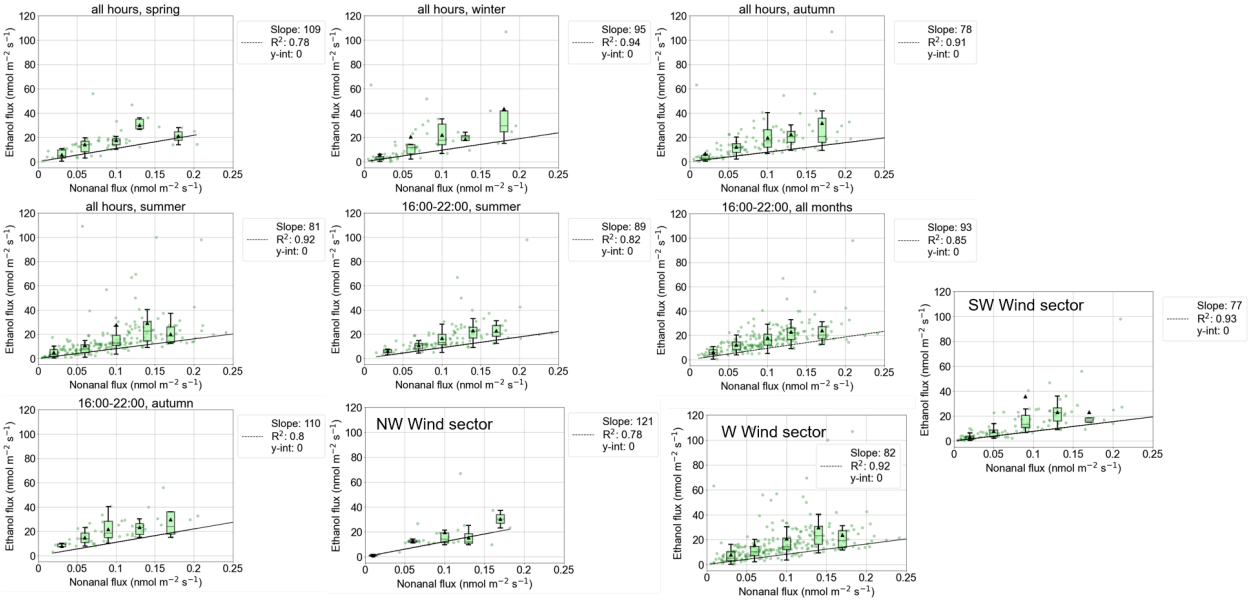
Line 326: Did the authors mean that cooking intervals are when ethanol/benzene are greater than 0.01? I would expect cooking periods to be when ethanol is high, but benzene is low, and even then I would think it's much more than 0.01. Please clarify this threshold.

> Thank you for catching this. We clarified the text. It was a typo and should have been the benzene/ethanol ratio instead of ethanol/benzene.
>
> "...when benzene/ethanol was less than 0.01 to avoid influence from on-road sources."

Line 327-328: The rationale for using the lower portion of the ethanol-nonanal relationship seems reasonable, since additional ethanol sources should only move points above the cooking relationship, and the similar slopes from the 5th to 20th percentiles provide a useful sensitivity test. However, I am still not fully convinced that the lower edge necessarily gives a representative ethanol/nonanal ratio for cooking. Even if the measurement uncertainty, cooking period selection, and binning are all reasonable, couldn't the lower edge still be influenced by differences among cooking sources or by non-cooking sources of nonanal? It would be helpful to show whether the slope is similar across different wind sectors, times of day, seasons, or reasonable cooking selection criteria, or by using quantile regression on the unbinned data. That range could then be used to show how sensitive the reported cooking contribution is to this assumption.

> We expect non-cooking sources of nonanal to be extremely minor in the flux footprint relative to the cooking source. These could include skin oil oxidation, indoor surface oxidation, (https://doi.org/10.1111/ina.12381) and ocean surface film oxidation (https://doi.org/10.5194/acp-24-3729-2024) (the ocean was not in our footprint). By taking the footprint average ethanol vs. nonanal relationship, we include all restaurants and residential cooking and thus provide a representative description of all types of cooking styles and types in the footprint.
>
> We ran a few sensitivity tests according to your suggestions. The original slope in the manuscript was 80, and the binned slope range was 71 to 92. The slope is similar across different wind

sectors (W slope=82, SW slope=77, NW slope=121). However, note that NW has much fewer data points available (see figure below), which reduces points that we can use to define the lower edge, which increases the uncertainty. NW data also therefore includes fewer cooking styles than W and SW. We also performed the fit (binned data fit through 10th percentile) at different times of day and different seasons (see figure below). Under these conditions, the slope ranged from 78 to 110 with an average of 94 and standard deviation of 12. These results show that the fit of the lower edge is stable across different conditions.



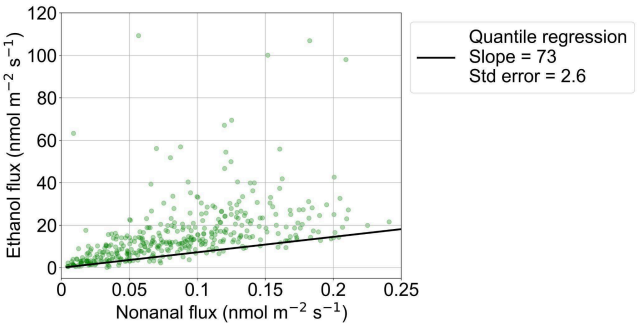
We appreciate the suggestion to use quantile regression on the unbinned data and found this to be a more robust fit. We decided to use this method in the revised manuscript instead of the original approach which involved using binned data. We performed quantile regression on the whole dataset, fitting the 5th percentile through the 20th percentile. With this fitting method, the slopes ranged from 64 to 90 with an average of 77 and standard deviation of 9. Ultimately, we believe that fitting the 10th percentile leads to the most accurate fit of the lower edge of the slope of ethanol versus nonanal, and that including all data (not limiting to wind sectors or certain times of day or season) allows for the most accurate definition of the lower edge and includes the greatest diversity of cooking styles.

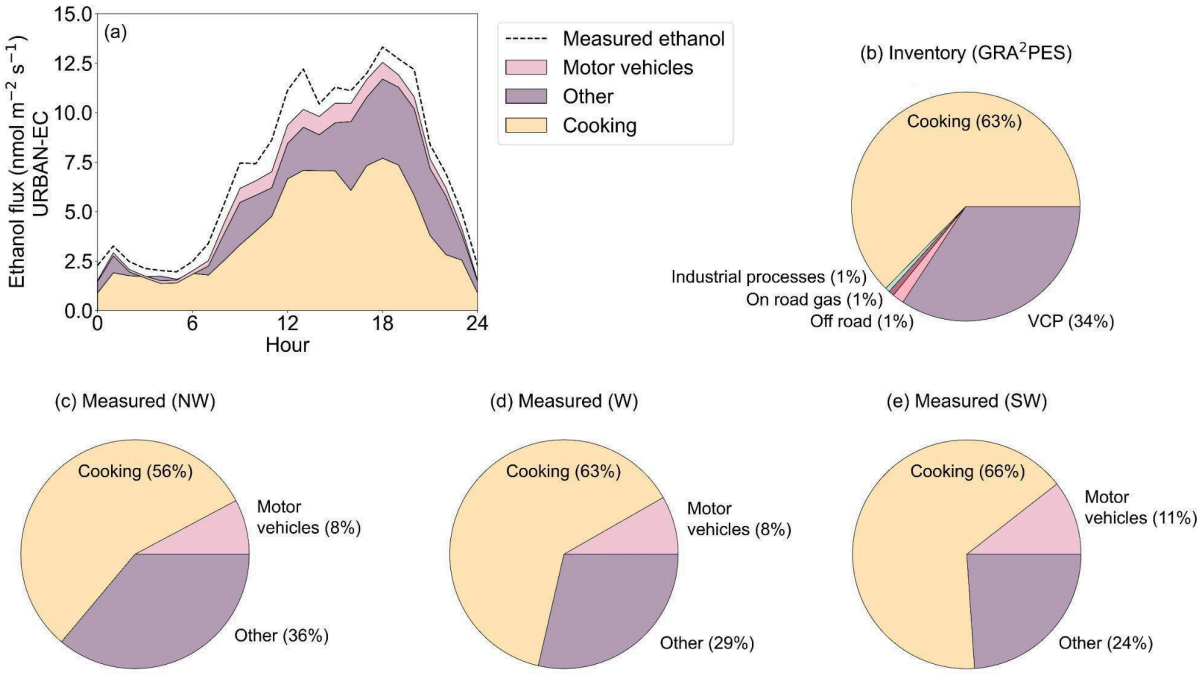
We estimated the uncertainty in the ethanol/nonanal ratio from cooking using bootstrapping. We sampled the dataset with replacement 10,000 times. For each iteration, we fit the ethanol vs. nonanal data using the quantile regression approach (through the 10th percentile). The median slope was 76 (95% confidence interval: 70 to 82). This result is slightly lower than the ratio we reported in the first draft, but it includes a more robust uncertainty estimate and it is closer to the estimate (ethanol/nonanal = 75) from Arata et al., (2021). The updated text, SI figure, and main results figure are below.

“The nonanal to ethanol ratios reported in the literature vary over two orders of magnitude. Ditto et al., (2023), who measured VOC emissions from cooking and cleaning in a restaurant, reported

ethanol to nonanal ratios of 479 to 588. The ethanol to nonanal ratio for a cooking profile derived using positive matrix factorization of ambient VOC data was 11 (Coggon et al., 2024a). Also, Arata et al., (2021) measured indoor air at a test house during cooking experiments and reported an ethanol to nonanal ratio of 75. Indoors, other sources of nonanal could be present, such as surface ozone interactions and skin oil oxidation (Kruza et al., 2017), though we expect their influence to be minor in these studies because the ratios were taken during active cooking. We estimated S_(EtOH_cooking) by examining the observed relationship between ethanol and nonanal when cooking emissions were sampled. Averaging intervals were assumed to be primarily influenced by cooking when nonanal/octanal was within 20 % of the observed nonanal/octanal ratio during dinner time (17:00-22:00) and when benzene/ethanol was less than 0.01 to avoid influences from on-road sources. The scatter plot of ethanol versus nonanal during these times had a clear lower edge with positive ethanol scatter (Fig. S13). The lower edge without scatter indicates that nonanal is co-emitted with ethanol. Non-cooking sources of nonanal (e.g., surface ozone interactions) were expected to be negligible relative to cooking and likely did not influence the lower edge. S_(EtOH_cooking) was the slope of this lower edge and was found by fitting the 10th percentile of ethanol flux using the quantile regression approach (median S_(EtOH_cooking) = 76, 95 % confidence interval = 70 to 82). The median value and the 95 % confidence interval (CI) were determined by bootstrapping the data. We sampled the data with replacement 10,000 times and calculated the slope of the lower edge of the ethanol versus nonanal scatter plot using quantile regression. Figure S13 shows quantile regression results using the entire dataset.

We tested the sensitivity of S_(EtOH_cooking) to a variety of factors. First, we performed the quantile regression using the 5th to 20th percentiles, which resulted in slopes ranging from 64 to 90 (average = 77, standard deviation = 9). Lower percentiles (20th and below) were chosen to capture the lower edge and because ethanol fluxes were influenced by sources in addition to cooking. We also considered binned linear fits through the 10th percentile of ethanol flux for certain subsets of the data, including selected wind sectors, dinner hours, and seasons. These conditions resulted in slopes ranging from 77 to 121. However, some subsets of the data (e.g., NW wind sector) had fewer data points available, which reduced the points available to define the lower edge, increasing the uncertainty. Considering each iteration, the average slope was 83 and the standard deviation was 12. The ethanol/nonanal ratio we determined through quantile regression and bootstrapping (S_(EtOH_cooking) = 76, 95 % CI = 70 to 82) is consistent with observations from cooking experiments at a test house (ethanol/nonanal = 75, Arata et al., 2021), the GRA^{2}PES inventory (ethanol/nonanal = 110), and average of possible ethanol/nonanal ratios from the additional cases tested from our dataset (ethanol/nonanal = 83). Ethanol/nonanal likely varies with food and cooking methods, so this process was chosen to represent all cooking emissions within this flux footprint. With this scaler-based source apportionment approach, 56 % of the median daytime ethanol flux was attributed to cooking."





Line 349: The VCP+ category is calculated as the residual after subtracting cooking and vehicle emissions. As described in the manuscript, it may include alcohol consumption, biogenic emissions, unaccounted sources, and errors in the other estimates. I think there should be some presentation of an independent tracer showing that VCPs dominate this term or it should be called "other" or "residual" to avoid being misleading. The statement that these sources are difficult to quantify also seems somewhat inconsistent with presenting this residual as a quantitative VCP contribution later on in the text.

> We thank the reviewer for the suggestion. We changed the category to be "other" instead of "VCP+".

Lines 368-375: The vehicle contribution is estimated using an ethanol/benzene ratio from GRA2PES, and the resulting source apportionment is then compared with the GRA2PES source fractions. The agreement in source fractions is therefore not fully independent, at least for the vehicle contribution, and I think this should be acknowledged.

> We thank the reviewer for acknowledging this. The following sentence was added.
>
> "We note that since F_(EtOH,mobile) is derived using the measured benzene flux and the ethanol/benzene ratio from GRA2PES, the comparison is not entirely independent."

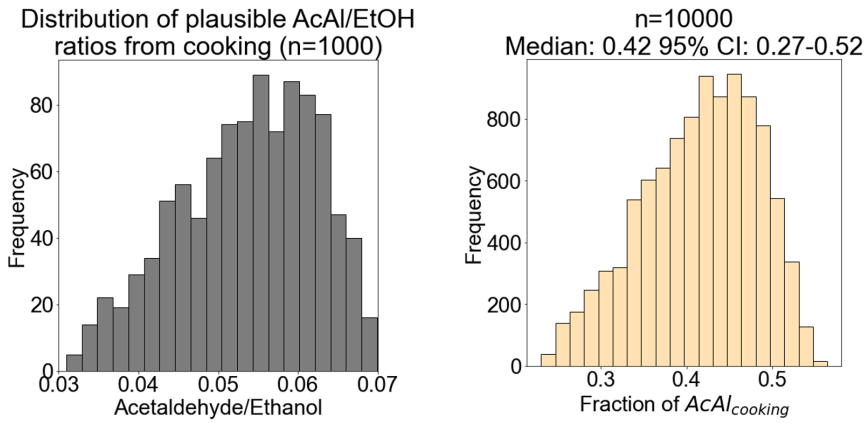
Lines 384-387: The GRA2PES ethanol emission in the lower population northern grid cell is closer to the tower observations, but I do not think this comparison alone demonstrates that the population-density

scaler is overestimated. Even if population scaling is the main reason the inventory differs between these two cells, the tower does not measure the northern cell, so the comparison only shows that a lower modeled emission magnitude agrees better with the observations. I think the language should be softened unless there is other supporting evidence.

> We thank the reviewer for the suggestion, we modified the text.
>
> "This grid cell comparison indicates that population density scalers used to compute ethanol emissions in this region may be currently overestimated. Direct comparisons in neighboring grid cells with variable population densities would confirm this assumption."

Lines 581-585: The acetaldehyde cooking estimate uses an acetaldehyde/ethanol ratio of 0.06, while the literature values span 0.03-1.33. Can the authors provide more context for this range? For example, how are the reported values distributed, and do they vary systematically with the food or cooking method? The finding that ratios above 0.1 produce more cooking acetaldehyde than the total measured flux provides a useful upper constraint, but it does not by itself establish that 0.06 is the most appropriate value. Since the reported 57% contribution depends directly on this ratio, please show how the estimated cooking fraction changes across a reasonable range of ratios and reflect this uncertainty in the reported contribution.

> We thank the reviewer for the suggestion and acknowledge that this uncertainty analysis is necessary. First, we tested acetaldehyde/ethanol ratios ($R_{AcAl/EtOH}$) between 0.03 and 0.1 (lower bound from the literature and upper bound for what makes sense physically with our dataset). Ratios of 0.08 and above resulted in some seasonal total acetaldehyde medians above the total measured flux, so we narrowed the range to 0.03 to 0.07. We performed a bootstrapping + Monte Carlo analysis to estimate the uncertainty in the estimated cooking fraction. Updated text from Sect. 3.4.1 is below.
>
> "To evaluate the uncertainty in this estimate, we performed a bootstrap analysis where observations were sampled with replacement 10,000 times. For each sample, we performed Monte Carlo sampling of $R_{AcAl/EtOH}$ by randomly varying $R_{AcAl/EtOH}$ within its plausible range (Figure S19). The plausible values were selected from a triangular distribution with a mode of 0.06 between 0.03 (lowest ratio from the literature) and 0.07. The range of for cooking in the literature spans three orders of magnitude (0.03 to 1.33) (Arata et al., 2021; Coggon et al., 2024a; Kumar et al., 2025). However, using above 0.08 in the tracer-based source apportionment approach caused the total calculated acetaldehyde flux to be higher than the measured acetaldehyde flux, indicating a lower ratio is more representative of the integrated cooking emissions sampled during this study. Through this method, 42 % (median value, 95 % CI = 27 – 52 %) of the acetaldehyde flux was attributed to cooking."



Lines 588-589: The combustion contribution also depends on inventory-derived acetaldehyde/benzene ratios. Please provide a reasonable uncertainty range for these ratios and show how it affects the reported 23% combustion contribution.

> Unfortunately, the uncertainty is not provided in the GRA^{2}PES inventory.

Lines 593-603: The correlation between the residual acetaldehyde flux and PPFD is interesting, and I appreciate the calculation and consideration of below-sensor gas-phase chemistry, but I do not think it is sufficient to identify photochemical degradation of surface materials as the source. Why was PPFD used rather than a surface UV measurement that is more directly related to the proposed process? Could thermal degradation also contribute to this process, which would be consistent with cooler surfaces in the winter? Is there evidence that ambient aerosols or urban surfaces can produce acetaldehyde under atmospherically relevant light and temperature conditions through heterogeneous chemistry?

> Our tower setup only included a PAR sensor, so we are using that as a proxy for overall UV exposure. We acknowledge that other sources are possible, but it is also speculative. We updated the text.
>
> "Instead, photochemical or thermal degradation of materials at the surface or heterogeneous aerosol chemistry could be sources of acetaldehyde."

Fig. 11: The interannual variability in both the cooking fraction and the magnitude of the acetaldehyde flux is notable. Can the authors provide some potential causes these differences? Are they related to differences in restaurant activity, wind direction and footprint, the estimated ethanol/nonanal ratio, or meteorology? This variability should be discussed because it affects how confidently the campaign-average source fractions can be interpreted.

> The average wind direction and footprint were not significantly different between years. The most likely cause of the difference is changes in restaurant activity between 2022 and 2023. However, we do not have the data to prove this. We will acknowledge this difference in the text and speculate the change is from restaurants.

"The fraction of acetaldehyde attributed to cooking was higher in 2023 than 2022, which may be caused by higher restaurant emissions, as meteorological drivers were not significantly different."

Minor Comments/Technical Corrections

Line 186: This is a really minor comment but the text says meteorology inputs were 20 Hz but the sonic data was actually downsampled to 10 Hz before calculating the covariance, correct?

The InnFLUX code upsamples the 10 Hz data.. Reported that detail in Sect. 2.2.3

"The inputs included 20-Hz sonic anemometer data (3D wind, temperature, pressure, etc.) and upsampled 10-Hz calibrated Vocus data."

Line 191: There is an extra "all" in "All data is all on Pacific Daylight Time."

Corrected.

Line 302: Please correct "they was highest" to "were highest".

Corrected.

Line 355: F(EtOH,cooking) should be subtracted, not added if this is a residual.

We thank the reviewer for catching this. Here is the corrected formula:

F_EtOH,VCP+ = F_EtOH - (F_EtOH,FF + F_EtOH,cooking)

Line 417: I know what the authors mean but it may be more straightforward to say a 200 x 200 m grid rather than "200 square meter grid" since it's technically not 200 m2.

Corrected.

Lines 421-422: The sentence stating that the tower could be "missing key sources of ethanol" could be made more clear. I think the intended point is that the footprint does not include some important regional hotspots, rather than that the local flux measurement is somehow incomplete.

Thanks, the corrected sentence is below.

"Importantly, the highest average ethanol concentrations observed via mobile monitoring were not within the flux footprint, indicating the flux observations could have missed key regional hotspots of ethanol emission (described below in Sect. 3.1.3)."

Reviewer 3

General comments:

Katz et al uses long-standing flux measurements at a site in Berkely California and couples them with intermittent mobile laboratory measurements throughout the broader region to identify OVOCs (ethanol, methanol, acetaldehyde) that contributed heavily to the molar flux and OH reactivity flux. The combined methodologies cover broad spatial and temporal scales and evaluates the results with current emissions inventories. This is a well-written manuscript, though, I have a few questions regarding the source attributions, assumptions, uncertainties, and spatial scale mismatches with the emissions inventories that should be addressed before publications. Overall, this is an impressive dataset and analysis, though limited in scope by the footprint representativeness.

> We thank the reviewer for the thoughtful comments.

Specific comments:

Cooking ethanol:nonanal ratio: The derived cooking-specific ethanol:nonanal ratio was determined as 80 using a 10th percentile "lower edge" daytime flux fit. While this matches specific indoor stir-fry experiments (Arata et al., 2021: 75), it is more than double the regional, observationally constrained ratios of 37 utilized in recent Los Angeles chemical transport models (Zhu et al., 2025), which are based on LA airborne fluxes and integrated HOMEChem profiles (Arata et al., 2021: 40.2). It is also much higher than the values determined from PMF source apportionment that the GRAPES inventory was initially built upon (Coggon et al., 2024;11). These differences can really drive the conclusions of this work, significantly modifying the daytime ethanol flux attributed to cooking and the subsequent residual determinations of VCPs.

> We thank the reviewer for pointing this out. While we were preparing this manuscript, we also noticed a large range of ethanol/nonanal ratios throughout the literature. We believe that the approach outlined in the manuscript and updated per the second reviewer's suggestions (see above) provides the most robust and representative ratio for this specific study domain. Other ratios may be more suitable for those specific studies. The ratio from Arata et al. also includes the Thanksgiving day experiment, which included many different cooking styles.

→Can you explain how these discrepant spatial scales reconcile (local high-density commercial tower footprint of 2.8 km² vs. integrated regional airborne flights).

> In this manuscript, we did not directly compare our quantitative flux results to flight campaign results. However, comparing the spatial scales should not be an issue because fluxes are normalized to the surface area, i.e., the unit is nmol m^{-2} s^{-1} or similar (e.g., mg km^{-2} h^{-1}). Since our flux footprint is more focused than flight campaigns, comparing e.g., campaign medians, should consider relevant metrics including population density, temperature, etc. Future analyses that involve such comparisons should consider comparing similar geographic areas. For instance, a

flight campaign that covers a dense commercial and residential city center could compare those specific geographical results to our results.

→Can you demonstrate that the 10th percentile fit isolated pure cooking rather than co-located, temperature-independent VCP solvent or mobile emissions, maybe color the scatter plot in Figure S9 by the benzene:ethanol ratio.

Thank you for the suggestion. We assess the lower edge of the ethanol/nonanal scatter plot in detail in our response to Reviewer 2.

→Can you provide a quantitative assessment showing how varying this ratio impacts the absolute molar flux distribution of ethanol.

We thank the reviewer for the question, but varying the ethanol/nonanal ratio would not impact the absolute molar flux we determined for ethanol.

→Can you confirm which specific ratio the GRA2PES inventory applies (i.e., is it PMF-based or scaled up using HOMEChem and flux measurements?).

The ratio we report from GRA2PES (ethanol/nonanal = 110) is directly from the inventory that we downloaded from NOAA (https://csl.noaa.gov/groups/csl4/gra2pes/datasets/). We apologize, but it is unclear how they go from the values reported by Coggon et al. (PMF-based) or Arata et al. (HOMEChem) to the value that is in the publicly available inventory.

Mobile and VCP tracers: It appears as though the total benzene flux is used as a mobile source scalar to establish an upper bound on vehicular ethanol emissions. However, the mobile data demonstrate that benzene co-emits with PCBTF and acetone, pointing to significant solvent/VCP sources of benzene in this urban footprint.

We thank the reviewer for the comment. We would like to clarify that the manuscript text does not state that benzene co-emits with PCBTF and acetone, the direct quote is “Acetone and PCBTF spike correlations were not observed with BTEX spikes, indicating at least two distinct solvent-based acetone sources were sampled: one that includes PCBTF and one that includes BTEX.” We made an adjustment to the language to ensure that correlations do not equal co-emission. It’s possible that solvent-related sources of acetone and PCBTF (e.g., construction) occur near BTEX sources (e.g., motor vehicles). Co-located sources could result in adjacent spike correlations in the maps of Figure 8.

To our knowledge, and following discussions with Brian McDonald, a key developer of the GRA2PES and FIVE-VCP inventories, benzene is not an ingredient in solvents in California. We suspect that the benzene mass used in the mobile data (C6H7) could be influenced by fragmentation of substituted aromatics. We added that note in Sect. 2.1.3.

> "Also, C6H7+, the formula for protonated benzene used in the mobile dataset, may be influenced by fragmentation of substituted aromatic VOCs."
>
> We also added a clarifying note to Sect. 3.3.2 to further address the question.
>
> "Acetone and PCBTF spike correlations were not observed with BTEX spikes, indicating at least two distinct solvent-based acetone sources were sampled: one that includes PCBTF and one that includes BTEX. Benzene is not an ingredient in solvents in California, so acetone correlations with BTEX could indicate that VCP sources were near on-road sources or that C6H7+, the ion used for benzene detection in the mobile dataset, was influenced by fragmentation of substituted aromatics."

→Can you discuss how much urban solvent emissions inflate the local benzene flux and quantify how this might lead to a systematic overestimation of the 9% upper-bound on-road ethanol allocation.

> We do not suspect that urban solvent emissions inflate the local benzene flux. Benzene is not an ingredient in solvents in California and the GRA2PES inventory does not include any benzene from VCPs.

→How much should we "trust" the GRAPES ethanol to benzene ratio for on-road sources and are there other studies that justify using this ratio for the mobile sector

> Previous studies (e.g., Pfannerstill et al. 2023 https://doi.org/10.1021/acs.est.3c03162) including this one have shown that VOC emissions inventories like FIVE-VCP and GRA2PES are better at estimating motor vehicle emissions than non-traditional sources of VOCs like VCPs due to decades of previous work in that area. Therefore, we feel confident in the GRA2PES estimates of motor vehicle emissions.

Residual lumping to VCP+: VCPs are treated as a simple residual ("other") without explicit tracers, acting as a grab bag for everything from indoor-to-outdoor exchange to solvent evaporation. Because any uncertainty in the cooking ratio (80) or the vehicle fuel ratio (7.1) directly propagates to this residual…

→are there sector-specific tracers within VCPs that could further elucidate these sources?

> Unfortunately, because VCPs are such a broad source category, there is not a specific tracer available. Following a suggestion from Reviewer 2, we decided to change the VCP+ category to be called "other" since we cannot isolate VCP sources using this dataset.

→Can mobile sampling data be leveraged to isolate specific ethanol plumes and associate them or their ratios with co-emitted industrial, construction, or cleaning tracers? Essentially using mobile measurements to "constrain" mobile sources, specific VCP sources, etc. Similar to how fingerprints can be used to constrain PMF.

We thank the reviewer for the thoughtful suggestion. We're focusing on similar ideas (mobile laboratory methodology for pinpointing direct VOC sources in urban areas) in a separate manuscript. Pinpointing precise sources and plume origins via the mobile laboratory is beyond the scope of this work. Instead, we chose to focus on source apportionment with the flux data and use the mobile laboratory to supplement our understanding of the broader urban area.

Biogenic emissions: It is assumed that biogenic emissions of ethanol, methanol, and acetaldehyde are negligible and can be ignored. However, biogenic inventories (MEGAN and BEIS) can systematically underestimate urban biogenic VOCs. Specifically urban biogenic emissions have been shown to be underestimated by MEGAN, do you have any localized tree coverage estimates within your footprint or can you justify whether assuming a strict zero contribution from biogenics is physically warranted. Can biogenics be integrated into the tracer methods for ethanol or methanol in any way? How does this vary in the surrounding regions?

In our previous manuscript, we estimated that the vegetation cover fraction was around 30 %. The predominant source of biogenic ethanol is decomposition, which we speculated was minor in the footprint relative to the other sources. We appreciate the reviewer's suggestion but performing a detailed evaluation of MEGAN and biogenic ethanol and methanol in urban environments is the subject of focused manuscripts and beyond the scope of the current study. We do not have data to directly answer the question of how this varies in surrounding regions.

Spatial scales and inventory evaluation: The flux footprint (2.8 km²) is compared to a much coarser inventory grid resolution (3 grids at 4×4 km). Can you justify the selection of 3 grids and were they renormalized to the footprint's area? Because the inventory distributes cooking emissions based on population density, *please discuss how population density compares to actual restaurant density in the area*. Can the authors downscale the inventory using restaurant density metrics to match the footprint, or scale up the flux measurements to match the grid?

We thank the reviewer for this comment. The three grid cells were chosen because they either contained the flux footprint or were the only adjacent grid cells with a large fraction of urban development (other adjacent grid cells contained a large fraction of water or forest). They were not renormalized to the footprint area. Surface fluxes could be compared directly to the inventory. In the manuscript, we acknowledge that the flux footprint area is small relative to the grid cell and that the measurements are representative of higher density urban residential and commercial zones. We also added a sentence in Sect. 3.1.1 to comment on spatial scales and inventory evaluation.

"This grid cell and flux footprint are unique in that they have a high restaurant density, high population density, and large influx of commuters and students during the week. The high population density of the flux tower could have contributed to the discrepancy with GRA2PES."

We appreciate the suggestion but think that adjusting the inventory is beyond the scope of this study. Our goal was to focus on the source apportionment results from our measurements and

assess these results against the existing inventory. Future studies which focus on inventory development may choose to adjust the inventory following your suggestion.

Methanol:ethanol: Methanol flux correlates strongly with ethanol flux, which is attributed to VCPs and indoor-to-outdoor exchange. While this holds at a monthly scale, does this correlation persist at the highfrequency 30-minute or diurnal scale? Showing this high-frequency correlation is necessary to prove whether these compounds share immediate emission sources or if the correlation is merely a seasonal artifact of temperature and boundary layer dynamics.

We thank the reviewer for this suggestion. We examined the relationship between methanol and ethanol fluxes at 30-min resolution and found that the correlation was weak (R^2=0.05), in contrast to the stronger relationship observed for monthly median fluxes (R^2=0.70). This indicates that the monthly-scale correlation does not reflect strongly coupled short-term emission events. Instead, methanol and ethanol appear to exhibit similar longer-term temporal variability, which may reflect shared seasonal influences. We have revised the text to clarify that the monthly correlation alone does not provide evidence that methanol and ethanol share the same immediate emission sources.

"Methanol monthly median fluxes were correlated with ethanol (R^2 = 0.70), indicating similar long-term variability, however, the high frequency (30-min) correlation was low (R^2 = 0.05), suggesting that this relationship does not arise from coupled short-term emission events."

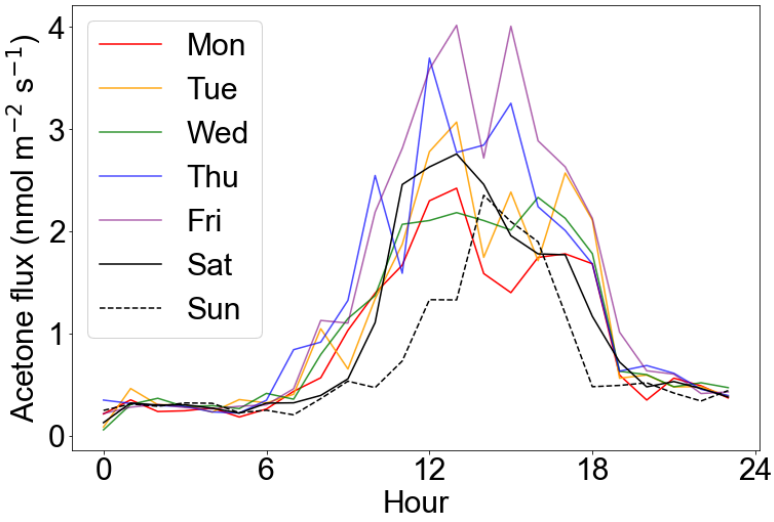
Grouping Saturday with weekdays (Monday–Friday) masks key differences, as commercial construction and industrial solvent use typically drop significantly on Saturdays compared to mid-week. What do you see when you separate Monday–Friday, Saturday, and Sunday in the acetone diurnal flux analysis.

We thank the reviewer for the suggestion. We wanted to clarify that acetone fluxes were significantly lower on Sunday when compared to every other day of the week individually, and we added text in Sect. 3.3.1 to clarify this.

"Day of week variability also suggests construction is a key source of acetone. Daytime acetone fluxes were significantly lower on Sundays than every other day of the week ($p<0.05$, significance was tested for each day individually)."

Below is a figure showing the median diurnal profile of acetone flux for each day of the week. The trend for Sunday is distinct from Saturday, while Saturday has a similar pattern as the weekdays. We added the figure to the SI and the following text to Sect. 3.3.1.

"Figure S18 shows the median diurnal flux of acetone for each day of the week; the trend for Sunday is distinct from all other days."



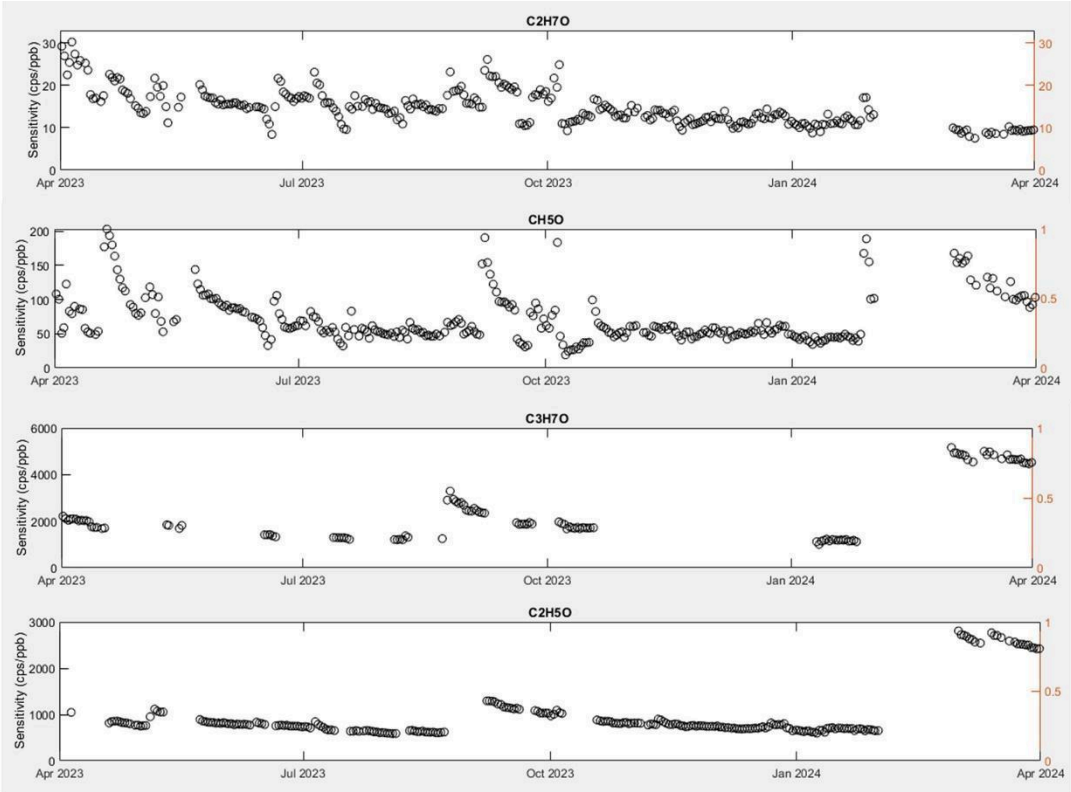
If cooking occurs year-round and restaurant density is constant, why does the daytime ethanol flux decrease by nearly 50% in winter? Is this more driven by temperature-dependent indoor-to-outdoor exchange (e.g., closed windows suppressing ventilation transport) or physical deposition effects. I'd imagine the levels of cooking stay the same through all seasons.

> We speculate that the seasonal difference is caused by changes in ethanol emissions associated with human activity. The change could be caused by lower indoor to outdoor emissions of both cooking and VCPs or lower VCP use (e.g., less construction activities) in the winter. Also, during colder months, the outdoor patios and restaurants in the footprint have fewer patrons due to the weather, so there are less evaporative ethanol emissions. We suspect that physical deposition is not a key driver of the changes, because the exchange velocity is lower in the winter, indicating a weaker source strength. (See Reviewer 2 response and Fig. S10).

Minor Comments/Technical Corrections:

Report the exact instrument operational parameters (specifically E/N and FIMR voltages) and the stability of the calibration factors across different periods in the SI

> We added the following text to Sect. 2.2.2.
>
> "The mass to charge (m/z) range of the Vocus was 10 to 497. The sampling rate was 10 Hz. The FIMR was 100°C and 2.5 mbar. The reagent flow was 15 standard cubic centimeters per second, the electric field gradient was 590 V, and the radio frequency was 450 V. The E/N ratio was ~ 125 Td (1Td = 10−21 V m2). The big, segmented quadrupole (BSQ) voltage was 200 V."
>
> We added the figure below to the SI. The calibration factors were interpolated across the dataset. Jumps in sensitivity correspond to changes in drift tube tuning or instrument maintenance. We alternated between cylinders, which is why there are periodic breaks in the acetone sensitivity.



L118 Monoterpenoids are summed at the parent and major fragment mass, how is fragment overlap from monoterpenoids and monoterpene alcohols (such as eucalyptol, which fragments to m/z 137) accounted for?

In this line, we refer only to the calibrant gases which included monoterpenes (no monoterpenoids). In the ambient data, C10H17+ could be impacted by monoterpenoids and we did not correct for this. We added a sentence to Sect. 2.1.3 to clarify.

"C10H17+, the formula for protonated monoterpenes, could have been impacted by fragmentation of monoterpenoids."

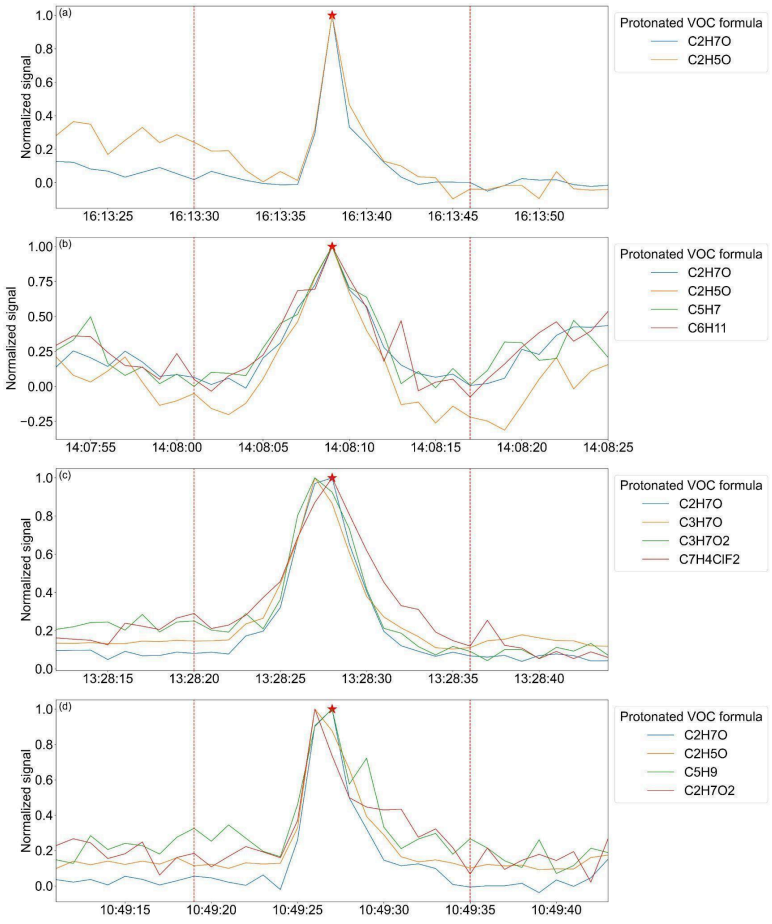
L119 Define what specific criteria are used to classify compounds as "well-behaved."

We define "well-behaved" in Sect. 2.1.2, and we added an additional sentence for clarification (see below).

"Well-behaved calibrants are those that do not fragment or those whose fragmentation products may be summed with confidence, such as monoterpenes (C10H17+ + C6H9+). Well-behaved calibrants have a linear relationship with the proton transfer reaction rate constant (kPTR)."

Can you explain further the physical or mathematical justification for choosing a 8 second window to determine co-emitted VOC spikes.

> After visual examination, this window was selected because it encompassed short-term spikes in concentration that were sampled while driving past a source or group of sources. We modified Figure S4 to include more examples of spikes in concentration (see below).



L166 Please clarify the phrase "default calibration factor scaled by the fraction of parent ion detected." Does this refer to a calculated k-rate calibration factor adjusted for instrument-specific fragmentation?

> Yes, that is correct. The calibration factor applied was the default sensitivity multiplied by the parent ion fraction, which accounts for instrument-specific fragmentation patterns. We added this sentence for clarity.
>
> "Thus, these VOCs were calibrated using the default sensitivity accounting for instrument-specific fragmentation patterns."

Are there any other known interferences at the propanol dehydration product (C3H7+) given its low molecular weight?

Interference from fragmentation on C3H7+ is possible. For this reason, we did not focus on this ion for further analysis. We added this text to Sect. 2.2.2.

"Fragmentation of larger VOCs could influence the signal at C3H7+ in addition to propanol, so it was not the focus of further analysis."

Were any corrections applied to acetaldehyde for ethanol and ethylene glycol interferences?

We did not correct for potential fragmentation from these VOCs.

L275 Tower ethanol fluxes in Berkeley were higher than those in Beijing. Given differences in cooking styles and oil usage, is this physically expected, or is it an artifact of Beijing's measurements representing a much broader regional scale?

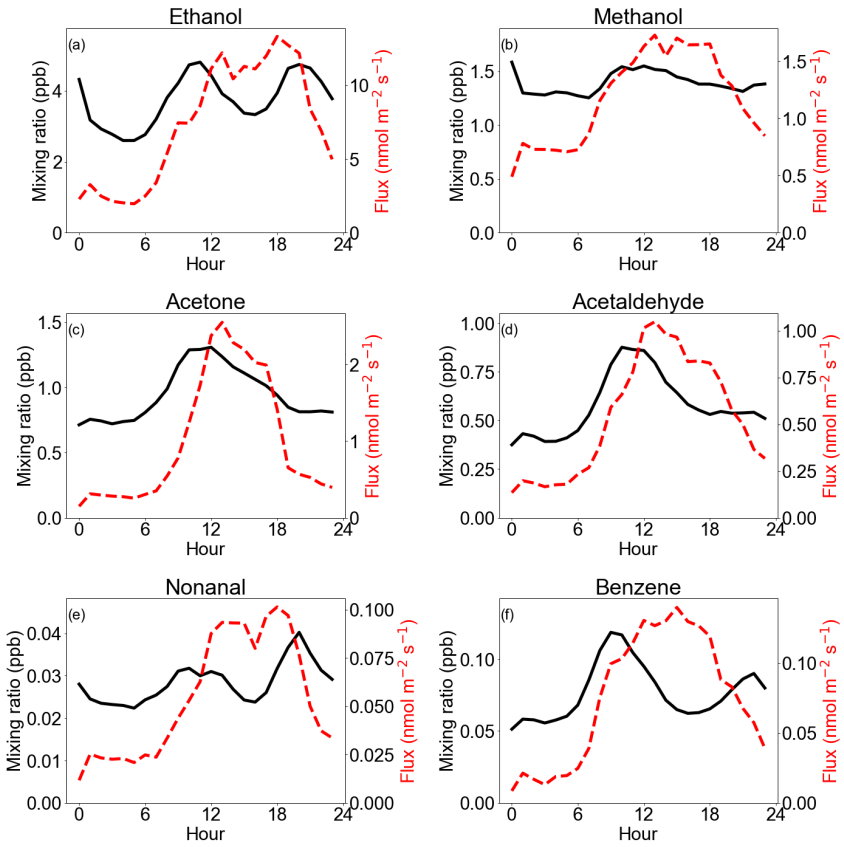
Our flux tower was located next to many restaurants, bars, and breweries, which likely contributed to the high ethanol fluxes we observed. Without knowing more about the Beijing flux footprint, it is challenging to compare the two.

Why was nonanal not quantified in the mobile laboratory?

We found that VOC fragmentation was sensitive to the skimmer voltage gradient in the drift tube (Katz et al., 2025 ACP, https://doi.org/10.5194/acp-25-15281-2025). At higher skimmer gradients, the nonanal fragmentation increased. During the mobile laboratory, our drift tube was not tuned sufficiently to allow for nonanal to be detected.

Show a diurnal profile comparing raw mixing ratios to fluxes to contrast nighttime boundary layer accumulation against active daytime venting (especially for nonanal and ethanol).

We thank the reviewer for the suggestion and added the following figure to the SI.



In Figure 6(a), shouldn't the measured ethanol flux be reached since VCP+ represents the full residual?

> We thank the reviewer for pointing this out. The measured ethanol flux is slightly higher than the sum of each ethanol source. The residual (i.e., VCP+) was calculated for each 30-minute averaging interval after accounting for the other two sources, so slight differences were expected between the median sum of each source and the actual measured median flux. Further, some of the VCP+ (now called "other") averaging intervals were negative, which explains why the sum doesn't exactly match the measured value.

Why not include nonanal spike co-occurrences in Figure 9?

> Unfortunately, nonanal spike co-occurrences could not be included because nonanal was not quantified in the mobile dataset due to instrument parameters.