In this document we provide our answers to joint comments we received 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**". We thank both organizations for the comments which resulted in improvements in the manuscript. Please find our answers in blue texts and changes we made in the manuscript in ***blue italic-bold texts***.

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 CH4 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.