Authors' answers to RC1's comments

This paper describes how metal oxide sensors can be used quantify the methane concentration in air. These concentrations are used to calculate an emission and the calculated emission rate compared against known rate to generate some idea of uncertainty in emission location and quantification.

We thank the reviewer for this general assessment and for his/her careful and insightful review of our manuscript which helped improve the presentation of our manuscript. We have carefully considered all the comments and revised the manuscript accordingly. Below, we provide our responses to all of the reviewer's comments.

General comments
The manuscript is poorly written and difficult to read in some parts (for example, the title is quite odd).

We have improved the writing of this manuscript as highlighted below. In addition we have modified the title. The new title is:

"Using metal oxide gas sensors to estimate the emission rates and locations of methane leaks in an industrial site: assessment with controlled methane releases"

The concept isn't very novel and several studies have presented the conversion of voltages output by Figaro MOS sensors into mixing ratios.

Indeed, the conversion of these voltages into mixing ratios is not the novelty of this publication. The specific conversion used here is adapted from Rivera et al. (2023).

The novelty is the methods by which the emission location and the emission rate are calculated.

The atmospheric dispersion inverse modelling method to derive emission locations and rates is not the novelty of this publication either: it has been documented in detail in the publication Kumar et al. (2022).

Details on both processes is lacking and it is currently unclear how novel the approaches used here are.

Our recent publications explicitly refer to Rivera et al. (2023) and Kumar et al. (2022) for more details on the overall rationale, strategies, concepts, and implementations of these two processes. We have modified the text to state it more clearly.

The two major novelties of this new publication are
   1) the application of the approach from Rivera et al. (2023) to on-site measurements during a real blind test of controlled releases with sub-optimal constraints that we did not manage
   2) a) the combination of the two processes (derivation of mixing ratios from Figaro voltages, and atmospheric dispersion inverse modelling to derive emission location and rates from these mixing ratios) b) with a key underlying question: is the accuracy of the mixing ratio

derivation satisfying for the purpose of the atmospheric dispersion inverse modeling? In previous publications, we had to derive notional targets for the accuracy/precision of the mixing ratio derivation, but the actual signal exploited by the atmospheric inversions from the timeseries of mixing ratios (through a complex binning of the mixing ratio data in practice, and then through the minimization of the model-observation misfits balanced by the modeling errors in the inversion process), and thus the requirements in terms of mixing ratio precision/accuracy are difficult to anticipate.

Also, this work would benefit from a deeper dive into the applicability of the approach in the real world.
Similar comments were already raised during the preliminary review of this manuscript, and some text had already been added in the introduction to clarify these points. We have the feeling that it is sufficient.

Specific comments
- **Abstract**
  - **Language needs to be improved throughout.**
  - **There are no details of the tests here: flow rates, meteorological conditions, time of year, how long the test were run.**
  - **Too much extraneous detail between lines 7 and 14. Should be condensed.**
  - **Last sentence (L14 to 16) doesn't make sense.**

We have updated the abstract to address the points highlighted by the reviewer. Here is the new version of the abstract:

"Fugitive methane (CH_4) emissions occur in the whole chain of oil and gas production, including from extraction, transportation, storage, and distribution. Such emissions are usually detected and quantified by conducting surveys as close as possible to the source location. However, these surveys are labour intensive, are costly, and fail to not provide continuous emissions monitoring. The deployment of permanent sensor networks in the vicinity of industrial CH_4 emitting facilities would overcome the limitations of surveys by providing accurate emission estimates, thanks to continuous sampling of emission plumes. Yet high precision instruments are too costly to deploy in such networks. Low-cost sensors using a metal oxide semiconductor (MOS) are presented as a cheap alternative for such deployments due to their compact dimensions and to their sensitivity to CH_4. In this study, we demonstrate the ability of two types of MOS sensors (TGS 2611-C00 and TGS 2611-E00) manufactured by Figaro® to reconstruct a CH_4 signal, as measured by a high-precision reference gas analyser, during a 7-day controlled release campaign conducted by TotalEnergies in autumn 2019 near Pau, France. We propose a baseline voltage correction linked to atmospheric CH_4 background variations per instrument based on an iterative comparison of neighbouring observations, i.e. data points. Two CH_4 mole fraction reconstruction models were compared: multilayer perceptron (MLP) and 2nd degree polynomial. Emission estimates were then computed using an inversion approach based on the adjoint of a Gaussian dispersion model. Despite obtaining emission estimates comparable with those obtained using high precision instruments (average emission rate error of 25% and average location error of 9.5 m), the application of these emission estimates is limited to

adequate environmental conditions. Emission estimates are also influenced by model errors in the inversion process."

- **Introduction**
  Throughout the introduction - It reads like a list of barely related bullet points. This needs to be rewritten and resubmitted for review. Currently, I find this very hard to comment on as there are too many inaccuracies and poorly written sentences. I have picked out some examples but this list is not exhaustive. For example:
  - **At L18: "Fossil fuel anthropogenic methane (CH4) emissions related to the production, exploitation and transport of coal, oil and natural gas, account for 35% of global anthropogenic emissions (Saunois et al., 2020)." Isn't all incorrect, only poorly written. It would be better written as: A recent study suggests that in 20XX methane (CH4) emissions from the production, transportation, storage and distribution of fossil fuels (e.g. coal, oil and natural) account for 35% of global anthropogenic CH4 emissions (Saunois et al., 2020). Then delete the next sentence.**

    We think that this is not an obvious example of a poorly written sentence. Sharing the statement from a publication and connecting this statement to the publication via simple parenthesis is very common in scientific publications to lighten the text, to avoid long series of "According to publication1…, Publication 2 indicates that….".

    However, we have accepted the correction and reformulated the phrase as suggested, and we have reviewed the whole manuscript to improve the writing.

    "*A recent study suggests that in the decade 2008-2017, methane (CH_4) emissions from the production, transportation, storage, and distribution of fossil fuels (e.g. coal, oil, and natural gas) accounted for 35% of global anthropogenic CH_4 emissions (Saunois et al., 2020).*"

  - **L26: "Atmospheric measurements" is too vague. Please describe what is being measured. "The downwind measurements CH4 concentration in the air" (or similar).**

    Added a clarification for "atmospheric measurements".

    "*Atmospheric measurements are increasingly used to detect and quantify CH_4 leaks from industrial facilities. These methods primarily involve measuring methane mole fraction downwind of the facility.*"

  - **L37: "limiting the impact of atmospheric transport modeling uncertainties". This is the key issue with networks of sensors.**

    We have rewritten the phrase:

    "*The use over long periods of time of dense networks of sensors deployed permanently increase the ability to identify the structures of the observed plumes, to improve the atmospheric transport modelling parametrization for the simulation of*

these plumes, and thus to improve the accuracy of the quantification of the leaks based on this modelling."

- **L38: "Chamberland and Veeravalli (2006)" this is an older study to be referencing. Also, the main issue with quantifying fugitive emissions on production sites is differentiating the vented and combustion emissions from fugitives.**

    Removed the phrase and reference from the manuscript.

- **L40: "dense network" another key issue with monitoring O&G productions sites. How many sensors? Typically, only four are deployed by a solution provider regardless of site size. Key question is how to optimize?**

    We have expanded the discussion section to cover the density of the sensor network. This discussion outlines our empirical approach in deploying a network of 16 tripods equipped with dual-sensor setups at strategic distances ranging from 40 to 50 metres from the emission source. Our findings suggest that while the industry standard might lean towards deploying only four sensors regardless of site size, our campaign demonstrates that a configuration of 4 to 5 sensors, strategically placed, can effectively minimise errors in methane emission quantification.

    We acknowledge that the question of optimising sensor count and placement is critical, especially considering varying site sizes and environmental conditions. Our analysis, particularly of cases with uniform emission rates (1 g CH4/s), indicates that a careful balance between sensor count and their strategic placement plays a pivotal role in achieving accurate emission estimations. However, we also highlight that determining the optimal configuration for a sensor network is complex and warrants a thorough investigation, potentially through simulations that explore typical emissions scenarios and the impact of adding or removing sensors. Given the scope of our study and the limited data points captured, a comprehensive exploration of optimal sensor network configuration was beyond our research boundaries.

- **L43: "metal oxide sensing material" not strictly true. The metal oxide doesn't sense anything, more physically changes in the presence of methane.**

    Rephrased to improve clarity. The new phrase is:
    "MOS sensors are composed of a metal oxide sensing material, incorporating an integrated heater. A chemical reaction affects the electrical conductivity of the sensing material in the presence of an electron donor gas such as CH_4 (Örnek and Karlik, 2012)"

- **L46: "suited to long time deployment" Again, not true. These sensors require relatively high amounts of energy (as you state in the following sentence), work poorly in low humidity and work poorly in low temperatures (as you state in the nest sentence).**

    Accepted correction. The content was reformulated, this is the new version of the text:

"However, MOS sensor sensitivity is also affected by other environmental parameters (including temperature and relative humidity) (Popoola et al., 2018); they also present low accuracy and may drift with time (in the form of a decrease in the conductivity of the sensing material), requiring periodic re-calibrations (Riddick et al., 2020; Shah et al., 2023, 2024), and the need for a constant power supply due to the heater material (Shah et al., 2023)."

■ **L62 – L71 Suggest you read: Riddick, S. N., Ancona, R., Cheptonui, F., Bell, C. S., Duggan, A., Bennett, K. E. and Zimmerle, D. J. (2022) A cautionary report of calculating methane emissions using low-cost fence-line sensors. Elementa: Science of the Anthropocene 10(1). <u>https://doi.org/10.1525/elementa.2022.00021</u>**

The study of Riddick et al. (2022) was added to the introduction. Here is the new version of the paragraph:

"The next logical step is to test the performance of the same sensors to reconstruct CH_4 mole fraction from real leaks, and to use reconstructed mole fractions to quantify and localise emission rates. Riddick et al. (2020) quantified CH_4 emissions from a gas terminal using a Figaro TGS 2600, deployed 1.5 km from the emission source. To reconstruct CH_4 mole fractions from voltage observations, the authors developed an empirical equation considering the measured voltage, temperature, and relative humidity from a nearby meteorological station and then applied a Gaussian plume model to quantify the emission rate using reconstructed CH_4 mole fractions and local wind information. Their estimated average emissions of 9.6 g CH_4 s^{-1}, with a maximum emission rate of 238 g CH_4 s^{-1}, given corresponding CH_4 mole fraction enhancements of between 2 ppm to 5.4 ppm within the plume. Their Figaro-based emission estimates were not confronted with corresponding emission estimates derived using a high precision gas analyser nor with an independent knowledge of the emission rate. Elsewhere, Riddick et al. (2022) studied the capabilities to detect and estimate CH_4 emissions of four Figaro TGS 2611-E00 sensors in a fence-line monitoring setup. Sensors were deployed closer to the emission source (30 m) and tested over a 48 hour period. Reported results showed detection consistency for emissions above 167 g CH_4 h^{-1} with an enhancement threshold of 2 ppm. However, the number of sensors used to compute the emission estimates was not specified, particularly given the spatial distribution of the sensors and varying wind speed."

■ **L78 – L109 Much of this is methods and should be condensed.**

The text was updated to present the context and contribution of the study more clearly.

"This study builds upon the research conducted by Rivera et al., (2022) and Kumar et al. (2022), demonstrating the potential for continuous monitoring of CH_4 emissions using cost-effective in situ sensors. Drawing from the insights derived from these two studies, it seeks to address the new challenges associated with the combination of both types of analysis, i.e reconstruction of CH_4 mole fractions from measured voltage variations and estimation of emission rates and location from CH_4 mole fractions. Firstly, the challenge arises in the deployment and management of onsite Figaro sensors, an issue not present in Rivera et al. (2022), as well as extracting CH_4 mole fractions from measurements that are impacted by more complex perturbations. For

instance, the background air in Rivera et al., (2021, 2022) was less polluted than air from an industrial site such as TADI. Moreover, the environmental conditions, especially in terms of temperature and water mole fraction, in these previous studies were smooth and not representative of field conditions as encountered in this new study. Secondly, the prescriptive precision and accuracy targets for CH_4 reconstructions outlined in Rivera et al. (2022) were established as generic targets, fitting for a variety of data processing strategies intended to quantify emissions from industrial sites. The specific observation and modelling strategy implemented in Kumar et al., (2022) to localise and quantify point source emissions carries its own set of precision/accuracy requirements. In particular, this strategy strongly relies on the characterisation of gradients across measurement stations of mole fraction averages over time or wind sectors, which makes the derivation of nominal requirements on the reconstruction of CH_4 spikes or the CH_4 time series quite complex. Furthermore, such requirements should be weighed against the modelling uncertainties associated with the corresponding Gaussian plume model inversions. Ideally, the uncertainties related with CH_4 mole fraction data would not significantly contribute towards the total uncertainty when combined with uncertainties from the modelling framework. This, however, does not necessarily mean that they should be much smaller than the latter. The direct comparison of the results obtained in this study with CH_4 mole fraction data derived from Figaro sensors and those from Kumar et al., (2022) provides insights into whether this objective is achieved.

Therefore, for 33 controlled releases at the TADI facility, we employed fixed-point measurements from both high precision CRDS instruments and low-cost TGSs. A considerable fraction of the TGS measurements were used for training models to reconstruct CH_4 mole fractions from measured TGS resistance and other variables. When reconstructing CH_4 mole fractions, we proposed a minimum accuracy target of 15% the amplitude of the largest observed mole fraction enhancement within a release. This corresponds to accuracies from 0.3 ppm for a release causing a maximum enhancement of 2.4 ppm up to 18 ppm for a maximum enhancement of 120 ppm. This accuracy is consistent with the accuracy requirement imposed in our previous study where we used TGS sensors to reconstruct CH_4 spikes created in a laboratory experiment (Rivera Martinez et al., 2022). However, the relevance of this target is implicitly re-assessed through the use of the reconstructed time series in the inversion scheme from Kumar et al. (2022)."

● **Methods**

The section was reorganised to present the methods in a more coherent and logical manner. In the current methods section we first introduce the TADI 2019 campaign (Section 2.1), then we present the Controlled releases and the sampling configuration (Section 2.2). The low cost logger system and the meteorological data is presented in Section 2.3. On Section 2.4 we talk about the preprocessing steps of the TGS data. Section 2.5 describes the reconstruction of Ch4 mole fraction from TGS voltage data and the metric used to evaluate the performance of reconstruction models. The selection of the training and testing sets are presented in section 2.6. Finally, Section 2.7 presents the inverse modelling framework used to estimate the release rates and locations.

○ **L112 – L128: Please condense this section to a description of the site. Please remove any extra detail superfluous to this study.**

The description of the site was updated, the corrected text is shown below:

"*In October 2019, TotalEnergies® performed multiple controlled releases at the TotalEnergies Anomaly Detection Initiative (TADI) facility, to investigate the capability of different detection and quantification techniques of CH_4 emissions from industrial facilities. The TADI test site is located northwest of Pau, France, with an approximate area of 200 m^2. It is equipped with infrastructure typical of oil and gas facilities (pipes, valves, tanks, etc) to simulate 'realistic' leaks. The terrain is flat but includes different obstacles that can affect the dispersion of the gases released to the atmosphere. Our experiment consisted of 41 controlled releases of CH_4 and CO_2, covering a wide range of emission rates of between 0.15 and 150 g CH_4 s^{-1}, with durations ranging between 25 to 75 minutes. We participated in this experiment to develop and test inverse modelling frameworks within the TRAcking Carbon Emissions (TRACE, https://trace.lsce.ipsl.fr/) program for the estimation of emission location and rates based on CH_4 mole fractions from high precision instruments (Kumar et al., 2022). We presented the inversion results for 26 releases from single point sources based on two inversion approaches, one relying on fixed-point measurements, and the other one on mobile near-surface measurements (the latter had already been documented in Kumar et al. (2021)). In both cases, the emission estimates relied on CH_4 mole fractions from high precision instruments, and on a Gaussian plume model to simulate the local atmospheric dispersion of CH_4. The results from Kumar et al. (2022) for point source emissions yielded an emission rate error of between ~23 to ~30 % and a localisation error (within a 40 m × 50 m area) of between 8 and 10 m. The controlled releases were emitted from heights of between 0.1 m and 6 m above ground level, and inside the 40 m × 50 m ATEX (ATmospheres EXplosibles) zone of the TADI facility (see Fig. 1).*"

○ **L129: Which heights? What is the ATEX zone?**

The heights and the ATEX acronym meaning were added to the manuscript. Here is the new text:

"*The controlled releases were emitted from heights of between 0.1 m and 6 m above ground level, and inside the 40 m × 50 m ATEX (ATmospheres EXplosibles) zone of the TADI facility (see Fig. 1).*"

○ **L131: I would like the information here instead of having to find a different paper.**

We acknowledge the remark and we have removed the phrase. The relevant information of the TADI test site for this study is included in section 2.1.

○ **L136: I have concerns over air drawn through 100m of tubing at 6 lpm. I assume transit time will be significant as will gas mixing in the tubing.**

The transit time for air through a 10 m length of 1/4" OD Synflex tubing is approximately 1.5 seconds, and for a 100 m length, it extends to approximately 14.5 seconds. We have empirically determined these transit times on-site by introducing a short exhalation of

breath into the inlet tubing and subsequently measuring the time taken to detect the resultant CO2/H2O spike. The precise timing of these spikes was captured using GPS timestamps, ensuring high accuracy in our measurements.

Regarding the potential impact of gas mixing within the tubing, with the Reynolds number around 1970. This indicates that the flow regime within the tubing remains laminar, significantly limiting the extent of mixing. Consequently, our findings suggest that gas mixing within the tubing is minimal and should not adversely affect the measurement of controlled methane releases. In addition, we have aligned our data to correct any time shift that might arise due to transit delays, thereby preserving the accuracy of our emission estimates.

○ **Tables 1 & 2 aren't adding much. Could go to SI**

Tables were moved to SI.

○ **L150 – L158: How were the supply voltage controlled? Also, sensitivity would be affected by the ADC used, can you comment on that?**

The power supply to our sensor system was fixed at a constant 5V throughout the duration of the experiment.

Regarding the analog-to-digital converter (ADC) and its impact on our measurements, it is important to clarify that while the ADC plays a critical role in defining the resolution and accuracy of the voltage readings, it does not directly affect the sensitivity of the sensor itself. The sensitivity of the sensor remains determined by its inherent properties and the environmental conditions under which it operates. The ADC's function in our setup was to accurately translate the analog signals from the sensor into digital data for analysis. Any potential disparities between the measured voltage and the actual voltage due to the ADC's characteristics were consistent throughout the experiment. This consistency ensures that while the ADC's resolution and accuracy are crucial for precise data capture, they do not alter the sensor's sensitivity to methane concentrations.

○ **Section 2.3 – Please filter this for unnecessary information. Why mention data that were not used?**

The text now belongs to section 2.2. The following text that mentions the controlled releases was updated to include only information relevant to the study and explain our motivation to use only a subset of 33 releases from the original 41 conducted during the campaign:

"A total of 41 controlled releases were conducted over a seven day period, between 2 October 2019 and 10 October 2019. Six releases corresponding to low wind speeds (< 0.6 $m\ s^{-1}$) were not used for the inversion as in Kumar et al. (2022), since measurements made in low winds are not suitable for atmospheric inverse modelling. They could however be used in the training of CH_4 mole fraction reconstruction models. The two largest releases produced high CH_4 mole fraction plumes that affected the amplitude measured by the TGS sensors, such that it was not possible to distinguish large CH_4 spikes from medium and

small spikes from voltage drop measurements (see fig A3) and, hence, they were also removed. Therefore, our study is focused on 33 from the initial 41 controlled releases conducted during the campaign. Table A1 details the releases that were measured by each chamber."

○ **L173 – Not being able to detect < 4 ppm is concerning.**

The challenge of not being able to detect methane concentrations below 4 ppm arises from a combination of factors that influence the sensitivity and accuracy of our sensor system. Firstly, the presence of a strong baseline signal at the Lacq site, caused by variations in relative humidity (RH), significantly impacts the sensor's ability to detect low methane concentrations. Additionally, discrepancies in the resistances of some loggers, which used 50 kΩ resistors instead of 5 kΩ, could have altered the sensitivity of the voltage measurements. These factors collectively contribute to the observed limitation in detecting methane concentrations below 4 ppm on some of the releases.

○ **L177 – Why 1-minute averages? Seems a long time. Was the sonic near anything else?**

Although the 3-D sonic anemometer measured variables at high frequency (20 Hz), due to logistical constraints on field, we would only have had access to the 1-min average data. However, our atmospheric inverse modelling framework relies on the process of average winds, especially since we use a Gaussian plume model to simulate the local dispersion. Therefore, there was no incentive to ensure that we could have access to higher frequency data.

Using data collected at shorter temporal scales, we commonly compute averages of meteorological variables such as wind speed and wind direction over longer durations or in specific wind sectors. However, computations of turbulent fluxes to use in the dispersion model to parameterize the diffusion parameters generally require high-frequency data.

○ **Table 3 -Why were these emission rates chosen? 1.4 to 18 kg h-1 are quite high emissions, what would be typical emission rates on a production site? What have other people measured as fugitives?**

The emission rates were chosen by the operator, TotalEnergies, to encompass the typical range of emissions encountered in production sites. This decision was based on the objective to ensure that our measurements and subsequent analyses would be representative of real-world conditions, as well as to provide insights into the scale of emissions that could be expected under various operational scenarios.

○ **Section 2.4 – This is particularly badly written. I am finding it difficult to follow. It also has too much information that could be put in the SI. Essentially TGS data were fitted to the methane analyzer data.**

This section, now in section 2.5, was updated to include only the relevant information to our study.

○ **L197 – Can you described the strengths and shortcomings of the two approaches. The main issue with approach 2 is the TGS sensors give erroneous output in low**

humidity conditions. This could strongly affect your linear interpolation if properly considered. At the very least, I would have thought the data were filtered for low RH.

Regarding the first approach, using H_2O and T data for correction, offers direct consideration of environmental variables affecting sensor accuracy. Its strength lies in its context-sensitive adjustment, aligning sensor data closely with real-time environmental conditions. However, its primary shortcoming in our study was the inconsistent availability of H_2O and T data due to logging issues, which limited its comprehensive application.

Respecting the second approach, involving linear interpolation to establish a baseline from non-peak voltages, provides a practical solution when environmental data are lacking. This method's strength is its independence from environmental data, making it applicable even with incomplete datasets. However, its shortcoming, as pointed out, is its vulnerability to erroneous outputs in low humidity conditions. TGS sensors' performance can be significantly skewed in such environments, potentially affecting the accuracy of the interpolated baseline. Implementing a filter for low RH would have significantly reduced the amount of usable data in our study, limiting the number of available releases for analysis

○ **Section 2.6 – Again, should be condensed.**

The text was rewritten to improve clarity:

"Defining an appropriate training data set is important to allow reconstruction models to derive sufficient information to generalise (i.e. to extend its performance to data not present in the training set) and to obtain good mole fraction reconstruction performances in the testing data set. In addition, the testing data set should be chosen to allow for the evaluation of model performances under a wide variety of conditions. Regarding inverse modelling, in order to provide a meaningful assessment of the estimation of emission rates and locations, inversion should be conducted using reconstructed CH_4 mole fractions from outside the training data set to avoid introducing bias in the evaluations of errors. Furthermore, depending on the magnitude of emission release rates, atmospheric turbulence, and locations/distances of the downwind active tripods from the emission sources, each of the six chambers could not be used to detect CH_4 mole fractions during each release; therefore a separate training and test set needs to be defined for each chamber.

These previous considerations constrain the selection of the training and testing data sets from each chamber. The testing data set of the releases for use in inversions was defined based on two criteria: 1) releases which yield reconstructed CH_4 mole fractions from at least three chambers simultaneously, and 2) releases corresponding to the more favourable wind speed conditions ($U_r \geq 1.4$ m s^{-1}) for inversions. We determined seven releases that meet these considerations (#2, #9, #10, #25, #26, #32 and #33). Because this testing data set was of insufficient size for all of the chambers, we decided to increase it by using data from four more releases with low wind speed conditions ($0.65 <= U_r <= 1.31$) (#5, #12, #21 and #29). This selection led to a testing data set of 40% of the releases. All remaining data were used as a training data set. The reconstruction models are trained and tested only once per chamber, following the distribution of the releases from table 2."

○ **Section 2.7 – This needs to be unpacked and explained clearly. Is there a figure that could help the reader? From what is written here, you are using the mole fraction (as calculated by a TGS) to infer four unknown variables (x, y, z and emission rate). Surely you need more than one measurement to do this, how many? What assumptions does this approach use? How realistic is it to well pad emissions? Again, this is impenetrable and needs to be written succinctly. A figure also helps the reader.**

The text was rewritten to improve clarity:

"Derivation of release locations and emission rates relies on an atmospheric inversion framework developed and tested by Kumar et al. (2022), which is based on measurements from high precision instruments. The details of the approach and implementation for this atmospheric inversion framework are given in this publication and they are just briefly summarised here.

This framework processes averages of the timeseries of the observed CH_4 mole fraction enhancements above the background from the different stations: from either the high precision measurements or from the reconstruction of CH_4 mole fraction based on the TGS sensors that is detailed above. For each station, these averages correspond to temporal averages within 2 to 7 bins of the observations defined by sectors of wind directions that are of equal ranges during the release (see below the practical definition of the wind sectors).

The atmospheric inversion relies on the simulation of these average CH_4 mole fractions using a Gaussian plume model. It uses the adjoint of this Gaussian plume model to simulate the sensitivity of these average CH_4 mole fraction enhancements above the background at a measurement location to the emissions at all potential source locations. This computation implicitly informs about the numerical operator corresponding to the simulation with the Gaussian plume model of the average CH_4 mole fraction enhancements above the background per wind sector and station as a function of the source location and rate. For each release, based on these sensitivities, the optimal horizontal and vertical location and emission rate are derived as the horizontal and vertical location and rate minimising the root sum square (RSS) misfits between averages of the observed and simulated CH_4 mole fraction enhancements above the background. In practice, optimal release location and emission rates are identified simultaneously, looping on a finite but large ensemble of potential locations, using an analytical formulation of the problem to derive the optimal rate and corresponding RSS misfits for each potential location and then identifying the optimal location and rate providing the smallest RSS misfits. The 40 m × 50 m (horizontally) × 8 m (vertically) volume above the ATEX zone is discretized with a high resolution (1 m × 1 m horizontally and 0.5 m vertically) 3D grid to define a finite ensemble of potential locations. The inversion exploits the change of wind direction during a release at the different measurement locations and the corresponding variations and spatial gradients in average mole fractions between the different measurement locations intersected by the plumes to triangulate the release location. The amplitude of mole fraction enhancements directly constrains the release rate estimate.

The Gaussian model (in practice: its adjoint) is driven by averaged wind directions and averaged turbulence parameters derived from 3D sonic anemometer measurements, using

the same bins for these averages as for mole fractions. Those bins are derived during each release based on the analysis of 1-min averaged wind directions: they correspond to a partition of the lower to upper range of potential 1-min average wind directions, into wind sectors of equal width in terms of range of wind directions. The total number of bins during this initial partition is defined as the rounding integer of the division of the release duration (in min) by approximately 7 min. However, only bins gathering at least four 1-min averages are retained. The aim is that the mole fraction and meteorological averages are representative of a timescale that is long enough for use in or comparison to the Gaussian model. This explains why, depending on the releases, the number of bins ranges between 2 and 7.

In this work, we slightly revise the reference computations of release location and rate estimates based on mole fractions from high precision instruments from Kumar et al. (2022). Indeed, in order to compare the release location and rate estimates from such a reference and to the results based on the TGS sensors, we restrain the set of high precision observations that are used in the reference computation to the station and time corresponding to data availability from the TGS sensors."

○ **L282: What data were used in the Gaussian plume model? Please be explicit.**

See the answer to the previous comment and the update of the section 2.7

○ **How do you overcome the GP's inability to work at distances less than 100 m?**

Gaussian plume models have regularly been used to model the local atmospheric dispersion over such short distances, see (Venkatram et al., 2004, Korsakissok and Mallet, 2009; Sharan and Kumar 2009, Korbeń et al. 2022, etc.). In fact, Kumar et al. (2022) evaluated the Gaussian plume model with actual source parameters for this experimental campaign at such distances and found it to be suitable for simulations of averaged plumes at such short distances. Here, we simulate average plumes for the different wind sectors during each release, which strengthens the relevance of using such a model. Furthermore, the good accuracy of our results demonstrates that the model was suitable for such an atmospheric dispersion inverse modelling problem. We have added a sentence in section 4 to highlight the use of GP models over such short distances in a series of past studies.

○ **L290 – L293: I do not know what this means. Please rewrite.**

The text was rewritten. See answer for Comment on Section 2.7.

● **Results**
  ○ **The current information presented here could be condensed.**

The result section was updated improving clarity and condensing the information.

○ **I would really like to know how applicable this method would be in the real world. For example, a well pad does not only have emissions from a single fugitive point which remain active until someone fixes it. Need to consider and comment on:**
  ● **There could be multiple emissions at the same time: vented, combustion and fugitive. How would this approach distinguish between types? Emissions (all**

types) can be intermittent. How long does an emission have to run to be detected?

Added a paragraph on the discussion section:

The inversion approach applied here, described for a single emission source only as required by the controlled releases experiments, can be easily extended to estimate emissions from multiple sources (see Singh et al. 2012). However, estimating emissions from multiple sources may require a more denser network of sensors to constrain a larger number of parameters for all sources. However, significant uncertainties may arise in emission estimates when measurements are taken in very close proximity to the emission sources. Our methodology requires that emissions be sustained long enough to be captured within the sampling intervals. The principal limitation is the requirement for at least four 1-minute averages, restraining the detection of short-lived emissions. Another challenge lies in the detection of emission types, such as vented, combustion or fugitive emissions. This aspect, out of the scope of our study, would require a detailed study of the characteristics of each kind of emission requiring additional tools to distinguish the particularities of them.

● **Gas coming out could be very hot, how does this affect your approach? Are any of the Gaussian Plume assumption violated? How does the GP approach work with distances < 100m?**

The elevated temperature of the gas emission is indeed a crucial factor to consider. In the Gaussian plume model, we need to consider the plume rise for a hot gas to include the effective source height in the computations. As mentioned before, the Gaussian plume model has been applied for distances less than 100 m. In fact, recently, Korbeń et al. (2022) applied the Gaussian plume model approach for the quantification of methane emission rate from oil and gas wells in Romania using ground-based measurement techniques less than 100 m. However, as mentioned previously and discussed in the revised manuscript, there may be significant modelling uncertainties for the measurements taken in close proximity to the emission source.

● **Density of sensors – is it realistic to have 16 sensor locations on a site? What is the optimum for your approach and what is realistic?**

The following text was added to the discussion section and table 3 was updated including the number of sensors used to compute the emission estimates for each release:
"***Density of sensor network***

In our campaign, we deployed 7 chambers connected to air inlets placed on tripods at distances of between 40 and 50 m from the emission source to capture methane plumes under various conditions. Table 3 details the number of sensors used for emission flux estimations across the controlled releases.

The optimal number of sensors for emission flux localisation and estimation is complex, influenced by varying emission rates, environmental conditions, and setup configurations. Notably, when examining cases with uniform emission rates (1 g CH_4/s), such as releases 12, 2, and 21 (with 3, 4 and 5 chambers respectively), a configuration of 4 to 5 sensors consistently produced the lowest errors for both sensor types. Yet, release 21 demonstrated that even five sensors may not guarantee low errors if the plume capture is suboptimal due to environmental factors or sensor placement.

We can contrast our setup with Riddick et al. (2022), who used four sensors approximately 30 m away from the source, but without detailing their individual contributions to emission calculations.

The optimal configuration of such a relatively dense network necessitates a thorough investigation, possibly through simulations of typical emissions and the strategic addition or removal of sensors to assess their impact. However, a comprehensive analysis of optimal network configuration was beyond the scope of our study due to the limited number of data points recorded."

- **Computational time. How long does it take to calculate the RSS matrix? Is this realistic?**

  The following text was added to the discussion section. Table A7 was added to SI.

  "*Computational efficiency of the inversion framework*

  Our inversion framework, developed in Python 3.8 utilising numpy, pandas, and scipy libraries, efficiently computes the RSS matrix through vectorized operations and a nested for-loop. This approach achieves an average computation time per release of 0.1 seconds for the RSS matrix and 1.46 seconds for full code execution, including data preprocessing on an 8-core Apple Silicon M1 processor. The framework, which can be further optimised with multiprocessing, is detailed in Table A7, showcasing computational times across different releases. It effectively estimates emission rates and source locations on a fine grid (40m x 50m x 8m, discretized at 1m x 1m x 0.5m), demonstrating practicality for real-world applications at minimal computational costs."

- **Technical comments**
    - **Currently, the manuscript is difficult to follow and I strongly suggest a comprehensive rewrite.**

  We have undertaken a comprehensive revision of our manuscript to enhance its clarity and coherence. This revision includes significant improvements to the grammar and style to ensure the manuscript is accessible and easy to follow.

References

Korsakissok, I., and V. Mallet, 2009: Comparative Study of Gaussian Dispersion Formulas within the Polyphemus Platform: Evaluation with Prairie Grass and Kincaid Experiments. J. Appl. Meteor. Climatol., 48, 2459–2473, https://doi.org/10.1175/2009JAMC2160.1

Sharan, M., & Kumar, P. (2009). An analytical model for crosswind integrated concentrations released from a continuous source in a finite atmospheric boundary layer. Atmospheric Environment, 43(14), 2268-2277.

Singh, S.K., Sharan, M. & Issartel, JP. Inverse Modelling for Identification of Multiple-Point Releases from Atmospheric Concentration Measurements. Boundary-Layer Meteorol 146, 277–295 (2013). https://doi.org/10.1007/s10546-012-9765-y

Venkatram, Akula, Vlad Isakov, Jing Yuan, and David Pankratz. "Modeling dispersion at distances of meters from urban sources." Atmospheric Environment 38, no. 28 (2004): 4633-4641.

Korbeń, Piotr, Paweł Jagoda, Hossein Maazallahi, Johannes Kammerer, Jarosław M. Nęcki, Julia B. Wietzel, Jakub Bartyzel et al. "Quantification of methane emission rate from oil and gas wells in Romania using ground-based measurement techniques." Elementa (2022).