This document provides our answers to the anonymous referee #2 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

The authors present an extensive study of ground-based mobile measurements of methane and several related tracers (C_2H_6, CH_4/CO_2 ratio, δ^2H-CH_4 and $\delta^{13}C$-CH_4 focused on quantifying and attributing methane emissions in two European cities, namely Utrecht and Hamburg, which both rely on subterranean pipelines as the delivery system for natural gas used in the households and otherwise. Such delivery systems are known to cause leaks that contribute to the anthropogenic global warming, and it has been demonstrated previously that fixing of these gas leaks can be a very cost-effective mean of climate change mitigation. Using a combination of in situ observations (with CRDS), discrete samples collected for identified leak sources as well as Gaussian plume modelling, the authors are able to identify approximately 100 leak sources in both cities over their study period, and thanks to the robust analysis of collected data, are able to differentiate them according to emission source (natural gas distribution system / microbial sources) and the respective source strength (with just several sources responsible for large parts of total emissions). A comparison of the results against previous studies conducted in US point to potential lower specific emissions for the studied cities. The authors also attempt to upscale the measurements performed over these limited campaigns in order to compare them to the publicly available aggregated data, albeit these results should be treated with care as the dataset is limited and much more robust studies are needed to achieve this goal (which the authors accurately point out).

The authors should be commended for the impressive amount of high-quality work that was put into design, execution and data analyses during those campaigns. This is no easy task, as the study encompassed simultaneously using many state-of-the-art techniques from very different scientific fields together with a very large amount of data (both measurement and supplementary) in order to achieved the stated goals. Simultaneous analysis of several tracers and isotopic composition is of particular interest, as it shows great promise in development the methods of precise small-scale emission estimations. I find this study to be a strong contribution to the discussion in the city-scale methane emissions, and the strategy developed here seems to be promising in developing both research-targeted and operational methods for leak detection and its strength estimation.

The article does suffer however, from this wealth of data and methods, and requires multiple improvements before final publication. For example, some sections of the text require more detailed information in order for the study to be considered reproducible. Also, the treatment of uncertainty in the source estimation and Gaussian plume modelling sections should be deepened. In the second case, the sensitivity of the method to the chosen meteorological parameters should be established. In methods section some restructuring is recommended, and the Discussion could benefit from introducing a clearly defined structure in order to appropriately focus attention.
I recommend publishing the article after addressing items listed below.

2 Major comments:

1. The Method description could do with an overhaul. In some places, more details need to be provided (see 'detailed comments' section below) for the experiments to be considered reproducible. In others, some information should be combined (2.6.4. and 2.4.).
- We have revised the method section accordingly, see our answers to the detailed suggestions below.

2. Discussion of uncertainties in the urban emissions is very limited, with authors stating that 'We used a Bootstrap method (Nelson, 2008) to estimate 247 emission uncertainties similar to Weller et al. (2018) for the US city studies by resampling from all recorded LIs randomly 30,000 times.' No further comment is given, and in the discussion section the authors quickly skim over this and analyze the statistics of LI, without providing information on how precise those classifications might be. The method described by Weller et al. (and earlier by von Fisher et al.) relies heavily on assumptions regarding the distance to the source and (calibration using control releases) it stands to reason that this simplistic approach must produce very large uncertainties if not supported by multiple measurement repetition. This is critically important here as the data from limited detections is interpreted and up-scaled. As it is now, it is not possible to get a realistic impression about the numbers given, and puts the resulting data analysis in question. The discussion of uncertainty and potential biases should therefore be expanded.
- As written in the manuscript, we follow the algorithm that has been described in detail in von Fischer et al. (2017) and Weller et al (2019). These studies include information on the precision, and we do not think that it is necessary to repeat all of the work that was performed there in our manuscript. We follow their approach so closely because we feel the inter-comparison with the US studies is important for a better overall understanding. The algorithm was applied to LIs that were surveyed at least twice, which we did not have as a strict criterion in our study. This difference is discussed in detail in our manuscript. Revisiting is not expected to have a large influence on the classification, since the Weller et al. (2019) algorithm only considers the maximum reading of the (at least two) identifications, so a revisit can only result in a higher emission rate estimate.
- Like other studies, we do not have information on the distance of the leaks from the point of observation, since we do not have access to the full grid of pipelines in both cities. We do mention, that individual leak rate estimates can have large errors, and have included this statement explicitly regarding the highest derived leak rates in the revised manuscript. Since we covered a very large fraction of the city, the error of the total emissions is relatively small, i.e., about 30 % and 15 % in Utrecht and Hamburg, respectively. The bootstrapping is a standard technique and it has been demonstrated that it performs well in emission quantification and the determination of emission factors. The bootstrapping is a standard technique and it has been proven that it performs well in emission quantification and the determination of emission factors.
- The following paragraph is now expanded in Sect. 2.2.2, L292-301:
To account for the emission uncertainty, similar to Weller et al. (2018) for the US city studies, we used a bootstrap technique which was initially introduced in Efron (1979, 1982), as this technique is adequate in resampling of both parametric and non-parametric problems with even non-normal distribution of observed data. Tong et al. (2012) indicated that bootstrap resampling technique is sufficiently capable in estimating uncertainty of emissions with sample size of equal or larger than 9. Efron and Tibshirani (1993) suggested that minimum of 1,000 iterations are adequate in

bootstrap technique. In this study, we used non-parametric bootstrap technique to account for the uncertainty of total CH_4 emissions from all LIs in each city with 30,000 replications. As mentioned above the algorithm is based on CH_4 enhancements of measurement with 5-40 m distance from controlled release location, and can produce large uncertainty for emission quantification of individual LI (Figure 4 in Weller et al. (2019)), but with sufficient number of sample size, the uncertainty associated with total emission quantified in an urban area is more precise.

3. The authors state that 'Emissions from facilities show significant contributions to the total emissions in both cities.' (supplement), but many details on the method used for estimating them are missing. This section needs to be expanded and more info should be given about the analysis as well as the uncertainty estimation and sensitivity of the method to the stated assumptions. For example, the use of measurement data from distant towers in order to drive the transport model raises an eyebrow, as these are critical for calculating the emission rate. What was the average distance between the tower and the measurement location and what was the elevation of that measurement? Are the wind speed and direction uncertainties reasonable? What about the elevation of the source, which is only very briefly discussed in the supplement? In the end, the reader should have a comprehensive view on whether the method is able to provide good emission estimation in a given setting, and at the moment the result with error bars (on the order of ±50 % 1-σ) suggest it is not. This should also be discussed in more detail.

- The focus of our paper is on the detection of methane elevations from unknown sources with the mobile vehicle approach and their attribution to natural gas, microbial and combustion sources. In addition to this new attribution component, we also decided to report rough emission rates derived for larger known facilities, in contrast to previous publications that excluded emissions from such facilities. The reason was to demonstrate that such elevations are also picked up by the mobile measurements, to estimate whether estimates on emission rates can be derived, even if the approach was not targeted to facilities, and to put the order of magnitude of emissions from the unknown sources into perspective of the known sources (e.g. waste water treatment plants).

We kept this section short on purpose to not distract too much from the main objective but realize from the referee reports that a more detailed is warranted. Some more description has been incorporated in the revised version of the manuscript, and the replies to the specific questions are included in our answers to the detailed comments below.

In addition to the information which had been already provided in supplementary section we now have expanded the paragraphs related to the emission height (L333-341) and wind speed (L351-358) in Sect. 2.2.3 as follows:

Neumann and Halbritter (1980) showed that the main parameters in sensitivity analysis of GPDM are the wind speed and source emission height in close distance and the influence of emission height become less further downwind compared to the mixing layer height. In this study, the heights of emission sources were low (<10m) and/or estimated during surveys and Google Earth imageries, and considering that such a larger measurement distance from the facilities, the main sources of uncertainty of the emission estimates for the WWTP and Compost and Soil company are most likely the mean wind speed and for the upstream facilities in Hamburg the major sources of uncertainties can be the mean wind speed and emission height. We considered 0-4 m source height for the WWTP in Utrecht, and for the upstream facilities in Hamburg we considered 0-5 m emission height for the Compost and Soil site, 0-2 m for the separator,

0-10 m for the storage tank, and 0-1 m for the oil extraction well-head. We used 1 m interval for each of these height ranges to quantify emissions in GPDM.

Due to technical issues, local wind data were not logged continuously and thus we used wind data from two towers which are 8 to 20 km away from the facilities we focused for emission quantifications. These distances introduce extra uncertainties in analyzing the emissions using GPDM mainly on the wind speed. By comparing some of the local high-quality wind data to data from the towers, we estimated that the local wind speed is within the range of ± 30 % of the collected tower data. This range was adopted to estimate the wind speed for emission quantifications for the set of plumes measured downwind of the facilities. The wind directions were aligned at local scale of each facility based on the locations of sources and locations of maxima of average CH_4 enhancements from a set of transects in each day's survey and we considered ± 5° uncertainty in wind direction for the GPDM quantification.

4. The discussion section would benefit from introducing subsections to provide focus for specific items under discussion.

- This has been updated, and we agree that this provides a clearer structure.

5. I find the overall quality of language very high, yet there are multiple minor deficiencies that still need to be addressed. Below I have listed some of them. I believe that this is mostly due to heavy editing during manuscript preparation, and I ask the authors to take special care of that issue before resubmitting.

- Thanks for pointing out several detailed issues, which have all been incorporated (see detailed replies below). The manuscript has also been carefully proof-read again.

6. The supplement is large and – I'm sorry to say that – poorly edited (tables are too large - the font can safely be made much smaller; order of figures and sections does not correspond to the manuscript reference order). In some cases, itis a source of important information that is also in some places missing from the main text (already mentioned section 2.4. and corresponding S.2.4, figure S16). If the authors want to keep some technical details apart from the main text (understandable with that much material), then I would ask to consider putting the more important parts in the Annex, in order to a) maintain the high editing standard and thus make reading easier, b) keep the important information together with the text. At the very minimum the editing of the supplement needs to be improved.

- The editing of the supplement has been updated according to the suggestions of the referee. We have moved some of the text from the supplement to the main text, as suggested.

3 Detailed comments:

Line numbers are given for identification. Comments for figures are given at the end of the list.

L25, also later in the text: ACP requires exponential notation of units, consult the 'manuscript preparation' on the ACP website for details (www.atmospheric-chemistry-and-physics.net/for_authors/manuscript_preparation.html)

- This has been adjusted in the revised version of the manuscript

L35: I'd suggest putting ppm outside of parentheses and mole fraction inside, as the ppm/ppb notation is the dominant one in the manuscript.

- The sentence was changed accordingly (L35-36): *The increase of CH_4 mole fraction from about 0.7 parts per million (ppm) or 700 parts per billion (ppb) ...*

L46-49: Sentence needs rephrasing
- The sentence was rephrased as follows (see Sect. 1, L47-50): *CH_4 emissions originate from a wide variety of natural and anthropogenic sources, for example emissions from natural wetlands, agriculture (e.g. ruminants or rice agriculture), waste decomposition, or emissions (intended and non-intended) from oil and gas activities that are associated with production, transport, processing, distribution, and end-use of fossil fuel sector (Heilig, 1994).*

L60-62: This paragraph does not fit well here, would be better if info given as part of previous or next.
- The information has been added to the end of previous paragraph (L60-63).

L64: 'high precision' used twice
- Corrected

L75-77: What were the main findings from these studies? Specifically, it would be good to comment on whether these methods can be useful for up-scaling.
- The following information has been added to the manuscript (see Sect. 1, L78-86): *Gioli et al. (2012) showed that about 85 % of methane emissions in Florence, Italy originated from natural gas leaks. Helfter et al. (2016) estimated CH_4 emissions of 72 ± 3 t km^{-2} yr^{-1} in London, UK mainly from sewer sesytem and NGDNs leaks, which is twice as much as reported in the London Atmospheric Emissions Inventory. O'Shea et al. (2014) also showed that CH_4 emissions in greater London is about 3.4 times larger than the report from UK National Atmospheric Emission Inventory. Zimnoch et al. (2019) estimated CH_4 emissions of $(6.2 \pm 0.4) \times 10^6$ m^3 $year^{-1}$ for Krawko, Poland, based on data for the period of 2005 to 2008 and concluded that leaks from NGDNs are the main emission source in Krawko, based on carbon isotopic signature of CH_4. Chen et al. (2020) also showed that incomplete combustion or loss from temporarily installed natural gas appliances during big festivals can be the major source of CH_4 emissions from such events, while these emissions have not been included in inventory reports for urban emissions.*

L78: 'We quantified emissions in this study using measured CH_4 enhancements above background, which were detected' - This needs revision; also, it feels like Weller et al. 2019 should already be quoted here, perhaps something like: 'In this study, we have quantified the CH_4 emissions using the method described by Weller et al., who demonstrated...'
- The following sentence has been added to the manuscript (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.*

L91: Was the reproducibility tested by the authors? Picarro currently gives 0.5 ppb for 5 s raw data. If the reproducibility was tested by the authors, please provide some details on the testing (either reference, or brief description of the experiment). Was the water correction modified or the factory settings were used? Please state that explicitly and also provide information if necessary.

- We now added more info about the G2301 to the manuscript (see Sect. 2.1.1, L100-108).
- The numbers represent the approximate range of the instrument noise when measuring background air. This information is provided to indicate to which order of magnitude we can identify elevations of CH_4 and C_2H_6.
- The factory settings were used for the water correction.

L99: Discard 'about' or the approximation sign
- Done

L100-101: Similar to comment for L91, please provide more info.
- Similar to the details provided for G230, information about the Picarro G4302 is now expanded. We also used factory setting to consider water correction (see Sect. 2.1.1, L109-117).

L104: Info on how the delay was calculated should be given here, but can be found later in L202. Please combine both (see major comment no. 1)
- Done

L111-112: Please spell out the main findings of the discussed comparison. Also, the reference to annex section number (S.2.) where it is discussed should be present (next to table S3 ref.). In general, sections of supplement should be referenced and not only tables or figures from it.
- The comparison sentence is now edited as follows with the main findings from the comparison (see Sect. 2.1.1, L135-142): *A comparison between the two instruments during simultaneous measurements showed that all LIs were detected by both instruments despite difference in instrument characteristics and inlet height. In the majority of cases CH_4 enhancements for each LI from both instruments were similar to each other. We note that there is likely a compensation of differences from two opposing effects between the two measurement systems. The inlet of the G2301 was at the bumper, thus closer to the surface sources, but the rather low flow rate and measurement rate of the instrument lead to some smoothing of the signal in the cavity. Because of the high gas flow rate, signal smoothing is much reduced for the G4302, but the inlet was on top of the car, thus further away from the surface sources (see* **Error! Reference source not found.** *in SI, Sect. S.1.3).*

L129: When reading the sentence for the first time I have understood that the gas pipeline network corresponds to the street network 1:1. Is that correct, or the general coverage of municipal areas is meant? Please clarify.
- That is correct. The sentence was changed as follows (see Sect. 2.1.2, L162-165): *The local distribution companies (LDCs) in Utrecht (STEDIN (https://www.stedin.net/)) and Hamburg (GasNetz Hamburg (https://www.gasnetz-hamburg.de)) confirmed that full pipeline coverages are available beneath all streets. Therefore, the length of roads in the study area of Utrecht and Hamburg are representatives of NGDNs length.*

L137: discard 'at the following links: Utrecht and Hamburg'
- Done

L147-148: Please briefly explain how the vehicles can be methane sources (with reference for subsequent discussion further in the text).
- The sentence was changed as follows (see Sect. 2.1.3, L183-185): *For example, they could be related to emissions from vehicles which run on compressed natural gas, or*

vehicles operated with traditional fuels but with faulty catalytic converter systems. Later we will discuss how to exclude or categorize these unintended signals (see Sect. 2.2.2 and Sect. 2.3.1).

L148: Please state clearly how many revisits were usually made.
- The respective sentence was changed as follows (see Sect. 2.1.3, L185-187): *Therefore, we revisited a large number of locations (65 in Utrecht (≈80 %) and 100 in Hamburg (≈70 %)) where enhanced CH₄ had been observed in during the first survey in order to confirm the LIs.*

L159-161: Have any cases where new leaks have occurred in-between surveys been observed?
- There have been some locations where we observed new LIs during revisits but not in the earlier visits or all the way round. This can be mainly due to the fact that not all LIs (mostly small LIs) are observable in all visits.

L166-167: Parentheses missing? 3 L bag for a price of 2 L bag is too good to be true.
- Corrected

L168-169: More details on sampling are needed.

- Was the data collection stationary or also mobile?
  The following sentence has been added to the manuscript (Sect. 2.1.4, L206-209):
  All the samples taken in the North Elbe study area and from most of the facilities were collected when the car was parked, but the samples inside the New Elbe tunnel and close to some facilities where there was no possibility to park were taken in motion while we were within the plume.

- How was the plume / non-plume location determined?
  The following sentences have been added to the manuscript (Sect. 2.1.4, L209-211):
  The sampling locations across the North Elbe study area of Hamburg were determined based the untargeted surveys, and the confirmation during revisits. The C₂H₆ information was not used in the selection of sampling locations in order to avoid biased sampling.

- What was the flushing time?
  The samples were taken using a pump with flow rate of 0.25 L min⁻¹. [info added to the sampling section (Sect. 2.1.4, L214)].

- Was the sample dried? How?
  Samples were dried in the lab followed by the CH₄ extraction. [info added to the lab analysis section (Sect. 2.3.2, L383-384 and L391-392)].

L173: See major comment no. 3.
- The relevant section has been expanded to provide more information (see Sect. 2.2.3). 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 wind speed.

- 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 referee, we increase this to a more conservative error estimate of ± 30 %. Regarding the wind direction (e.g. the oil wells), for several sources the emission point is relatively certain and can 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, L352-359).

L191-192: Case shown in S5 is special and I strongly recommend to remove it from here and discuss later. As it is now, the text does not explain it, and thus may imply that all the cars are potential sources of CH_4, which is certainly not the case.
- Section 2 has been rearranged and this part has been moved to the sub-section of '2.3. Emission attributions' (Sect. 2.3.1, L367-370). The text has been revised as follows to mention that not all vehicles emit CH_4 but vehicles running on compressed natural gas. *Compressed natural gas vehicles can be mobile CH_4 emission sources (E. K. Nam et al., 2004; Curran et al., 2014; Naus et al., 2018; Popa et al., 2014) and in this study we also observed CH_4 signals from vehicles. For example, the point to point C_2H_6:CH_4 ratio (C_2:C_1) calculated from road measurements of a car exhaust shown in **Error! Reference source not found.** is 14.2 ± 7.1 %.*

L194: Consider providing these standards in the parentheses or the supplement section S.7.
- Done

L196: How many such cases were observed? Could they be important for the overall budget?
- In the study are of Hamburg (north Elbe), there were 34 cases and in the study area of Utrecht (inside the ring) there were 7 cases which we excluded based on this constraint (C_2:C_1 >10). Note that the methane/ethane instrument (Picarro G4302) was used for all surveys in Hamburg, while this instrument was not available for all surveys in Utrecht.
- The CH_4 enhancements measured by the G4302 for these cases were 0.37 ± 0.24 ppm in Hamburg and 0.26 ± 0.03 ppm in Utrecht. Based on the quantification from Weller et al. (2019), these LIs are not important to the overall budget but they will of course affect the total number of leak indications (LIs). Therefore, it is important to exclude them from the evaluation.

L198: Please state the reasons for this exclusion, briefly.
- The speed constraints are now reformulated as follows (see Sect. 2.2.2, L253-256): *Following the algorithm from von Fischer et al. (2017), measurements at speeds above 70 km h^{-1} were excluded, as the data from the controlled release experiments (von Fischer et al., 2017) were not reliable at high speed (Weller et al., 2019). We also excluded measurements during periods of zero speed (stationary vehicle) to avoid unintended signals coming from other cars when the measurement car was stopped in traffic.*

L199-200: Just a small comment, no action needed: I don't see the benefit of this artificial increase of the data frequency. This brings no new information at the cost of tripling of the data that needs to be processed.

L200-201: Wording. If the time was just converted to UTC, then calling it 'a correction' is not warranted. Consider changing to: 'Following the interpolation step, the data was converted to UTC, and subsequently corrected for ...'

- We used the word 'correction' as the clock on the Picarro instruments had a drift over the period of the campaign and we needed to correct this drift to set all the measurements to the correct UTC time. The text has been adjusted as follows:
The clocks on the Picarro instruments were set to UTC but showed drift over the period of the campaigns. We recorded the drifts for each day's survey and corrected to UTC time. The data were also corrected for the delay between air at the inlet and the signal in the CH_4 analyzers. This delay was determined by exposing the inlet to three small CH_4 pulses from exhaled breath, ranging from 5-30 seconds, depending on the instrument and tubing length. We averaged the three attempts to determine the delay for each instrument and used the delays for each instrument. Individual attempts were 1 to 2 s different from each other. For the G4302 the delay was generally about 5 s and for the G2301 it was about 30 s; the difference is mainly due to the different flow rates. The recorded CH_4 mole fractions were projected back along the driving track according to this delay.

L202: About the delay time estimation: 5–30 seconds is a very broad range. Were the ranges so variable for both instruments, or was it 5 for one and 30 for the other? Also, how was the pulse generated? Can you estimate precision of that delay estimation (even grossly)?

- The answers to this comment are merged with the answers to the previous comment (see Sect. 2.1.1, L122-129).

L204: Reference order needs correction. Previous reference supplement figure was S5 (L190).

- Corrected

L207: In CO_2 signal (Fig. S8), it can be clearly seen that the background line is some-times higher than the observed signal. Since this plot is about the background, it would be good to change the limits of y axis to make the calculated background visible clearer, especially for methane. Please give some comment about the possible negative enhancements after subtracting such background (can it affect the estimation of emissions?).

- For CO_2, individual plumes, e.g. from vehicles, can overlap and create a locally enhanced background according to our background extraction procedure (5th percentile of CO_2 measurements in a ± 2.5 min time window). Negative deviations are not considered, but the enhanced background may result in a potential small underestimate of the CO_2 enhancement. Such small changes will not affect the categorization of the CH_4 enhancement based on the $CH_4:CO_2$ ratio, very wide ranges were assumed for this source attribution. We have updated the figure in the revised version as suggested (see Figure S7 in SI, Sec. S.2.1).

L217: Please add 'peak' after enhancement, to make it clear that it's not about the release height.

- We changed this to: *...to demonstrate that the magnitude of the observed methane enhancement...* (see Sect. 2.2.2, L264)

L237: Wording. Why should results from different cities be comparable? The authors clearly mean that the analysis software used on a given dataset should be comparable. Please clarify. Actually, this whole paragraph can be limited to information that 'Our software was compared

to analysis tools developed by CSU (von Fisher et al.2017, Weller et al. 2019) and no significant differences were observed (see SI, section S.2.7)'.

- Thank you for the suggestion, we updated the paragraph as follows (see Sect. 2.2.2, L284-286): *We compared the outputs of our software to the one developed by Colorado State University (CSU) for the surveys in US cities (von Fischer et al., 2017; Weller et al., 2019). 30 LIs were detected and no significant differences were observed (linear fit equation y = 1.00 * x - 0.00, R² = 0.99) (see SI, Sect. S.2.4, Figure S10).*

L251: Erase 'areas'.
- Done

L252: Erase empty parentheses.
- Done

L253 and L384: 'drive-by'- I propose 'mobile'. I was surprised to find it used in Fisher et al. (albeit only once), as in U.S. this word is sometimes used to describe something much more nefarious then GHG observations.
- Done

L254: 'We report (...)' – Unclear what is meant in this sentence, please rephrase it.
- The sentence was rephrased as follows (Sect. 2.2.3, L309-310): *In this study, we also report the data obtained from larger facilities, since rough emission estimates from facilities can be obtained in the city surveys.*

L256-257: Erase 'both' and 'each day's'
- Done

L269-274: How was the release height determined? How is the uncertainty of this determination included in the uncertainty of emission?
- Information on the release height and uncertainty associated with each facility has been added to this section (see answer to the major comment 3).

L283: This section should be combined with 2.4.
- Section 2 has been rearranged and information regarding air sampling collection is now provided in Sect. 2.1.4 and details on analysis of samples are provided in Sect. 2.3.2.

L286: Info on the isotopic scales used in this study needs to be given.
- The information on the scale and instruments for isotopic measurement is now added (see Sect. 2.3.2, L380-383 and L393-395).

L293-294: Please provide explanation on why these particular ranges were selected. For signatures, specifically also provide references supporting the choice of isotopic signatures. Please keep in mind that for fossil fuel signatures, figure 7 from Rockmann et al. 2016 doesn't give a full picture – see e.g. Sherwood et al. 2017 for a broad overview of isotopic signatures for fossil methane.
- We acknowledge that the full range of isotopic signatures of natural gas from different reservoirs is wide, in particular for $\delta^{13}C$, as documented in Sherwood et al. (2017), and we have indeed adjusted this range in a recent publication of Menoud et al., (2020). Nevertheless, this does not mean that the full range of signatures would be encountered in one region in a single campaign. No data are available for Hamburg, but Levin et al., (1999) reported a mean $\delta^{13}C$ value of -40 ‰ for the NGDN in Heidelberg, with a

seasonal cycle of ± 10‰. Hoheisel et al. (2019) reported $\delta^{13}C$ = -43.3 ± 0.8‰ for the NGDN in Heidelberg the period 2016-2018, but without the strong seasonality, and -46.1 ± 0.8‰ for natural gas storage tanks and compressor stations. These values are compatible with the selected range that we have chosen based on Röckmann et al. (2016).

L314-315: 'were correspond' – corresponded
- Done

L315: What is the uncertainty here? This relates to major comment 2.
- See the answer to major comment 2. We have also added the following sentence (Sect. 3, L422-424): *Noted that estimates for individual leaks with the Weller et al. (2019) algorithm can have large error, thus these results are indicative of large leaks, but the precise emission strength is very uncertain.*

L378: Why uncertainty given only for wind direction?
- Corrected

L394: von Fisher – V should be capitalised at the start of the sentence (Von Fisher).
- Done

L404: 'About 50 %' - please give the specific number that was used in the calculation.
- We used precisely 50 % for both cities for the calculation. The wording has been adjusted.

L420-422: These numbers are in fact very similar. The variability of $\delta^{13}C$ in the natural gas can be quite substantial. See e.g. Fig 4 in Sherwood et al, 2017.
- We have updated this sentence as follows (see Sect. 4.2, L533-537): *These numbers do not agree within combined errors, but are also not very different. $\delta^{13}C$ values of CH_4 from the NGDN can vary regionally and temporally, e.g. due to differences in the mixture of natural gas from various suppliers for different regions in Germany (DVGW, 2013). It is also shown that how $\delta^{13}C$ values of fossil fuel CH_4 have significant variabilities in different regions at global scale (Figure 4 in Sherwood et al. (2017)).*

L426: % used twice
- Corrected

L433: Reference to Figure 1?
- Corrected

L436-438: a) Fig S16 also points to the local sewage system as potentially important source, but this is not mentioned here. b) Please be more precise in the argument here- i.e. explain why measurements around the lake point to anaerobic methanogenesis specifically. Linking to a), please include info on the potential role of the sewage system if needed. Is it possible that the sewage is seeping into the lake?
  - a) We changed this sentence as follows (see Sect. 4.2, L553-559): *Many of the microbial LIs encountered in Hamburg are around the Binnenalster lake (**Error! Reference source not found.**), which suggests that anaerobic methanogenesis (Stephenson and Stickland, 1933; Thauer, 1998) can cause these microbial emission in this lake, as seen in other studies focused on emissions from other lakes (DelSontro et al., 2018; Townsend-Small et al., 2016). Microbial CH_4*

emissions from sewage system (Guisasola et al., 2008) can also be an important source of in this area, as seen in US urban cities (Fries et al., 2018). Fries et al. (2018) performed direct measurement of CH_4 and nitrous oxide (N_2O) from a total of 104 sites, and analyzed $\delta^{13}C$ and δD signatures of samples from 27 of these locations, and attributed 47 % of these locations to microbial emissions in Cincinnati, Ohio, USA.

    b) We described in the text that the isotope and multi-tracer observations point to microbial sources. Unfortunately, we do not know if and/or to what degree the sewage system seeps into the lake.

L443: 'because there is no publicly available activity data for associated activity data'– please rephrase.

- The sentence was rephrased as follows (see Sect. 4.3, L564-566): *Also, it is not possible to calculate a robust city-level estimate using the nationally reported emission factors because there is no publicly available associated activity data, i.e., pipeline materials and lengths for each material, at the level of individual cities.*

L452: Too many parentheses. I suggest '(...) 40 kg km⁻¹ yr⁻¹ (for other material, p <200 mbar; see p. 130 in Peek et al., 2019, for details)

- The sentence was split in several parts and rewritten as follows (see Sect. 4.3, L571-577): *The Netherlands National Institute for Public Health and the Environment (RIVM) inventory report derived an average NGDN emission factor of ≈ 110 kg km⁻¹ yr⁻¹ using 65 leak measurements from different pipeline materials and pressures in 2013. This weighted average ranged from a maximum of 230 kg km⁻¹ yr⁻¹ for grey cast iron pipelines to a minimum of 40 kg km⁻¹ yr⁻¹ for pipelines of other materials with overpressures <= 200 mbar (for details, see P. 130 in Peek et al. (2019)). This results in an average CH_4 emissions of ≈ 70 t yr⁻¹ (min = 30 t yr⁻¹ and max = 150 t yr⁻¹) for the study area of Utrecht, assuming ≈ 650 km of pipelines inside the ring, and further assuming that Utrecht's NGDN is representative of the national reported average (see qualifiers above).*

L473: 'credibility interval' – confidence interval

- Done

L491: '(...) factors can be gas pipeline age and material, sewer system.' Part of sentence missing? Please rephrase.

- The sentence was changed as follows (see Sect. 4.3, L611-613): *CH_4 emissions can vary among different cities, depending on the age, management and material of NGDNs, and/or the management of local sewer systems.*

L531: 'were' – where

- Corrected

L533: 'as shown' used twice.

- Corrected

L542-544: The scheme from S18 cannot be treated as a 'protocol' without a proper description of the method. In reality, it describes the main components of the method applied in the study, so in fact the manuscript itself is more of such a protocol. As it stands now, consider either expanding the description in the supplement (so gas companies might actually use it as a protocol) or discard it altogether.

- We agree with this comment and removed the term "protocol". We reformulated the sentence as follows (see Sect. 4.4, L668-669): *Figure S19 (see SI, Sect. S.5) illustrates how the individual measurement components can be efficiently combined in a city leak survey program.*

L561: 'corresponding to emissions of about 107 tCH$_4$/ yr' – exactly 110 t yr^{-1} is given in L332.
- Corrected

L562: Please state the method, e.g. 'These estimates, based only on the studied area, were then up-scaled for the total municipal area, using the road network map as a proxy to (...)'
- The up-scaled emissions are related to the emissions across the study area; inside the ring in Utrecht and north Elbe in Hamburg and not the total municipal area of these two cities. The text has been adjusted as follows (see Sect. 5, L688-690): *These estimates, based on the streets covered, were then up-scaled to the total study area, using the road network map as a proxy for the length of the pipeline network which then yielded total emissions of 150 t yr^{-1} and 440 t yr^{-1} across the study area of Utrecht and Hamburg respectively.*

L567: 'were from' - I suggest 'originated from'
- 'were from' has been changed to '*originated from*' (L695)

Figure 2: a) Please fix the x axis description - extra arrow unnecessary. b) the plot is cropped in the lower part, by several pixels. c) extra grid dashed lines (green) are unnecessary, make the labels difficult to read.
   a) We prefer to keep the arrow. It indicates that the total emissions are attributed to four different emission sources and how much these sources contribute to the total emissions individually
   b) Done, this probably happened during conversion to pdf format.
   c) The extra green dashed lines on the x-axis have been removed
   The caption has been changed as follows: *Total CH$_4$ emission rates from different sources in (a) Utrecht and (b) Hamburg; the arrow shows how the emissions are attributed to different sources*

Figure 3: a) again, the arrow from total emissions to 'Road classes' seems unnecessary. b) please explain the arrows from the plot in the caption.
   a) The arrow is used to indicate that the total emissions are categorized into six different road classes and how much these classes contribute to the total emissions individually
   b) the caption has been changed as follows: *Total CH$_4$ emissions in Utrecht and Hamburg; the arrow shows how the total emissions are distributed over different road classes*

Figure 4: I recommend plotting all US cities in a single colour (grey?) and simply label the line as ('15 US Cities', Weller et al.) or similar. The colours are indistinguishable anyway.
- Done

Figure 5: Excellent plot! a) Please add comment about the uncertainty of δ^{13}C and δ^{2}H signatures in the caption (they were only plotted for C2/C1. I also softly suggest to label the plots with a-b-c-d and move the labels into figure caption. b) On previous plot the units were placed in parentheses, consider keeping notation consistent.
- Both suggestions have been included

Figure 6: a) Plumes of what? b) Please also provide information on whether the peaks are on the same scale; if yes, then what is the plotted range of mole fractions (if those are mole fractions)?

- Done

References

DelSontro, T., Beaulieu, J. J. and Downing, J. A.: Greenhouse gas emissions from lakes and impoundments: Upscaling in the face of global change, Limnol. Oceanogr. Lett., 3(3), 64–75, doi:10.1002/lol2.10073, 2018.

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.

Hoheisel, A., Yeman, C., Dinger, F., Eckhardt, H. and Schmidt, M.: An improved method for mobile characterisation of δ 13 CH 4 source signatures and its application in Germany, Atmos. Meas. Tech., 12(2), 1123–1139, doi:10.5194/amt-12-1123-2019, 2019.

Levin, I., Glatzel-Mattheier, H., Marik, T., Cuntz, M., Schmidt, M. and Worthy, D. E.: Verification of German methane emission inventories and their recent changes based on atmospheric observations, J. Geophys. Res. Atmos., 104(D3), 3447–3456, doi:10.1029/1998JD100064, 1999.

Menoud, M., van der Veen, C., Scheeren, B., Chen, H., Szénási, B., Morales, R. P., Pison, I., Bousquet, P., Brunner, D. and Röckmann, T.: Characterisation of methane sources in Lutjewad, The Netherlands, using quasi-continuous isotopic composition measurements, Tellus B Chem. Phys. Meteorol., 72(1), 1–19, doi:10.1080/16000889.2020.1823733, 2020.

Peek, C. J., Montfoort, J. A., Dröge, R., Guis, B., Baas, K., Huet, B. van, Hunnik, O. R. van and Berghe, A. C. W. M. van den: Methodology report on the calculation of emissions to air from the sectors Energy, Industry and Waste, as used by the Dutch Pollutant Release and Transfer Register., 2019.

Sherwood, O. A., Schwietzke, S., Arling, V. A. and Etiope, G.: Global Inventory of Gas Geochemistry Data from Fossil Fuel, Microbial and Burning Sources, version 2017, Earth Syst. Sci. Data, 9(2), 639–656, doi:10.5194/essd-9-639-2017, 2017.

Townsend-Small, A., Disbennett, D., Fernandez, J. M., Ransohoff, R. W., Mackay, R. and Bourbonniere, R. A.: Quantifying emissions of methane derived from anaerobic organic matter respiration and natural gas extraction in Lake Erie, Limnol. Oceanogr., 61(S1), S356–S366, doi:10.1002/lno.10273, 2016.

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.