This document provides our answers to the anonymous referee #1 for "**Methane mapping, emission quantification, and attribution in two European cities; Utrecht, NL and Hamburg, DE**"

We would like to extend our appreciation to the anonymous referee for the insightful and point-by-point comments from careful reading of our paper. In the following sections, comments from the referee are provided in normal black text, our replies are in normal blue, and changes is the manuscript are in *blue italic* format.

1 Overview

This study presents an extensive set of vehicle-based measurements of methane and other compounds to investigate methane emissions in Utrecht, NL and Hamburg, DE. The authors used empirical equation developed von Fischer et al. (2017) and updated by Weller et al. (2019) to estimate the emissions from the natural gas distribution network using the methane enhancements they observed during their surveys. They also tested several approaches to determine the origin of these enhancements (biogenic, thermogenic, pyrogenic) such as the isotopic signature of methane, C_2/C_1 ratio or CH_4/CO_2 ratio. Finally, they used a Gaussian dispersion model to estimate methane emissions from larger sources.
Overall, despite the large ranges and uncertainties presented here, the study is a valuable contribution to the literature as the state of knowledge of urban methane is not very advanced compared to other pollutants. The measurement campaigns seems carefully done and well designed, I appreciated the authors described their interactions with the local distribution companies and showed how their work helped reducing the emissions. It was also nice to find a list of all the acronyms in the supporting information as there are so many of them in the text and I was a bit lost at first. I recommend publishing it after addressing few minor points (which are also mentioned in the detailed comments):

1 General comments

1) Structure: I would reorganize section 2 a bit and group all the source attribution approaches together in one subsection instead of having two subsections about isotopic analysis and information about the ratios scattered throughout the rest of the subsections. It would also make it easier for the reader if there was a table in this section summarizing these approaches and the limits used to attribute the emissions.

- We have followed this suggestion and section, Section 2 is now rearranged to the following sub-sections: 2.1. Data collection and instrumentation, 2.2. Emission quantification, and 2.3. Emission attribution.
- In the "data collection" section we keep sub-sections of mobile measurement of C_2H_6 and CO_2 separated from sampling for isotopic analysis as the analytical techniques for these two attribution methods are very different. Nevertheless, we now combine data evaluations of C_2H_6, CO_2, and isotopic analysis in sub-section 2.3.

2) I would have liked more discussions about the uncertainty on the emissions estimated with the approach developed by Weller et al. (2019). I understand that this method is the reference for mobile surveys at the moment but the fit of the calibration curve presented in figure 4 of Weller et al. (2019) makes me wonder about the uncertainties associated with these estimates. Also, the authors used this empirical equation to estimate emissions from microbial and combustion sources whereas it was originally designed to estimate emissions from NGDN.

While the classification into small, medium and large LIs depending on the maximum amplitude of the enhancement remains correct, I am not very comfortable using the empirical equation to estimate the emissions of these other sources. Biogenic emissions are very different from NGDN emissions, they are way more sensitive to atmospheric conditions (especially temperature) and are likely to vary in time unlike NGDN emissions which should be more constant. They could also potentially be located further away than the usual road side emissions. Figure S16 examples illustrate this: microbial emissions from the water body are likely localized further away and the sewage system seem to emit at a higher level than the road level.

- Yes, the uncertainties of the quantification algorithm introduced by von Fischer et al. (2017) and improved by Weller et al. (2019) for individual LIs are large. It is indeed evident from figure 4 of Weller et al. (2019) that individual LIs can be strongly under- or overestimated. The rationale is that when a complete city is surveyed, the contribution from the different LI categories and the total emission rate can be estimated more precisely. Following the request from the referee, we have explicitly clarified this better in the revised version (see Sect. 2.2.2, L293-302).
- We acknowledge that the algorithm in the original papers was designed to quantify pipeline leaks. Attempts to exclude emissions from other sources were to restrict the spatial extent of a CH_4 plume to <160 m, to require a minimum enhancement of 10% above background, and to require multiple detections. In our study, we add explicit attribution for many LIs by evaluating the co-emitted (or not) tracers C_2H_6 and CO_2. Rather than simply flagging and neglecting these LIs for the quantification of potential pipeline leaks, we use the same algorithm for quantifying emissions from other categories, namely microbial processes and combustion. We thank the referee for pointing out that this needs to be spelled out specifically in the paper. Whereas biogenic emissions, e.g. from the sewage or wastewater systems, will be released through manholes, waste water drains and other cavities that are also important for the release of CH_4 to the atmosphere, combustion related emissions may come from vehicles or houses and thus have different release pathways. The quantification of emissions from these source categories is thus based on the assumption that the same conversion equation can be used. This is especially the case for microbial emissions from manholes, where the enhancements and distance of the emission release point is very similar to NGDN leaks, hence the emission quantification approach applies for both source categories. In the revised version, we have stated this explicitly. We now suggest that the number of detected enhancements that have been attributed to microbial and combustion sources is more reliable than the emission rates.
- We have considered at an early stage of our research to use the information of co-emitted species (especially C_2H_6) to focus the paper only on NGDN leaks but decided against this. We feel the biogenic emissions are part of the overall anthropogenic urban CH_4 emission and, while being more uncertain, it is relevant to have an approximation of the importance in relation to NGDN. We thank the reviewer for reminding us that we need to be extra careful with this category of emissions (and we agree) but still think it is valuable information.

3) The presentation of the GPDM approach used to quantify the emissions from larger facilities should be reworked and expanded a bit. For example, the authors should talk about the wind data they are using as it is a critical parameter in this approach. Did they adjust the wind direction so that the maxima of the observed plume is aligned with the maxima of the modeled plume, etc. The part about the selection of the sigma y and sigma z is also not very clear. The author should also specify here which observation they are fitting with the model

(the measured concentrations? One plume at a time or all the plumes measured during all the surveys?)

- The relevant section has been expanded to provide more information. The reason for using data from the two mentioned towers in Utrecht and Hamburg is that the online data logging setup failed to continuously record all the local wind measurements during the surveys. The distance of the towers to the facilities ranges from 8 to 20 km, and indeed these distances introduce extra uncertainties in emission quantification, mainly related to wind speed (see Sect. 2.1.5). When we compare the data that were recorded on the vehicle with the tower, we derive a difference in wind speed of ± 10 %. After considering the remarks of the reviewer, we increase this to a more conservative error estimate of ± 30 % (see Sect. 2.2.3, L352-359).
- Regarding the wind direction (e.g. the oil wells), for several sources the emission point is relatively certain and confirmed by analysis of Google Earth images. In addition, we passed several sources during different wind conditions and did a "triangulation" based on the observed plumes and wind data (see Sect. 2.2.3, L324-333).
- For explanation on how sigma_y and sigma_z were derived, see the detailed comment below.
- Regarding the last question, we fit each plume individually (since they were often observed during different days at different locations) and average the individual results. This has now been specified (see Sect. 2.2.3, L343-344).

3 Detailed comments

L18-19 and L30: Should be consistent with number notation (whether letters or numbers).

- Done

L22: This should be phrased differently, the largest emission rate in Utrecht is actually coming from the wastewater treatment plant.

- This sentence focuses on the emissions from LIs found in the normal mobile surveys, and doesn't include emissions from larger facilities.

L64: Typo? "high precision" is written twice in a row.

- Done

L76: Typo? A comma is missing after "(Giolo et al., 2012; Helfter et al., 2016)".

- Done

L78-80: This sentence is too vague, most methods quantify emissions methane enhancements! The authors should specify which approach they used and which type of sources they used it for.

- The respective sentence was changed as follows (see Sect. 1, L88-90): *In this study, we quantified LIs emissions using an empirical equation from Weller et al. (2019), which was designed based on controlled release experiments from von Fischer et al. (2017), to quantify ground-level emissions locations in urban area such as leaks from NGDN.*

L83: Typo? Should it be "across the urban areas in these two cities" (rather than "across the urban areas is these two cities")?

- Done

L93: Why specify the time needed to flush the cell for the G2301 but not for the G4302? Could you add a sentence about how the methane enhancements measured by the two instruments compare? This discussion is actually in the SI, the authors could add a sentence to refer it.

- This sentence describes the smoothing effect in the cell of G2301. Information on G4302 is added to the revised manuscript as follows (see Sect. 2.1.1, L110-111): *The flow rate is 2.2 L min^{-1} and the volume of the cell is 35 ml (operated at 600 mb, thus 21 ml STP) so the cell is actually flushed in 0.01 s, which means that mixing is insignificant given the 1 s measurement frequency of the G4302.*

L122: What do the level 2 and 3 roads correspond to?

- Information on the level 2 and 3 are now added to the paragraph and the respective sentence was changed as follows (see Sect. 2.1.2, L151-154): *Level 1 roads are primarily larger roads connecting cities, level 2 roads are the second most important roads and part of a greater network to connect smaller towns, level 3 roads have tertiary importance level and connect smaller settlements and districts.*

L137: The authors should remove "at the following links: Utrecht and Hamburg", the citation "(Maazahalli et al., 2020b)" is enough.

- Done

Section 2.4: I would merge sections 2.4 and 2.6.1 into a source attribution section that details the multiple approached used in this study. I would incorporate in this section a table summarizing the different ratios/isotopic measurements and the ranges used to distinguish between fossil, combustion, microbial and unclassified sources.

- Section two is now rearranged to three main sections: '2.1. Data collection and instrumentation', '2.2 Emission quantification', and '2.3. Emission attribution'. In Sect. 2.1 we present information on the mobile measurement of CO_2 and C_2H_6 and sample collection. In Sect. 2.2, we provide information on emission quantification of LI and larger facilities separately. In Sect. 2.3, we combine the attribution approaches as suggested by the referee.

L168: Did the authors only took samples for isotopic analysis in Hamburg? Why not in Utrecht?

- Due to time and budget limitations, we were only able to take a sufficient number of samples for attribution in Hamburg.

L179: How far are the measurements tower from the studied sites? Wind parameterization are large sources of uncertainty in Gaussian plume dispersion model, especially since wind close to the surface can be very different from the wind measured at 10 meters at these towers.

- The distance of the towers to the facilities ranges from 8 to 20 km. The reason for using the tower data was that wind data were not logged by the 2-D anemometer continuously (instrument failure) and the limited data collected were not sufficient to analyze emissions from these facilities.

L184: "It has been demonstrated that the algorithm adequately estimates the majority of emissions from a city (Weller et al., 2018)." The authors should specify that this method was specially developed to quantify methane emissions from the natural gas distribution network. In this sentence, the authors seem to imply that they could estimate the emissions from any type of sources from a city.

- The respective sentence was changed as follows (see Sect. 2.2.2, L246-L249): *This algorithm was designed to quantify CH_4 emissions from ground-level emission release locations within 5-40 m from the measurement (von Fischer et al., 2017), such as pipeline leaks and has been demonstrated that the algorithm adequately estimates the majority of those emissions from a city (Weller et al., 2018).* As mentioned in the reply to the general comment 2, in this study we use the same algorithm to provide indicative estimates of emission rates of microbial and combustion emissions as well. We note that emission pathways to the atmosphere are partially different for such emissions. Therefore, the emission rates should be seen as indicative, whereas the LI numbers from the different categories are more reliable.

L192-194: How did the authors know about the mole percent of CH_4 and C_2H_6 in the NGDN in Hamburg and Utrecht? Is it based on measurements or did the NG suppliers give them this information?

- This data was indeed provided by the network operators. This is now indicated as "personal communication" in the manuscript (see Sec. 2.3.1, L3612-364) as follows: *During the Utrecht campaign, the overall mole fraction of CH_4 and C_2H_6 in the NGDN was ≈ 80 % and ≈ 3.9 % (STEDIN, personal communication) and in Hamburg the mole fraction of CH_4 and C_2H_6 in the NGDN was about ≈ 95 % and ≈ 3.4 % (GasNetz Hamburg, personal communication) respectively.*

L196: If I understand correctly, this whole part is used to explain how you differentiate car exhaust signals from NG signals. This is not really clear, the authors should introduced it up front to help the reader follow the organization. This could probably also be moved to the source attribution section.

- Done

L204-207: I don't understand why do the authors use different approaches to estimate the CH_4 and CO_2 backgrounds? This should be explained.

- The background determination method for CH_4 from Weller et al. (2019) was used to stay compatible with the quantification algorithm for the urban studies. But this algorithm doesn't include background extraction for CO_2 and here we chose background detection methods commonly used in the literature (see Sect. 2.2.1 and Figure S7 in SI, Sect. S.2.1).

L229: Did the authors really need to convert decimal degrees to Cartesian coordinates in order to cluster enhancements? Doesn't it introduce additional uncertainties than directly estimating the distance between enhancements using decimal degrees?

- The constraint for clustering based on the von Fischer et al. (2017) algorithm is 30 m, thus we need to have the data in metric system. There are many ways to convert the decimal degrees to metric system, we used this way as it gave a very good one-to-one correlation with R^2=1.00 when we compared output of the equation we used for converting to cartesian system to e.g. EPSG:32632 projection. Easy implementation of this equation in the code we wrote to evaluate the data is another advantage of using this method to convert decimal degrees to metric system.

L233: Why did the authors assigned the maximum observed enhancement to the cluster rather than a weighted average just like for the location? Wouldn't that artificially increase the emissions?

- This follows the algorithm from Weller et al. (2019).

L240: The "visited at least twice" criterion in von Fischer et al. (2017) and Weller et al. (2019) was implemented to identify enhancements from the natural gas distribution network that are considered to emit continuously. I would mention that you are using another source attribution method instead.

- The following sentence was added to the paper (see Sect. 2.2.2, L287-288): *Instead, we used explicit source attribution by co-emitted tracers.* This topic was also discussed earlier in (see Sect. 2.3, L189-193): *Due to time and budget restrictions, it was not possible to cover each street at least twice, as done for the US cities. After evaluation of the untargeted first surveys that covered each street at least once, targeted surveys were carried out for verification of observed LIs and for collection of air samples at locations with high CH_4 enhancements. The rationale behind this measurement strategy is that if an enhancement was not recorded during the first survey, it obviously cannot be verified in the second survey.*

Section 2.6.3: I was surprised that the authors did not talk about wind measurements in this section given that this is one of the biggest source of uncertainty of this technique. Maybe they should move part of section 2.5 here.

- We acknowledge that the wind speed is a large source of uncertainty in the GPDM. The section is now revised and more information has been added in Sect. 2.2.3. See our answer to general comment 3.

L252: What do the authors mean by "These data were evaluated using a simple point source GPDM"? What are the authors evaluating?

- We meant that the data collected downwind the larger facilities were analyzed using GPDM. The respective sentence was changed as follows (see Sect. 2.2.3, L306): *We applied a standard point source GPDM (Turner, 1969) to quantify methane emissions from these larger facilities.*

L252: Typo? "()" should be removed.

- Done.

L265-266: The authors should be consistent with the notation: zsource (which is equal to 0 in the text) and h are to the same thing.

- Corrected.

L276-279: This part is not very clear. Do you select sigma y and sigma z separately? Could you end up with a sigma y of a given Pasquill-Gifford stability class and combine it with a sigma z from another stability class?

- We first determined sigma y based on the width of the plume observed during the measurement and the source location. From the distance between the source location and the maximum of the plume location and sigma_y we chose the most suitable Pasquill-Gifford stability class and then we chose the corresponding sigma_z value from the respective Pasquill-Gifford stability tables.

L288: It would be appropriate to at least in a sentence or two explain the isotopic analysis so the reader doesn't need to go back and read these papers (which analyzer, how long were the samples measured...).

Following this comment, we have now added and combined all information related to the isotopic analysis in Sect. 2.3.2 including analyzers, measurement time scales, calibrations, etc.

L314: Typo? “Utrecht and Hamburg correspond to” rather than “Utrecht and Hamburg were correspond to”

- Done

L321: Typo? “Figure 2” looks weird.

- Done

L332: You showed previously that different types of road had very different LI rates per km depending on cities, why didn’t the authors use these road-specific emission factors to upscale their emissions?

- The evaluation showed that different types of road have different LI rates per km in these two cities, which means that the smaller or bigger LIs can happen on different road types. In this study we aim to compare cities based on total emissions derived from LIs, so for the upscaling we used total length of road no matter what road types those are.

Figure 5: Typo? “of collected air samples” instead of “of air samples collected”. The authors should also show the microbial and pyrogenic clusters on these figures (L342).

- Done
- We tried to add additional boxes to the figure, but this makes the figure quite busy and therefore we prefer to highlight the “gas leaks” category only.

L352: Typo? “combustion-related” instead of “combustion, related”.

- Done

L360: Not clear which criteria for CH_4/CO_2 ratios the authors used to classify LIs as combustion-related in the end. $CH_4/CO_2 > 0.2$ ppb/ppm?

- The criteria we used to identify combustion-related signals are based on CH_4 enhancement to CO_2 enhancement (CH_4:CO_2 ratio (ppb:ppm)). If the ratio is between 0.02 and 20 ppb:ppm and linear regression enhancements of these two species has R^2 greater than 0.8, we attributed those LIs as combustion-related sources. This has been specified in the revised version as follows (see Sect. 3.2, L469): *Based on the CH_4:CO_2 ratio (ppb:ppm) criterion defined above (see Sect. 2.3.1),…*

Figure 6: This figure is relatively difficult to interpret, it is difficult to visualize the shape of the observed plumes when they superimposed like this. It would have been interesting to see how and where you triangulated the location of the source for this site. How many sources did you find for this site? In wastewater treatment plant, the main methane source usually correspond to the sludge treatment areas that can be spotted with Google Earth.

- Figure 6 gives the overview of measurements around the WWTP. An example of the shape of a plume is given in Figure 7.

L375-377: The definition of the error estimate is very confusing, what are the 5 sets of measurements if there were only 3 days of measurements at the wastewater treatment plant?

- On some days there were two sets of measurements per day; e.g. one in the morning and one set in the afternoon. We have now defined the definition of measurement set in the paper which described back to back measurement downwind each facility as follows (see Sect. 2.2.3, L344-345): *A set of plumes is defined as a back to back transects during a period of time downwind each facility on different days.*

L395: Typo? Extra space before “74%”.

- Done

L426: Typo? One of the “%” should be removed.

- Done

L413-432: The author should expand the discussion about the different source attribution approaches, is it necessary to use all of them? Which approach would the authors recommend to use in the future?

- The following sentence was added to the manuscript (see Sect. 4.2, L547-549): *Overall, C_2H_6 and CO_2 signals are very useful in eliminating non-fossil LIs in mobile urban measurements and with improvements in instrumentations, analyzing signals of these two species along with evaluation of CH_4 signals can make process of detecting pipeline leaks from NGDN more efficient.*

L479: Shouldn’t it be the “annual natural gas leakage rate per capita” rather than the “annual natural gas consumptions per capita”?

- No, this sentence refers to the annual gas consumption provided in the previous mentioned sentence and intends to give a comparison between consumption per capita in Utrecht, Hamburg, and US.

L480: Typo? “per km of pipeline” rather than “per km pipeline”?

- Done

L491: The authors already explained several times that natural gas emissions depends on the age of the pipelines and the type of material used for these pipelines. I am not sure it is useful to repeat it here, especially since it will be discussed again later (L514).

- Here (Sect. 4.3), we mention the pipeline material and age, as these have important influences on the emissions from NGDNs in different cities, and later we give more information on different types and age of pipeline (see Sect. 4.4).

L545-549: The authors should choose one unit for the emissions and use it for all the sources, it would make easier for the reader to compare these emissions (wastewater treatment plant in t/yr, wells in kg/h...).

- Done

L557: Typo? “For emissions from the NGDN, the urban...” rather than “For emissions from the NGDN the urban...”.

- Done

L545-557: Did the authors also looked at the ratios of these larger facilities? It could be also be an interesting information.

- Correlations between CH_4:CO_2 for the facilities were not very good, which may be due to the relatively small enhancements of CH_4 downwind the facilities and the expected ratio of CH_4:CO_2.
- For the Utrecht WWTP, a ratio of 0.4 ppb:ppm of CH_4:CO_2 with R^2 of about 0.52 was observed. The sludge treatment part of the WWTP emits both CH_4 and CO_2 while CO_2 is also emitted from other parts of the WWTP, e.g. power generation, anoxic/anaerobic treatment part, which explains why the correlation is not very high.

- Downwind the Compost and Soil Company in Hamburg the CH_4 enhancement was low and no clear correlation between CH_4 and CO_2 was observed.

Supplementary information: Section 1: "Figure S2a and Figure S2b show total length..." rather than "In Figure S2a and Figure S2b total length...are shown". Same for "In Table S1 and Table S2".

- Corrected

Section 2.1: Typo? Should it be "CH_4-only mode, which show" (rather than "CH_4-only mode. which show"). It is indeed very strange that the higher inlet measures higher methane enhancements than the bumper inlet. Would it possible that this source was located above the ground ("chimney" emissions or like the sewer pictures showed below)?

- The typo has been corrected, thanks for spotting this.
- Based on the CO_2 and C:C1 analysis this LI can only be attributed to a source of natural gas emission, likely from a pipeline leak in the ground.
- We are presently investigating the influence of intake height and instrument response in more detail for an upcoming publication, where measurements in several cities will be compared. Qualitatively, the relatively slow flush time of the cavity and lower measurement rate in the G2301 relative to the G4302 instrument (see comment above) lead to generally higher maximum enhancements in the G4302 instrument compared to the G2301, which for our measurements in Hamburg and Utrecht counteracts the fact that the inlet of the G2301 is closer to the ground and thus closer to most emission points. For individual plumes, turbulence in the street from driving cars can occasionally lead to higher mole fractions at the top inlet.

Section 2.2: What does "the ratio of the sum of CH_4 enhancements (in ppb) to the sum of CO_2 enhancements (in ppm)" mean? Does it correspond to the area under the plume? There is no mention of Figure S7 in the text.

- The respective sentence was changed as follow (see Sect. S.2.6): *In Figure 12, the ratio of the area under the CH_4 enhancements along the driving track (in ppb*m) to the area of CO_2 enhancements along the driving track (in ppm*m) is 5.5 ppb:ppm which is much higher than reported in previous studies, possibly indicating incomplete combustion.*

Section 2.4: "Errors in wind speed are estimated to be ±10% and for wind direction ± 5°" this seems low to me considering that the wind was not measured on site but at a tower located away from the site. Table 5 caption should be better isolated from Table 4, this is a bit confusing at the moment.

- By comparing some of the recorded measurement from the 2-D anemometer next to the facilities with the data from the towers we noticed that the local wind speed data were within ± 10% of the data from the towers; as described above, following the comment of the referee we now use a more conservative estimate of ± 30%.
- Given that for many sources the emission point is known, either from satellite imagery or from triangulation, wind direction (between emission point and maximum of observed plume is quite well known and here we think that the error estimate of ± 5° is adequate.
- The caption for Table 5 has also been corrected.

Section 2.6: In Figure S10a, shouldn't the authors constraint delta^{13}C, deltaD, C_2H_6 and CO_2 before clustering? It would avoid clustering enhancements from different types of sources. Figure S11: caption not very precise.

- Based on the Weller et al. (2019) algorithm, it is assumed that LIs which are clustered together should be from the same source. Thus, based on the algorithm if one of the LIs within a cluster belongs to a specific emission class (e.g. microbial or combustion, etc.) then all the others should have fall into that source class. Based on the multi-tracer and isotope data, we have no evidence that in our dataset this is not the case. Therefore, we kept the analysis this way to keep consistency with other studies where no attribution techniques were used to attribute the LIs.

References

von Fischer, J. C., Cooley, D., Chamberlain, S., Gaylord, A., Griebenow, C. J., Hamburg, S. P., Salo, J., Schumacher, R., Theobald, D. and Ham, J.: Rapid, Vehicle-Based Identification of Location and Magnitude of Urban Natural Gas Pipeline Leaks, Environ. Sci. Technol., 51(7), 4091–4099, doi:10.1021/acs.est.6b06095, 2017.

Turner, D. B.: Workbook of Atmospheric Dispersion Estimates, U.S. Environmental Protection Agency. [online] Available from: https://nepis.epa.gov/Exe/ZyPDF.cgi/9101GKEZ.PDF?Dockey=9101GKEZ.PDF, 1969.

Weller, Z. D., Yang, D. K. and von Fischer, J. C.: An open source algorithm to detect natural gas leaks from mobile methane survey data, edited by M. Mauder, PLoS One, 14(2), e0212287, doi:10.1371/journal.pone.0212287, 2019.

Weller, Z. D., Roscioli, J. R., Daube, W. C., Lamb, B. K., Ferrara, T. W., Brewer, P. E. and von Fischer, J. C.: Vehicle-Based Methane Surveys for Finding Natural Gas Leaks and Estimating Their Size: Validation and Uncertainty, Environ. Sci. Technol., acs.est.8b03135, doi:10.1021/acs.est.8b03135, 2018.