In this document we provide our answers to comments from ‘Anonymous Referee #1’, ‘Anonymous Referee #2’ and joint comments from DVGW (German Technical and Scientific Association for Gas and Water) and Gasnetz Hamburg for “**Intercomparison of detection and quantification methods for methane emissions from the natural gas distribution network in Hamburg, Germany**”. Please find our answers in normal blue text and changes in the manuscript in ***blue italic-bold texts***.

Answers to the comments from Anonymous Referee #1

In line 333, are the equation 3 taking into account the meteorological factors? If not, please evaluate the impact of meteorological data on the final results.

This equation was introduced by Weller et al. (2019) based on controlled release experiments. It uses only the mole fraction measurements (and in fact the observed maximum) to quantify leaks from individual mobile transects close to a gas leak. This equation does not include the meteorological information and it is based on statistical analysis. Von Fisher et al. (2017) stated that incorporation of meteorological information did not improve the emission rate estimate. It is acknowledged as a deficiency, still the same equation has been used in many different studies to derive comparable leak rate estimates.

Whether the large distribution of the maximum enhancement mentioned in lines 471-473 will affect the judgment of the threshold, and thus affect the results

The 10% threshold is a cutoff which excludes leaks smaller than ≈0.5 L min^{-1} emission rate. The rationale is that one should exclude transects in which the plume is almost “missed” due to unfavorable meteorological conditions. As we discuss in our manuscript, this threshold results in overestimation of the smaller leaks because then only the transects with the largest peaks are used for quantification, not the entire population. The 10% threshold has a minor impact on the bigger leaks.

Are the values of relative uncertainty mentioned in lines 479~486 too large and Whether they will affect the overall degree of confidence of the data

Uncertainties are indeed very large for individual passes. This has been investigated in detail by Luetschwager et al. (2021). According to their analysis the uncertainties in a quantification reduce to 10 % after about 8 transects.

What is the cause in lines 756-760 that the emission rates of the locations provided by the LDC were much lower than the locations detected by mobile measurements

These are the gas leak locations classified into the B and C categories. Indeed, emission from leaks in this category are lower compared to the A1 and A2 category. We do not know the causes, but we show that this can lead to a low bias of the gas leak emission rates reported in the German inventory. This is because leaks quantified with the suction method are most likely only from the B and C category, as the other leaks should be fixed either in one day or within a week.

About the two C_2H_6 signals mentioned in lines 789~794 that are not confirmed as the location of leakage by LDC, you suggest two reasons that they are related to the distant leakage and transmission, or surrounding emission sources. For the first reason, is it possible to compare the wind speed and direction when C_2H_6 signals are measured to find the location of leakage, For the latter reason, can you match the signal with the sources may produce both CH_4 and C_2H_6.

We have looked into that, but wind analysis is often not conclusive in cities. The wind conditions in the urban area are influenced significantly by the built-up environment (e.g., houses, trees) and traffic. Although the general wind direction can be determined if streets are aligned with the wind direction, this cannot be easily determined if this alignment does not exist. The determination of wind direction in narrower streets is also influenced by the street canyon effect.

References:

Luetschwager, E., von Fischer, J. C., Weller, Z. D., Characterizing detection probabilities of advanced mobile leak surveys: Implications for sampling effort and leak size estimation in natural gas distribution systems. Elementa: Science of the Anthropocene; 9 (1): 00143. https://doi.org/10.1525/elementa.2020.00143, 2021.

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, 4091-4099, https://doi.org/10.1021/acs.est.6b06095, 2017.

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: Mauder, M., PLoS One, 14, e0212287, https://doi.org/10.1371/journal.pone.0212287, 2019.

Answers to the comments from Anonymous Referee #2

The quantification of CH_4 emissions from gas leaks is highly important and an interesting dataset has been obtained using different methods. While parts of the manuscript are well written and also interesting conclusions have been drawn, other parts need major improvements. In my opinion, this paper can only be published after these improvements have been made and a reviewer (ideally from the inventory community) has seen the revised paper. The suggested improvements are mainly concerning the suction method and its implications on the inventories.

We thank the referee for the general comment. We had indeed discussed our results with experts who are directly involved in development of the emission inventory in Germany. We realized that there is lack of accessible information about the methods, assumptions and uncertainties. We reported the references based on our best knowledge and could not find clearer references. This is stated more clearly in the revised version of the manuscript.

1) Line 135-136: "The reported uncertainty range of this method is ± 10% based on 23 measurements in the 1990s (E.ON, personal communication, 2020)."

This is only an appropriate reference if there is no published data about the uncertainty of the suction method. In this case, it should be stated that to your knowledge there are no publications about the uncertainty of the suction method. Otherwise, an appropriate reference should be used.

Despite intensive effort, we have not been able to retrieve a published reference for this statement, and therefore we need to report it as personal communication.

2) Line 518-524: "At several of the locations where the mobile method had indicated high emission rates subsurface accumulation was widespread, and the suction method was either not deployed or the measurements were incomplete because of either safety reasons or because the suction team estimated that they would be unable to complete the measurements within a day."

I do not understand the requirement to complete the measurements within a day. If the aim of the study is to compare methods, the measurements should be completed even if it takes longer than a day. This constraint resulted in only one data point (fully completed measurement) for the suction method. In line 1002-1003 it is stated "While the mobile and tracer methods have been evaluated previously, this is the first peer-reviewed study that includes the suction method." I doubt that one can state this given just one completed measurement. It could be justified if the partially completed sampling and its comparison to the other methods is discussed in more detail

Due to logistical and financial reasons the campaign period was limited to two weeks. To assess many different leaks, it was decided to stop the measurements after one day and report an upper leak rate estimate instead of a full quantitative estimate. Based on conversation with the suction measurement team, the same practice applies when the suction method is used outside of this present study. We are of course fully aware of the consequences, and in our opinion, we discuss them clearly and openly. While we agree with the reviewer that "the measurements should be completed even if it takes longer than a day" to improve the accuracy of the suction measurements, the implementation in our study reflects the current practical reality and the consequences for reporting given the safety regulatory conditions. Regardless, we still consider this the first peer-reviewed study where the suction method was included, even when only one full quantification was derived; the upper estimates are also valuable.

Change in the statement:

While the mobile and tracer methods have been compared previously, this is the first peer-reviewed study that includes the suction method, although suction measurements could not be completed in one day at most locations.

3) Line 960-997 “4.4 Possible suction method sampling bias with implications for emission inventories”

After reading it the first time I became interested in the topic and started reading the referenced reports. After doing this, my conclusion is that 4.4. needs major revision.

You write in line 961-962 “The national inventory for CH_4 leakage from the gas distribution network in Germany is based on measurements with the suction technique (Umweltbundesamt, 2021).” I wanted to find out to which degree the inventory is based on the suction method and had a look at the referenced report. What I found is a reference to Gottwald et al., 2012 and Müller-Syring & Schutz, 2014. I had only a look at the latter reference and this only discusses theoretical emission estimates. It could be my fault that I missed the appropriate section in this extended report. However, it would be good to clarify what is meant with “based on” and it would be good to state where in this very extended report this can be found. (similar to what you do in line 966)

The statement is now rephrased as follows:

Following our communications with the emission inventory experts (personal communications with Christian Böttcher, 2022), we cannot fully reconstruct the methods that are used in the existing national inventory report to establish the emission factors due to lack of transparency. However, the German environmental agency (UBA) is considering to use the results of the recent large scale measurement campaign based on the suction method (MEEM, 2018) in future publications of the national emission inventory in Germany (Federal Environment Agency, 2021).

The next reference I was looking at was “MEEM 2022” (line 963). According to the reference section this is “MEEM, Analysing the Methods for Determination of Methane Emissions of the Gas Distribution Grid (2022). [online] Available from https://www.dbi-gut.de/emissions.html. (Last Accessed: 25 January 2022)” The link refers to a webpage of a project and the “MEEM Project - Phase I” has the title as given in the stated reference. I only found a “Management Summary GERG Project Phase I with the title Analysing the Methods for Determination of Methane Emissions of the Gas Distribution Grid” published in May 2016. The second phase under a different name has been completed in 2018. Therefore the reference to the “ongoing project”, which is “underway to refine these emission estimates” needs to be updated.

The statement referring to the ongoing project is now removed.

Line 965-966: “This implies that this method is not applied at locations of the A1 category, which demand immediate repair (P. 27 in GERG, 2018).” Though the statement is correct in Annex 5 of the given reference it is stated “Three measurement principles are considered for direct measurements on underground pipelines: Tracer Method, Suction Method, High Flow Sampler”. Though I see that the suction method could “have a location sampling bias towards leaks in the B and C category”, I am not convinced that it is not accounted for this bias in the inventories by using other methods for the A1 and A2 categories.

To the best of our knowledge, gas utilities do not quantify gas leaks themselves and the quantifications are performed by a third party. Due to logistics and time required for a gas leak emission quantification from the third party, the quantifications are performed most likely only at the B and C locations for which repairs are not urgent. For the leaks reported by public (recognizable by the odor), larger emission factors are used, however it is not clear how large these emission factors are. For the A1 and A2 leaks during this campaign, we could not smell the odor, and detection was only based on the signals (CH_4 and C_2H_6) from the instruments.

In summary: I think this section is interesting but needs to be re-written. The references of the reports should be revised (correct year, link,...), Also, the reference GERG, 2018 and GERG 2020 are the same. In addition, it needs to be clearer what the suction method is used for and if this has really implications for the emission inventories or not.

References have been updated (changed to MEEM (2018)), and to the best of our knowledge (and the best knowledge of the NIR compilers that we contacted), the suction method has been used so far which is described in our manuscript. We believe that the implications for emission inventories are thus a logical interpretation as stated in this paper.

References:

MEEM, Methane emission estimation method for the gas distribution grid, [Online], Available from: https://www.dbi-gut.de/emissions.html?file=files/PDFs/Emissionen/Final%20Report_MEEM%20DSO_end_signed.pdf&cid=5804, (Last Accessed: 12 December 2022), 2018.

Federal Environment Agency: National Inventory Report for the German Greenhouse Gas Inventory 1990–2019, available at: https://unfccc.int/documents/194930, (last access: 15 December 2022), 2021.

Answers to joint comments we received from DVGW (German Technical and Scientific Association for Gas and Water) and Gasnetz Hamburg.

Comment on line 1 to 2 (headline)

"Intercomparison of detection and quantification methods for methane emissions from the natural gas distribution network in Hamburg, Germany" The suction method is not a system for the detection of leaks. To check the pipelines and detect existing leaks, inspections are carried out in Germany in accordance with technical regulations (G465). This was not part of the method comparison.

The detection comparison refers to the comparison analysis of the carpet and the mobile method which we also performed in our manuscript.

Comment on line 24 to 26

"The quantitative intercomparison of the emission rates from the three methods at a small number of locations is challenging because of limitations of the different methods at different types of leak locations." The comparison of the measuring methods has been carried out with few measuring points. The main goal was the scientific exchange. For a representative statement on the sample type and sample size, the measurements would have had to be prepared and carried out consistently and reproducibly. A quantitative comparison is not possible on the basis of the planned 10 measurements.

Due to logistical and financial constraints, it was not possible to carry out a more extensive comparison. In fact, as we report, a direct comparison is not possible from our dataset. Nevertheless, our analysis still reveals systematic effects that are worth publishing, in addition to actually documenting the challenges and limitations of the different methods, where published information is particularly scarce, as noted by the referee. It is important to point towards a possible sampling bias exists for the suction method because of a regulatory/safety constraint to preclude a measurement. This regulatory/safety constraint will still be present even if the sample size increases.

Comment on line 33 to 39

"The suction method could not be completed or applied at locations with widespread subsurface CH_4 accumulation, or due to safety measures, and this sampling bias may be associated with a bias towards leak locations with low emission rates. The leak locations where the suction method could not be applied were the biggest emitters as confirmed by the emission rate quantifications using mobile and tracer methods and an engineering method based on leak's diameter, pipeline overpressure and depth at which the pipeline is buried." This is not correct. There is no limitation for measurements with the suction method. More time or equipment is needed for large gas accumulations. In other measurement campaigns, large gas accumulations were also measured using the suction method.

We acknowledge that there is no principle limitation to the suction method, and have added the statement that in principle measurements can also be continued for more than a day until an equilibrium is reached. Still, in practice, this is likely not done a lot (see our response to Reviewer 2).

We are interested in the remark that larger emission rates were also measured by the suction method. These data are not available to us, and in fact we had been informed previously by the operators of the suction equipment that such large leak rates were never observed by the suction method. In fact, this statement was one of the key motivations to carry out this intercomparison campaign as stated in our manuscript: “The discrepancy between these rather low leak rates compared to leak rates inferred with the mobile method calls for further investigation.” This underlines the lack of transparency of leak rate estimates by the suction method.

Comment on line 98 to 99
“Gas pipelines in a city with the scale of Hamburg are monitored every 5 years with the carpet method. The leak emission rate is not quantified and thus also not a parameter affecting the course of action ” Every 4 years according to national regulation for low pressure lines (HH 6.500 km) and medium pressure lines (HH 250 km). High pressure line monitored every year and additionally controlled by helicopter.
The sentence is now edited based on the information provided as follows:

There are 6,500 km of low pressure and 250 km of medium pressure gas pipelines in Hamburg which are monitored every 4 years with the carpet method based on the national regulations in Germany. Gas leaks in cities are not quantified and thus also not a parameter affecting the course of action. Moreover, high pressure pipelines are monitored on annual basis with additional helicopter-based measurement platform.

Comment on line 134 to 140
“Suction measurements normally find leak rates that are < 2 L min -1 (E.ON, personal communication, 2020). The reported uncertainty range of this method is ± 10% based on 2 measurements in the 1990s (E.ON, personal communication, 2020). The discrepancy between these rather low leak rates compared to leak rates inferred with the mobile method calls for further investigation, since the suction method is also employed to derive network-wide emission factors for the German co untry-wide gas distribution network (Federal Environment Agency, 2020). ”

See also line 111 to 115: There are stated that 10 % of the leaks are responsible for between 30 % to 70 % of the emissions. Therefore, the average value is not a contradiction. The emission factors from 1990 were updated by a large-scale national measurement program. Due to investments in the pipeline network (PE pipes, removal of gray cast iron, regular inspections, etc.), emissions have been greatly reduced since 1990. The updated emission factors confirm this.

Thank you for the interesting comment. Indeed, the very large leak rates are likely not normally distributed, and the average is of course smaller. Unfortunately, the underlying dataset is not publicly available, and it is not possible for us to compare the statistics directly. This would be an interesting future project. We still think that the above statement: “The discrepancy between these rather low leak rates compared to leak rates inferred with the mobile method calls for further investigation” is valid, in fact it was the driving force that initiated this collaborative project.

Comment on line 447
“Tabel 1 Results of gas leak quantification with different methods in Hamburg, Germany ” There is no big difference between the results.

We evaluate the differences in the text in detail. We acknowledge that the 1:1 comparison is not useful, and we discuss the reasons for different local settings. Part of this analysis is the comparison between leak rate categories, and we find that all three methods showed higher emission rates at A1 and A2 locations compared to the B and C locations.

Comments on line 518-522

"At several of the locations where the mobile method had indicated high emission rates, subsurface accumulation was widespread, and the suction method was either not deployed (n = 3; HH003, HH04, HH011) or the measurements were incomplete (n = 7; HH001, HH002, HH008, HH009, HH010, HH015 and HH101) because of either safety reasons or because the suction team estimated that they would be unable to complete the measurements within a day." For higher surface accumulatios the measurement with suction method is possible. It takes more time to pump out of the ground via injection ground lances surrounding the underground leaks until an equilibrium CH_4 mixing ratio is reached in air out flow. In only a few cases these measurements go beyond one working day. The suction method is the most accurate method in the comparison of the three systems. It is generally known that source-level measurement systems are more accurate than extrapolations from side-level measurements.

We think that the statement in the text is still correct, we state how it was handled in our campaign, and for these measurements the locations were flagged incomplete when not finished in one day. We do acknowledge (see above) that this is not a principle limitation.

In our campaign this happened at most locations, whereas the referee states that this happens only in a few cases. This may indirectly support the sampling bias, i.e. such locations are usually not investigated by the suction technique.

Comment on line 575 to 577

"Based on the previous experience at locations with widespread subsurface accumulation it was concluded that the suction method could not be applied at this location. The other case in this category was HH009. " See above – suction method can be applied. In this cases more time was needed. In this field trial the suction team has scheduled only 8 days for 10 measurements. It turns out that this was not enough.

Again, we acknowledge that this is not a principle limitation.

Comments on line 704 to 708

"Although the number of quantified leaks is limited, all the three methods show that the emission rates from category A1 and A2 leaks are higher than category B and C leaks . This indicates that the site selection bias of measurements for the suction method due to safety concerns (see qualifier above), can lead to a bias in the emission rate in this method. " The statement that it is not possible to measure AI or AII safety categorized leakages is not correct. These leaks can also be measured with simultaneous concentration measurement inside the building. This has nothing to do with the methodology. An investigation exclusively in urban areas was not representative. For general statements, different types of pipelines, leaks or environments would have to be considered.

We are grateful for the additional information on how/that AI or AII category leaks can in principle be measured when they are accompanied by simultaneous measurements inside the building. Given practical limitations we assume that these additional requirements again support our statement that they are not regularly measured with the suction technique so that a site selection bias indeed exists.

Comments on line 900-904

"Further research is needed to identify the physical mechanism(s) to explain the observed correlation between A1 and A2 leaks and high emission rates. As a hypothesis, the presence of soil cavities associated with leak category A1 may result in higher permeability, i.e. lower underground resistance, which then leads to higher emission rate for the same pipeline hole size compared to locations with no cavity. "

Correlations between A1 and A2 leaks as well as B and C leaks could not be formed due to the small number of measured values. Such a result would also be very surprising, as we remain of the opinion that a leak occurs accidentally. It is also random in terms of size and emission intensity, so it cannot be predicted. To explain the categories: The categorization was developed in DE in order to standardize a reaction time based on the distance of a leak to a living area. In the case of A1 and A2 damage, we react immediately because personal protection has absolute priority in this case. This is also prior to proving the measurement accuracy of the source-level measurement method, which is higher than side-level measurement methods. An emission rate depends on the leak size and the soil permeability for natural gas. In contrast, the soil permeability of natural gas varies and cannot be predicted because very variable soils and soil densities can be found. For a research based statement on this, serial examinations according to a standardized procedure are necessary. In particular, differences in emissions in countries with predominantly sandy soils compared to countries with predominantly loamy soils would be easily explained.

We clearly acknowledge in the manuscript that the statistical basis is small, but the differences are seen in all three methods, even when they did not quantify the same locations and when they have different systematic and random errors.

We understand the definition of the categories and are grateful for the clear explanation provided here again. One possible process that we have mentioned to explain such a difference between the categories that the presence of cavities itself may induce larger leak rates. I.e. when a leak occurs (spontaneously) close to a cavity the resistance to the atmosphere is much lower than when it is in solid soil, so the emission rate could increase. This is in addition to the influence of soils as mentioned in the comment, and possible other factors (the presence of buildings and cavities may have an effect on the stability of the soil.

Comments on line 1033 to 1034

"Our results therefore stipulate that representative site selection includes sampling at all leak safety categories (GERG, 2020). Otherwise, this could lead to a sampling and emission rate bias in the national inventory of gas leak CH4 emission in Germany."

It is generally known that source-level measurement systems are more accurate than extrapolations from side-level measurements. The comparison of side-level measurements with source-level measurement results usually serves to calibrate the less accurate side-level measurement. Since only one source-level measurement system was used in the method comparison, this is a very limited comparison.

We acknowledge the larger errors for the site -level measurements, especially from the mobile method. Nevertheless, the difference between the categories was observed for all three methods, including the (incomplete) suction measurements. We state repeatedly that the statistical basis for comparison is limited. The upper limits from the suction measurements after one day should already partially distinguish between larger and smaller leak rates.