<u>Author responses to reviewer comments on "First validation of high-resolution satellite-derived methane emissions from an active gas leak in the UK" by Dowd et al.</u>

We would like to thank the reviewers for their time and feedback to help make improvements to our manuscript. The author responses are highlighted in blue and the line numbers respond to the updated manuscript.

Referee 1

Inverse anthropogenic methane emissions from large point source are challenging but significant to manage fugitive emissions and design of control strategies. The study by Dowd et al. estimates methane fluxes from a large pipeline gas leak based on high-resolution satellite and ground-based mobile observations. The authors use and compare three different models to estimate methane emissions from pipeline leaks, assess the leak's contribution to observed concentrations at a downwind tall tower site, and further successfully help the local utility company solve the leakage problem, which is of great application value. Nevertheless, from a research article perspective, there are certain aspects that require careful re-evaluation, such as the comparison and uncertainty of different top-down methods. Additionally, the overall structure of the paper could be improved. Only after appropriately addressing the following issues can I recommend this paper for publication in AMT.

We would like to thank Reviewer 1 for their comments, which have helped to improve our paper and clarify the uncertainty on our results. We hope that we have addressed the reviewer's concerns appropriately.

Comments to introduction

The introduction section only consists of two paragraphs, which may not have a strong sense of logic and hierarchy, and there are too many words in each paragraph. The author can rewrite it as 3-4 paragraphs. And I suggest the section should supplement relevant background of previous methods for inverting point-source emissions, and the comparison of different methods.

We have restructured and updated the introduction taking into account reviewer suggestions, improving the logic and hierarchy of information. We have included information on different flux estimation methods for both satellite and surface-based observations. It describes these topics in order:

- Introduction to methane and anthropogenic emissions globally and in the UK.
- Summary of how fugitive emissions from gas transportation are estimated in the UK.
- UK does not have a methane monitoring system for fugitive emissions, so we describe a current system provided by MARS which uses different types of satellite data to aid mitigation.
- Description of GHGSat and its ability to detect and quantify smaller emissions than satellites used in MARS.
- Brief summary of different flux estimation methods for satellite data and method best suited for the GHGSat constellation.
- Introduction to the Cheltenham gas leak and validation using a ground-based mobile survey.
- Description of the two main methods to estimate fluxes from ground-based mobile surveys.
- Introduction to methods used to quantify Cheltenham gas leak and the tall tower analysis.

We have also included in the Methods section (Section 2.3) more details on the different flux estimation methods and reasons why we use the IME and Gaussian Plume Methods.

The satellite-derived flux methods start on line 194 and the ground-based mobile survey starts on line 260.

Comments to estimation methods

1. In the flux estimation method, the authors do not provide a clear description of the calculation process and model parameters, and how the observation data are used in this method. I suggest the author should present specific calculation formulas and parameters explanation. Besides, how the source location is determined by the GEOS-FP wind direction? I find this sentence confusing. What is the resolution of the GEOS-FP dataset, and I suggest supplementing the relevant information.

   We have improved the description of the IME method in Section 2.3.1 to include formulae and further description of parameters used. We have also included the differences between flux estimation methods and reasoning behind why we think that the IME method is the best for estimating fluxes from GHGSat observations.

   The source location determined by GHGSat is not solely derived from the wind direction. The GHGSat operator uses the shape of the plume (plume tail direction) and area most concentrated at the beginning of the plume tail to identify the location. When the plume and concentration gradient in the plume does not show a typical directional plume-shape, we can use the wind direction from Goddard Earth Observing System Forward Processing (GEOS-FP) to determine which side of the emission is more likely to correspond to the source location.

   We have moved the location estimation method to Section 2.1, line 100: "A GHGSat operator determines the source location by the shape of the plume (plume tail direction) and the area most concentrated at the beginning of the plume tail. When the plume shape and concentration gradient in the plume does not show a traditional directional plume-shape, the wind direction from Goddard Earth Observing System Forward Processing (GEOS-FP, NASA GMAO, 2023) is used to determine which side of the emission will most likely correspond to the source location."

2. In the Gaussian plume model, the authors use a mean of the source locations provided by the satellite retrievals as the source location, which is different from the assumption of the above flux method. What is the formula for calculating the diffusion coefficient and atmospheric stability classification? What is the unit of the modeled downwind concentration C, is it consistent with the observation. If not, how converted? Is the height of control surface related to atmospheric stability?

   To address a number of questions and comments regarding the Gaussian Plume Model set up and uncertainties made by Reviewers 1 and 2, we have made some changes on how we estimate the fluxes from the surface-based mobile survey.

   The Gaussian Plume Model used for the analysis of the vehicle-based measurement data to determine methane fluxes has been redeveloped to perform as a Monte-Carlo simulation. This will give a more robust result regarding how the uncertainties are propagated through the flux calculation and also increase the reliability of the overall results. The measurements and input parameters are allowed flexibility to interplay and thus should produce a more realistic overall uncertainty in the final calculation.

The following parameters are now assigned according to:

Leak Location: The location is randomly assigned to any point within an approximate 10 m × 10 m box which was used as the release point in the NAME model and identified as the most likely point of emission. The box is bounded by longitude and latitude coordinates: -2.099682, 51.9506799; -2.0997039, 51.9506635; -2.0997021, 51.9506682; -2.0997023, 51.9506718.

The five source locations determined by the satellite observations, from the above method (described in response to Reviewer 1, Main Comment 1), do not vary much apart from one observation which is situated on the other side of the road. We took a mean of the four satellite derived locations (excluding the outlier) as a best estimate of the location for the NAME model and a centre point for the Gaussian Plume bounding box.

Wind speed: The wind speed for each individual plume is determined using the mean wind speed during the traverse +/- a random uncertainty with a mean of 1 m s^{-1} which is allowed to vary according to a Gaussian distribution. As the wind speeds are on the order of a few m s^{-1}, this may be an over-estimation of the uncertainty, especially for the wind speeds as measured on the vehicle at 1 Hz resolution. The sonic anemometer was used on 26th May and UKV wind data on 12th June as described on line 219.

Measurement height above ground: Unusually, the vehicle height varies due to the presence of a road bridge (over a railway line) during the interception of the plume. Taking into account the difference in height between ground height and inlet height results in an expected range of 2.5m to 6.5m. As the model levels are 1m in vertical thickness, the central model inlet height is uniformly randomly allowed to vary between 2.5 and 6.5m inclusively in discrete 1m intervals for each traverse. The addition of this uncertainty into the simulation has had the effect of reducing the overall estimated flux.

The presence of the bridge causes a difficulty in assessing the appropriate height above ground to use as some of the plume will flow under the bridge, and there will also be the impact of the bridge structure on the plume flow which cannot be captured in the simple modelling conducted here.

The atmospheric stability classification is based on the standard Pasquill classification using meteorological conditions (Ref Pasquill, 1961) along with diffusion parameters calculated using Briggs (1973). The dispersion parameters σ_y and σ_z are given as functions of downwind distance (x) and stability class:

General form of parameterization of plume width parameters according to Briggs (1973):

$$\sigma_y = \frac{\alpha x}{\sqrt{1+\beta x}} \text{ and } \sigma_z = \alpha x(1+\beta x)^{\gamma}$$

Where constants are defined according to Pasquill's atmospheric stability classification (e.g. see Briggs 1973).

We now automatically select atmospheric stability classification based on the wind speed selected by each Monte-Carlo run (as described above) and manually input sky conditions. As the sky conditions did not significantly vary during the period of measurement on either day, this is effectively variable on the wind speed only – see below:

	Daytime insolation			**Night-time conditions**	
Surface wind speed (m s^{-1})	**Strong**	**Moderate**	**Slight**	**Thin overcast or > 4/8 low cloud**	**<= 4/8 cloudiness**
< 2	A	A - B	B	E	F
2-3	A - B	B	C	E	F
3-5	B	B - C	C	D	E
5-6	C	C - D	D	D	D
> 6	C	D	D	D	D

The Monte-Carlo simulation was run 1,000 times for each suite of transects and gives the following outputs.

Results:

May 26th:

Average flux (kg hr^{-1}): 845.51

Uncertainty: ±452.83

June 12th:

Average flux (kg hr^{-1}): 634.35

Uncertainty: ±299.31

A description of the Monte-Carlo simulation has been added to Section 2.3.2.

We have added a sentence and above equation to line 227: "σ_y and σ_z are given as functions of downwind distance (*x*) and stability class. Equation x shows general form of parameterisation of plume width parameters according to Briggs (1973)."

The modelled downwind concentration, C, is in µg m^{-3} and the measurements in ppb are scaled to µg m^{-3} to match units using standard temperature and pressure (STP) conditions. We have added a sentence to line 231: "The observations are measured in parts per billion (ppb) and are scaled to µg m^{-3} at standard temperature and pressure (STP) conditions."

The height of the control surface is set to best match the height of the inlet on the vehicle. During standard operational conditions, this would be 1.5m agl to best match the height of the measurement from the vehicle inlet. The vertical resolution of the model is 1m and the inlet height is 1.8m (therefore the control surface between 1 and 2m is the most appropriate comparison). However, as we are dealing with an elevated position compared to the leak location with the plume able to flow through the bridge structure this significantly complicates the effective height above ground of the measurement. To best account for this we have allowed the control surface and effective height of the inlet to vary between 2 and 7m in the Monte-Carlo model set up.

We have added a line to 236: “The height of the control surface is allowed to vary between 2 and 7 m a.g.l to best match the height of the vehicle inlet above ground level when accounting for the effect of the bridge structure.”

3. Since the authors can obtain the UKV meteorology to drive the NAME dispersion model, why not use this high-resolution data to drive the other two models? If the author reconciles the wind fields in the three methods, it may be more persuasive to compare the difference of retrieved fluxes.

   GHGSat is a commercial company who provided us with emission estimates through the ESA Third Party Mission (TPM) programme. The emission estimates by GHGSat are a Level 4 data product provided through the TPM programme which uses the GEOS-FP wind information. It would not be a realistic assessment of GHGSat’s flux estimations if we use UKV winds with the integrated mass enhancement method because they would not normally have access to UKV data. During the ground-based mobile survey we utilised the sonic anemometer on board the mobile laboratory on 26th May to provide more accurate wind information than the UKV model. Unfortunately the instrument was unavailable during the survey on 12th June so the UKV model was used in this case (see line 219).

   In order to provide some continuity between the different observation and flux estimation methods we provided independent estimations of the gas leak using the NAME simulations (see Section 2.4). We have added a sentence on line 289 to clarify this: “We did this to provide a flux estimation for both observation methods, using the same model and meteorology to provide some continuity between the satellite and mobile survey-derived estimates.”

   We have also added the line to Section 4, line 506: “The GHGSat Level 4 data, provided through the ESA TPM programme, these data give estimates the flux of a source using GEOS-FP wind data as standard. GHGSat would not normally have access to the higher resolution UKV wind data.”

4. The section lacks a description of uncertainty for each method. What aspects of uncertainty were considered and how is the range calculated?

   For the GHGSat flux estimation we have updated the final line in Section 2.3.1, line 191: “The uncertainty on the source rate is the 1σ standard deviation based on the uncertainties on the wind speed, measurement uncertainties, and the IME model parameters, where the wind speed is the dominant source of uncertainty.”

   For the Gaussian Plume model we have done a Monte-Carlo simulation which varies various parameters (described in full in response to Reviewer 1, point 2) to provide a more robust estimate and uncertainty value derived from the ground-based mobile survey estimates.

   For the NAME-derived flux estimations we have varied the plume definitions, described on lines 296-303, Section 2.4, in order to provide a range of fluxes estimates from the model simulation. The UKV model does not provide an uncertainty on the wind speed values so we are unable to propagate the uncertainty of the wind speed into our flux estimate. Therefore, we describe the caveats of this method in the discussion (Section 4).

Additionally, I suggest the author can consider adding a flowchart framework in the section to clearly show the calculation process of the three methods.

Thank you for this suggestion, we have added two flow charts to the Supplement which shows the calculation process for the different flux estimation methods for each observation type. We reference this in the text on line 318.

More minor issues:

There are different fonts in Line 144, 149 and 225.

Thank you for noticing this. We have adjusted the font on these lines.

Line 289: What is "The LI-7810 data"?

Thank you for noticing this, we have corrected it to read Licor-7810. The Licor-7810 is an instrument described in Section 2.2.2.

The x-label in Figure 2 should be revised to "YYYY-MM-DD". How to plot the uncertainty range? I recommend supplementing the uncertainty estimation for each method in the method section.

We have revised Figure 2 to be "YYYY-MM-DD". The graph is plotted in Python using the flux values and their uncertainties listed in Table 1 under columns "Mobile Survey Flux (kg h^{-1})" and "GHGSat Flux (kg h^{-1})". We have added a sentence to the Figure caption to clarify: "The error bars represent the uncertainty on the flux estimates, as described in Section 2.3.1 and Section 2.3.2."

The uncertainty on the GHGSat-derived emission estimate is described in Section 2.3.1 and the new ground-based survey Monte-Carlo simulation with associated uncertainties is described in Section 2.3.2.

Line 327: Perhaps the author can assess how much uncertainty of the wind fields. Recommend discussing the consistency of results under the same wind fields.

The uncertainty on the GEOS-FP wind speeds is based on a comparison of GEOS-FP modelled 3-hour winds with measurements from U.S. airports in the University of Utah MesoWest Database (Horel et al., 2002, Varon et al., 2019). Varon et al. (2018) found relative error standard deviations of 15-50% for source rates between 0.5-2025 tons per h^{-1} with 10 m wind speeds 2-7 m s^{-1}, where larger errors correspond to lower wind speeds (see Supplement of Varon et al., 2019). We have referred to the Supplement of Varon et al., 2019 in Section 2.3.1.

There is not an uncertainty provided for the UKV wind data so we perturbed the plume criteria as stated on line 300 in Section 2.4. See our response to Reviewer 1, Main Comment 4.

See Reviewer 1's Main Comment 3 as to why we have not used UKV winds in the IME method and first ground-based survey estimate using the Gaussian Plume model.

Line 333-335: The Gaussian plume model assumes that the wind field is constant, which only uses wind vector of a single point, so I think the winds driving the NAME and Gaussian plume model are not fully consistent.

The Gaussian plume model takes average wind data for each transect when utilising the on-board anemometer (visit 1, 26th May), so is using an averaged wind speed for each transect where possible. When using the hourly UKV wind data due to unavailability of the anemometer data (visit two, 12th June), the wind data is selected to the nearest hour and therefore will be constant, or near constant, throughout the mobile measurements.

We have clarified this on line 219: "We then took the observed concentration data and wind speed data to create the initial model plume using Eq. 3. On 26th May we used a mean wind speed observed by the vehicle's 10Hz sonic anemometer for each transect, whereas on the 12th June we used wind speed data, averaged to the nearest hour, from the Met Office's UKV model due to the unavailability of the anemometer."

On line 402 we have also clarified that the Gaussian Plume Model and NAME used the same UKV wind speeds on the second visit: "The NAME-derived fluxes use the same wind speeds as the Gaussian Plume Model on 12th June so differences between the model and the mobile survey fluxes are likely due to differences in the peak location along the road and the model resolution."

It is confusing in the 3.3 section what "pollution event" refers to. Two events defined in line 355 and line 360 seem the same event, but their definitions and criteria are not identical. I would suggest keeping consistent with clear definition.

Apologies for the confusion here. On line 355 'a pollution event' should read 'a leak pollution event'. On line 421 we are describing the number of times two different criteria are met. The reasons for selecting these two criteria are described on line 423 and line 418. First we count the number of times the gas leak concentration was at least 2σ larger than the background concentration at RGL and then we also count the number of times the gas leak contributed to a 'leak pollution event' during the NAME simulation. A 'leak pollution event' is when simulated concentrations of the leak are at least two standard deviations (2σ, 14 ppb) larger than the observed background concentrations and contributing a significant percentage (>=90%) of the above-background concentrations.

We have defined a 'leak pollution event' as 'LPE' and have updated the text.

We have clarified this by editing line 421: "We calculated the number of times the gas leak concentrations at RGL met two different criteria. We first calculated the number of times the gas leak concentration was at least 2σ larger than the background concentration at RGL, i.e. when the leak's contribution was above the noise of the background concentrations. We also calculated number of times the gas leak contributed to a 'leak pollution event' (> 2σ and > 90% above-background) at RGL."

Line 453: The difference between the NAME and the Gaussian plume model is not only from the wind fields. The uncertainty of the Gaussian plume model also comes from the atmospheric stability parameter, the source location and other variables.

The other factors you suggest are described in the sentence on line 482.

Line546: What is the "$CH4_{Tall\ t}$" ?

Thank you for noticing this, this was a typo and it has been removed.

Notice the unit formats (quite a few instances for the full text). For example, "ms-1" in Line 179 should be "m s-1".

We have gone through and checked unit formatting throughout.

Referee 2

Referee Comments on First validation of high-resolution satellite-derived methane emissions from an active gas leak in the UK

Note – for simplicity the line numbers in this review relate to the online version, though this reviewer was provided with a revised version of the manuscript containing a missing section 2.2.3 containing a description of the tall tower measurement methodology.

General comments

This paper presents a series of measurements undertaken to confirm a methane emission identified during a satellite measurement targeted on another source. This is a useful and important study to show independent methods confirm the identified emission and to compare the emission rate estimates obtained by the different approaches. The study deployed ground mobile sampling (MS) based concentration measurements to provide confirmation of the emission location and emission rate. A comparison to emission retrievals using a NAME model based approach applied to both the mobile and satellite concentration data was also carried out to 'normalise' a model approach for both observational methods. As an additional study the impact of the emission on concentration levels observed at a tall tower was also assessed to determine if the tall tower would have detected the leak.

This paper is a valuable addition to the literature on methane emission detection and quantification.

We thank Reviewer 2 for their time and effort in reviewing our manuscript. We have addressed the comments below and think that these revisions have clarified and improved our paper.

I do have a slight concern over the title and the use of the term validation, which although it doesn't have a formal definition, does imply performance evaluation and independent assessment. There are three aspects of the satellite measurements that this paper covers; leak detection, the identification of leak location, and leak emission rate determination. As noted by the authors, this provides one of the first assessments of the satellite with a real world 'live' leak. With some additional information the paper could provide more evidence to assess and validate the satellite results.

We believe that the word validation is suitable in this case as we provide a performance evaluation of the satellite data using independent observations. Due to the limited visual clues in the area of the gas leak we have used the best information available for the leak location from the satellite data. To reduce the reliance on the satellite information we have randomly assigned the location of the gas leak within a 10m box (see response to Reviewer 1, Main Comment 2) and also perturbed the location in the NAME simulation to assess the impact on the estimated fluxes.

The paper also assesses whether the leak is detectable in the time series of concentration measurements at one of the UK DECC network tall tower stations at RGL. The is a very interesting addition to the study, however, it does not add much information to the validation of satellite derived measurements. The authors could consider whether this would be best separated into two papers – one addressing the satellite validation, and the other assessing the potential methods for detecting such a leak, including tall tower sites, satellites, and combined observation systems.

Thank you for this suggestion. We believe that the assessment of the UK DECC network and the measurements taken at Ridge Hill are related within the framework of this specific gas leak. A full assessment of potential methods for detecting leaks in the UK is out of the scope of our study. We believe that the overall manuscript is strengthened by the inclusion of the analysis of the tall tower observations at Ridge Hill. It highlights that the current observation network would not have spotted this leak, emphasising the importance of validating satellite data and having a multi-instrument and multi-scale approach to monitoring CH_4 emissions. We have emphasised this further in the revised text, line 596: “This highlights the importance of validating other observations methods, such as the GHGSat satellite constellation.”

Specific comments and suggestions

The following suggests some comments on leak detection, leak location and emission rate quantification. Addressing these in turn:

Leak detection

The performance claimed for the satellite is 50% probability of detection for a 100kg/h emission at wind speeds of 3m/s. The paper does not provide the meteorological information (e.g as a minimum the wind speed and direction) obtained from the different sources used by the different techniques, for the different days that the leak was measured. This should be included, at least in the SI.

The wind speed values were provided in the Supplement and can be found in Table S2. We have added the wind direction to the table.

Some indication of how sensitive the detection probability is to wind speed, leak rate, and other factors such as land cover would be informative and the expected detection threshold for the actual surface type (inhomogeneous farmland), and range of wind speeds during this study would be very helpful.

The sensitivity of the detection threshold has a linear relationship to wind speed. The detection threshold is ~100 kg h^{-1} at 3 m s^{-1} and ~200 kg h^{-1} at 6 m s^{-1}. We have included this in the text to clarify the instrument’s detection threshold and referenced the White Paper by McKeever and Jervis 2022. See line 465: “The GHGSat detection threshold has a linear relationship with wind speed and is 100 kg h^{-1} at 3 m s^{-1} and 200 kg h^{-1} at 6 m s^{-1} (see McKeever and Jervis, 2022). GHGSat has demonstrated it can detect down to 42 kg h^{-1} (McKeever and Jervis, 2022) and up to 79,000 kg h^{-1} (GHGSat, 2022).”

The GHGSat constellation has not undergone any studies on the detection threshold dependency on surface type. GHGSat has a high spectral resolution and is not affected by surface type as strongly as other satellites such as Sentinel 2 and Landsat. The column precision of GHGSat was tested across all their observations (with a mixture of surface types) and found a column precision of ~2% (MacLean et al 2023 and Jacob et al 2022). See line 468: “The high spectral resolution of GHGSat means that it is not affected by surface type as strongly as other satellites such as Sentinel 2 or Landsat. GHGSat has been tested across a mixture of surface types and found to have a column precision of ~2% (MacLean et al., 2023; Jacob et al., 2022).”

The wind speed data can be found in the Supplement, Table S2. We have added the wind direction.

Leak detection is clearly demonstrated, as the presence of a leak from a distribution pipeline was confirmed by the relevant utility company. The initial identified plume is plotted in Figure1. This and all subsequent satellite plots presented in the paper are roughly centred on the emission. The targeted landfill consists of a number of potential methane sources – a composting and household waste (plus closed landfill) (~1.2 km to the south of the leak) and an active area of landfill tipping (which at the time of the leak the active tipping area appears to have been 900m to the southeast of the emission) and other potential sources such as closed landfill area and a biogas plant. In order to understand the 'screening' or leak detection capability of the satellite it would be useful to show/describe what area was within the original targeted/ tasked region.

Therefore, it would be useful to add information on the following points:

- What was the full field and centre of the original target area of the tasked satellite survey – is that equivalent to the areas shown in Figure 1

  The field of view of the satellite is described in Section 2.2.1 and is 12 km × 12 km. The leak was detected in the northern section of the target centred at 51.9402°N, -2.0998°E and has been specified on line 346: "The satellite was centred on 51.9402°N, -2.0998°E with a 12 km × 12 km field-of-view."

  Figure 1 is not the full field-field-of view of the satellite. It shows the pixels within the field-of-view that were assessed by GHGSat to contain elevated methane concentrations from the leak. We clarified this in Figure 1 caption: "Total column CH_4 (ppb) observations from the GHGSat satellite showing the variation in strength and size of the plume from the gas leak on six dates between March and June 2023. The geographical area shown is not the full field-of-view of the satellite and contains only the area where enhanced CH_4 concentrations were identified."

- Was the satellite re-tasked/focussed to the leak area?

  The satellite was re-tasked to monitor the leak as described in Section 3.1. The satellite target was not changed from the observation because the gas leak remained in the 12 km × 12 km field-of-view. We have edited the line 349: "After detecting the leak, GHGSat continued to monitor the site with the same field-of-view to quantify how much CH_4 was being released."

- Was the detection/identification of the emission plume manual or was a threshold automatically triggered?

  All of the column-averaged concentrations of CH_4 retrieved from the satellite observations are reviewed by experts at GHGSat to assess whether or not an emission is present. We have clarified this on line 133: "The raw images collected by the satellites are processed through GHGSat's proprietary toolchain and reviewed by experts at GHGSat."

The paper also compares the potential for detection of the leak by two continuously screening approaches, the MARS approach utilising the Tropomi satellite which provides detection of leaks above > 25,000 kg h-1 and the observation of enhanced concentration using the Tall Tower network in the UK. In order to place detection into context the paper could also provide a comparison against the 'baseline' leak detection mechanism in the UK for distribution networks – which relies on the fact that natural gas in the medium/low pressure network is odourised (the leak was apparently from a Medium Pressure (MP) 350 mbar – 2 bar pipeline which would contain odorised gas). Given the close location of a road, it seems likely the leak would be reported by the public – it could be informative to understand from the utility operator if any reports had been received.

Wales and West Utilities (WWU) confirmed that members of the public had reported the smell of gas several weeks before the work began and they were aware of the leak prior to the notification from GHGSat. They did not provide the date of the first report. We have added a sentence on line

372: “A member of public had also reported the smell of gas to the utility company prior to the notification from GHGSat and the utility company started work on assessing and repairing the leak on 27th April..”

Other pipeline leak screening approaches are already used in different countries for example the DVGW Set of Rules used in Germany that mandates screening of pipelines by ground or airborne surveys. While this may be beyond the scope of this paper, if a screening approach is being suggested, it should be compared to other existing approaches.

Thank you for this suggestion. A comparison of different screening techniques is, however, beyond the scope of this paper. Detailed information on DVGW was inaccessible, therefore we were not able to add it as a recommendation for monitoring fugitive emissions in the UK.

I do have a further very speculative suggestion – what was the original driver for the selection of the landfill for the satellite tasking? Was it selected because of pre-existing concerns about emissions? A quick Google search indicated press articles on complaints related to the odour from the landfill – an odour described as rotten eggs. Notably the mercaptan/sulphide based New Blend odorant used in the UK also has a ‘rotten egg’ odour. Could it be that local odour issues had been miss-attributed to the landfill when it was in fact the gas leak? This would have two implications – one, the leak had been occurring for quite a long period, and secondly the satellite provides a very useful tool for targeted investigation of such reports. But it would imply the underlying detection was indirectly down to the odorant in natural gas.

We identified the landfill as a potentially large source using the UK Environment Agency’s Pollution Inventory and it was relatively close to the nearest tall tower site (Ridge Hill) so we could do some model based analysis of the tall tower network (UK DECC network). We did not task the satellite due to reports of odours in the area. In our analysis we were unable to identify the start of the leak using modelling techniques and Ridge Hill observations so we do not know how long the gas had been leaking.

Note – the reference (Wales & West Utilities, 2023) links to a website with a generic description the utility company’s upgrade programme and does not give information on the pipe identified as the source of the leak – the utility company should be asked directly if this pipe was an older metal pipe or not.

We contacted Wales & West Utilities and they do not know the exact cause of the leak but think it is likely due to the age of the gas network in the area. They have been investing in a gas pipe replacement programme which replaces old metal pipes within 30 m of buildings with long-lasting plastic pipes.

Leak location

The general identification of the leak location is again confirmed by the utility company finding a leak in the area identified by the satellite. The paper presents the determination of the location of the leak from the satellite as (51.95088N, 2.09962W) reported in Line 94. Google maps shows this to the west of the railway line and ~50 m north of the road. The location of the repair of the pipe, and therefore presumably the location of the actual underground leak, is clear from Picture S1 and the aerial image (S3). The rough location of the repair as per Picture S1 is (51.95068N, 2.09902W) again from Google maps In the SI a picture is also presented of vegetation die back (S2) which could be an indicator of the location where the sub-surface leak entered the atmosphere. This is some distance from the repair, but is possible, as this is along the underground route of the pipe (as available from the utility company). I have estimated, based on the Picture S2, the region of vegetation die back is at (51.95063N, 2.09943W) (from Google maps). Note, I did not find any discussion or mention of S2 in the paper.

The method to determine the location of the source is described in Section 2.3.1 (see response to Reviewer 1, Main Comment 1). The location estimated from the satellite data shows the location at which the leak enters the atmosphere and not necessarily the pipeline break, see line 115. Also due to the proximity to the railway line, the location of the repair shown in Figure (Picture) S1 is not necessarily the location of the leak.

We have included some information in the Discussion about the vegetation die-back, see line 559: "During the 2nd ground-based mobile surveys we discovered an area of dead vegetation which could be due to plants being suffocated by the amount of CH_4 (see Supplement Figure S2), however this is circumstantial."

These locations are ~ 40-50 metres from the initial satellite leak location, which could be discussed in the context of the location uncertainty claimed for the satellite. The utility company may also be able to confirm the precise location of the leak and whether they surveyed the area and found the point the leak entered the atmosphere.

GHGSat do not assign an uncertainty to the source location because it depends on the plume shape, concentration gradient and wind direction (see method described in Section 2.1). Due to the age of pipes in the area WWU did not determine the exact location of the pipeline break because they replaced the whole pipe with a new one. As a result, we are unable to confirm the exact source location. We have added a sentence to line 556: "WWU confirmed they replaced the whole pipe at once so could not confirm the precise location of the leak."

The location is described as being determined from the initial satellite measurement when the leak was first identified, however, later in the paper the location used for NAME model, which is the same (51.95088N, 2.09962W), is described as being based on the average of four of the five source locations identified by the satellite. More information on the locations identified for each satellite measurement should be provided. The references for the satellite state ~25m location accuracy, but the location accuracy is also expected to vary with various factors (presumably meteorology, emission rate, surface type). It would be useful to provide and discuss the locations provided by the satellite and the uncertainty in these for each of the five passes that identified the emission.

We have updated the text in Section 2.1 to clarify the estimated locations, see line 98: "The gas leak was first detected by GHGSat during its first cloud-free overpass on 27th March 2023 and the location of the leak from the satellite was estimated to be 51.95097 ^{o}N, 2.09956 ^{o}W at approximately 33 m above sea level (m.a.s.l)." and line 117: "In our analysis, we use a mean location for the leak, 51.95088^{o}N, 2.09962^{o}W, estimated by the satellite. The individual estimated locations for each satellite observation can be found in the Supplement."

GHGSat retrievals have a pixel size of ~25 m and a georeferencing accuracy of ~25m which is stated on line 135.

We have also added a list of the locations estimated by GHGSat to the Supplement, Table S1.

Throughout the paper the location identified from the satellite data is used, by the satellite team and also by the NAME model and the MS retrieval. A discussion on the the sensitivity of each of the emission estimation approached to the location would be useful as a separate discussion. At various points throughout the paper (e.g Lines 180, 416, 471) uncertainties due to the source location are mentioned, so it would be useful to investigate the uncertainties in the location in more depth. When investigating the sensitivity of the NAME model results to the location a series of NAME model runs with the location varied by 10 m were carried out. The reason for the choice of 10m should be stated, the uncertainty in the satellite location determination could have been used, or some variation based on the range of locations determined by the satellite. Similar sensitivity analysis of the effect of the location on the other emission estimation methods could also be carried out.

This paper is assessing GHGSat's flux estimation and it is a data product they provide so we are unable to perturb the leak location for the GHGSat-derived flux estimates. We are aware that the flux estimations for the mobile survey and the NAME simulations rely on a mean location estimate from GHGSat and for this reason explore, the impact of this choice and the uncertainty it adds to the calculations in NAME by perturbing the location and in a Monte Carlo simulation for the Gaussian Plume Model.

We have updated our method to estimate the flux of the gas leak using the ground-based mobile survey observations (see response to Reviewer 1, Main Comment 2). The location of the leak is randomly selected within a box centred on the mean location of the leak and this is incorporated into the emission uncertainty.

In Section 4 we describe the impact of perturbing the location by 10 m in the N/S/E/W directions. A perturbation of 10m was selected because the NAME simulations are run at 10 m resolution and the observations around the peak of the median plume are approximately 10 m apart. The perturbations were kept the same for the satellite location simulations. We have added a sentence to line 563: "We selected 10 m location perturbation to match the resolution of the NAME simulations using the surface-based observations. We kept the perturbations the same for the NAME simulations which use the satellite observations."

It is also mentioned that the repair work may have made the emissions more diffuse, however, if the leak had been migrating underground before entering the atmosphere, the repair work may also have moved the effective leak location. More detail on the individual location information reported for each measurements would be informative in discussing this. The implication of a more diffuse emission is also not discussed in detail.

In Section 2.1 we describe the estimated leak location as being the approximate location where the gas is emitted into the atmosphere and not necessarily the pipeline break. Figure 1 shows that concentrations from the leak and size of the plume are smaller from 22nd May. The enhancements also remain to the west of the railway line but the trench was situated to the east of the railway line (Supplement Figure S1) suggesting that the largest emissions are still being released from the same location as first detected on 27th March. WWU said there could be more than one source of the leak but gas could be emitted into the atmosphere at a different location to the pipeline break.

In the text we refer to diffuse emissions from a wider area of excavated soil. We have added a line to the discussion, line 494: "More diffuse emissions could result in a wider CH_4 plume with lower concentrations which may not be above the threshold for enhanced CH_4 in the satellite retrievals."

Emission rate

Three different approaches are used for quantification:

Satellite concentration data with IME method, using GEOS-FP wind.

MS concentration data with a gaussian plume model using either onboard wind data or the Met Office’s UKV model.

NAME Lagrangian dispersion model applied to Satellite and MS concentration data, wind data from Met Office UKV model.

These sections should include more information on the sources of uncertainty in each emission estimation method and the impact of assumptions made.

We have clarified how the uncertainty is calculated for the IME method in Section 2.3.1. The IME flux rate estimate is highly dependent on wind speed (which is incorporated into the uncertainty quoted). We have referred to the Supplement of Varon et al. (2019) for full details of how the error on wind speed is incorporated into the flux uncertainty.

We have also changed our analysis method for the surface-based mobile survey estimates to provide a more robust estimate, see changes described in response to Reviewer 1, Main Comment 2.

For the NAME model we describe how the modelled and observed plumes do not overlap well so we have to apply some criteria to the NAME plume in order to scale it to match the observations. To account for this we changed the plume criteria to give a range of values for our estimate. This may result in an over- or underestimation of the flux due to the modelled plume not matching the satellite-observed plume but it is difficult to improve on this due to the reasons described in line 513.

To provide validation of the satellite emission data using alternate methods (MS and the NAME model) when no direct measurement of the emission rate is possible, it is important that all sources of uncertainty in all the methods are assessed. In particular this should address what sources of uncertainty may be correlated between the methods. In this study, the methods used are not independent – both MS and NAME make use of the location provided by the satellite in their retrievals. In addition, all the methods (apart from one MS measurement) use wind data derived from models. Given the importance of wind direction and wind speed on the results it would be very useful for the paper to assess the meteorological data in more depth. The local scale of the measurements, and the local impact of features such as the rail bridge, and the effect that the meteorological data has on the retrievals.

The emission estimates using the Gaussian Plume Model and NAME are guided by the mean location of the gas leak, estimated by from the satellite data. At the gas leak site there was no notable infrastructure and lack of access meant we were unable to determine location of the leak during the ground-based surveys. We decided to use the information we had available in order to estimate the fluxes using the surface-based mobile survey observations and the NAME model. During the second mobile survey, we noticed an area of dead vegetation close to the location of the leak. While this is circumstantial evidence, it could be due the vegetation being suffocated from the gas leak. WWU replaced the whole pipe instead of repairing a section of it so could not give us an exact location of the pipeline break.

The GEOS-FP wind data has a resolution of 0.25° lat x 0.3125° lon and the uncertainty on the wind speed is determined by comparing local wind speeds with GEOS-FP (see author response to Reviewer 1, Minor Comment related to line 327). This method captures the uncertainty related to the resolution of the wind speed value and is propagated into the satellite-derived emission estimates. We have added some information in the Supplement about the uncertainty of the wind speed and the satellite-derived estimates.

The UKV wind data has a higher resolution (1.5 km × 1.5 km) than GEOS-FP but the size of this plume means that the estimated wind speed is assumed to be constant across the whole plume in all three flux estimation methods (IME, Gaussian Plume Model and NAME). It is likely that the CH_4 will experience some variability in the wind from small-scale turbulence which is not fully captured in the wind speeds used in the flux estimations. To account for the uncertainty in the wind speed for the surface-based mobile survey we apply a random uncertainty with a mean of 1 m s^{-1} in the Monte-Carlo simulation (see author response to Reviewer 1, Main Comment 2).

We have added a line in the discussion on the impact of the resolution of the wind speed on the flux estimations, starting line 547: "In all three flux estimation methods the resolution of the wind data is much coarser than the size of the observed plume. Atmospheric transport at the surface through small-scale turbulence and influence of the local terrain may not be well represented, contributing to the uncertainty in the flux estimate."

The impact of local features, such as the railway bridge, are difficult to account for in the flux estimations. It is likely that the plume will be dispersed up and over the railway bridge and through the tunnel. In the Gaussian Plume Model we have perturbed the measurement height above the ground to account for the railway bridge. The resolution of the satellite data means we cannot see the impact of dispersion of the plume due to the railway bridge and often the wind direction is blowing the plume away from the bridge. We have included a discussion of this in Section 4, beginning line 579: "Although the ground-based mobile survey-derived fluxes and NAME-derived fluxes used independent observations and/or methodologies for the flux estimation, the assumed leak location was taken from the mean of GHGSat-derived location estimates. Despite using independent observations and models from the satellite data, it is noted that the flux estimates are not fully independent because we were unable to determine an independent estimate for the leak location."

While the uncertainties in the different estimated emissions are given, it would be useful to state if these are expanded uncertainties and with what degree of confidence these are quoted, and what coverage factor is used.

For the IME method, we have updated the description in Section 2.3.1 (see response to Reviewer 1, Main Comment 4).

We have performed a Monte-Carlo simulation to provide a more robust estimate of flux uncertainty using the surface-based observations. Full details can be found in response to Reviewer 1, Main Comment 2.

For the MS – at least 12 passes along the road are made for each measurement some more details would be useful - for example the time period of the traverses, how the plume average transect presented in the SI is determined and whether the emission retrieval is carried out on these averaged concentration data or on each run, and if the average is used. For measurements this close to the source, would it be expected that local wind fields, the effect of local terrain such as the trees and railway bridge would have a significant impact, and the concentrations observed by the MS be sensitive to small changes in local wind. Also the fact that the sampling height of the MS will vary as the car traverses the bridge may also have an effect. The discussion on the NAME model states the UKV wind data has a resolution of 1.5 km and an hourly temporal variability, some discussion of the potential influence of this on the very local wind field relevant to the local dispersion would be useful.

The median data are presented in Figure S7 on the composite mobile methane figure for each measurement period for clarity (due to subtle shifts in wind direction with time, the plumes do not perfectly line up spatially).

To show the variability in the measurements, the methane mixing ratio time-series data from the May campaign can be found in the Supplement, Figure S6. It shows 13 transects, with the median plume displayed in Figure S6 shown with the orange outline. We have referred the reader to the Supplement, Section S3 on line 313.

We have added this text to the Supplement, Section S3: "The variability is expected to be driven by a combination of meteorology and varying flux rates. This inherent variability in measurements made during mobile transect measurements can be seen in other studies of this nature, such as Caulton et al., (2018). Averaging of the fluxes derived from each individual transect has been demonstrated to be an effective method to estimate a true flux under controlled release conditions to within approximately 40% (Kumar et al., 2021)."

It is very difficult to quantify the impact of these very local influences such as the impact of the rail bridge.

The NAME model provides an alternative set of mass emission estimates for both approaches using a common model. The mass emission estimation methodology is tailored for each observation approach – for the satellite total column data is produced to match the satellite observations. It is noted on line 209 that the modelled plume and the satellite observations do not overlap well. Section 2.4 discusses the determination of model bounds, however, these relate to the definition of the plume area to be integrated for a given model output. The NAME model was also run with different locations to test the sensitivity to source location. Interestingly for most runs the perturbed model runs are all higher than the default location. This seems counter intuitive and it would be worth checking this is correct and discussing what might explain this.

We have double checked our values and improved the consistency of the number of significant figures of the main estimate to be in line with the number of significant figures used to calculate the fluxes in the location perturbation simulations. We have updated the results tables.

The plume criteria is reliant on the maximum value and the values below 1% and 5% of the maximum value. When we perturb the release location the maximum concentration of the plume can vary by 0.5 ppb which influences the size of plume we integrate over. The wind speed remains the same for the perturbed simulations, as a result, the difference in the maximum value between these simulation is mostly influenced by the particles being advected by unresolved motions, such as turbulence, which are simulated by a random walk technique. This contributes to the uncertainty in the NAME-derived fluxes.

We has added some text to line 569: "The wind speed remains the same as the original simulation for each perturbed location and as a result, the maximum value of the plume was influenced most by the particles being advected by unresolved motions, such as turbulence, which are simulated by a random-walk technique. This contributes to the uncertainty in the NAME-derived fluxes and shows that the flux estimation is dependent on the precise location of the leak when comparing with the IME and Gaussian Plume Model-derived fluxes, particularly when the fluxes are large (e.g. 20th April and 20th May)."

Some further investigation of the sensitivity of the NAME model plumes to other input parameters would be interesting, to provide more understanding of the uncertainty in this process. This would allow the differences between the model and satellite emissions to be put into context. It would also be worth exploring whether any of the constraints used to match the model to the observed satellite plumes could also introduce correlations between the two approaches. This could help understand how independent the NAME approach is. This is important as the NAME time series is used to support the suggestion that the emissions are varying with time.

The other parameter that could be changed in NAME is the maximum number of particles that are released during the simulation. We need a large number for the maximum number of particles (9×10^{7}) in the simulation to get a contiguous plume because of the high spatial and temporal resolution. We optimised the maximum number of particles in consultation with Met Office NAME scientists to get a typical plume shape output from NAME at high resolution. If we increased the maximum number of particles in the NAME simulations the plume shape would become wider as more particles are dispersed in the simulation. The final flux value would not be sensitive to the increase in the number of particles due to the unit release rate. We have included this in the discussion, line 574: "In the NAME experiments, the maximum number of particles in the simulation can be adjusted so that the model does not stop producing particles during the simulation. We conducted all NAME runs with a maximum of 9×10^{7} particles. This value was selected in consultation with Met Office NAME scientists and was determined to be the optimal number for our high-resolution simulations. The final flux value is not sensitive to the number of particles because the total mass released (determined by the release rate) is distributed across the number of particles released."

The modelled and observed satellite plumes do not overlap well so we do not try to match the model to the observed plume apart from the length of the observed plume. We apply criteria to the NAME plumes, which rely on the maximum value of the modelled plume and the length of the observed plume. This method assumes that CH_4 emitted in NAME has travelled the same distance and speed detected by GHGSat, stated on line 297 in Section 2.4. We have added a sentence in Section 2.4 line 303: "Apart from the release location and length of the plume, no other constraints from the observed satellite plume are applied to the modelled plume".

Similarly, the NAME model was configured to provide concentration path data to match the MS measurements – constrained to be at a height of 2m. This does not take account of the different elevations for the MS sampling point due to the terrain (e.g rail bridge). Some investigation of the influence of relevant parameters in the model for the mobile monitoring would again be useful.

The peak of the plume from the MS, defined in Section 2.4, is the maximum value and the observations taken before and after the maximum value. The maximum value observed during the mobile survey are on the tail end of the railway bridge and the NAME simulation was simulated at a 0-4m grid box, where the output is centred at 2m which is a sufficient grid box to account for the tail end of the bridge. We have corrected our description on line 309, Section 2.4: "The model was output at a horizontal resolution of 10 m × 10 m with a single 4 m layer to capture the volume observed by the mobile survey."

The time series of measurements is presented clearly with the graphic in Figure 2, the other events mentioned in the text could also be added, the data the utility company were informed and the data the investigation and repair work began. Any further information from the utility company on the leak, and whether they did vary the pressure in the pipe after being informed of the leak would be important information to also aid the interpretation of the results.

Figure 2 shows the date that GHGSat informed the utility company and it is stated on line 370.

The discussion notes the gas leak is well within the detection threshold of the satellite quoting 42 kg/h, however, in other areas of the paper a detection threshold of 100 kg/h is used, and the references given state 200kg/hr as the smallest detection. As mentioned earlier a discussion on the theoretical detection limit for the satellite for the conditions of these measurements would be useful.

In the discussion, we state that GHGSat has managed to detect an emission rate of 42 kg h^{-1} as an example of how low the emission rates can be that GHGSat can detect, however the theoretical detection limit given by GHGSat of the satellite which is used by the company is 100 kg h^{-1} (described in Section 2.2.1). GHGSat's detection threshold of 100 kg hr^{-1} is defined as the rate of emission where GHGSat would have a 50% probability of making a detection when the wind speed is 3 m s^{-1}. This means that statistically GHGSat can observe emissions smaller than 100 kg hr^{-1}, where the smallest emission detected is 42 kg hr^{-1} as detailed in McKeever and Jervis, 2022. See line 465: "The GHGSat detection threshold has a linear relationship with wind speed and is 100 kg h^{-1} at 3 m s^{-1} and 200 kg h^{-1} at 6 m s^{-1}, see McKeever and Jervis, 2022."

In line 91 the impact of wind speed is mentioned as likely cause of differences in the flux observed by the mobile system – however, as mentioned the wind data is not reported in the paper and should be added. It is suggested that the difference between the mobile and satellite measurements could be due to real changes in the emission. However, the two mobile measurements in May and June overlap with their uncertainties as do the two corresponding satellite measurements. Both techniques show a reduction in emissions from May to June – which may be real – though for both measurement techniques the difference between pairs of measurements lie well within their uncertainties. The difference between the two methods does look to be more systematic and probably not due to variations in emissions (though this can't be ruled out). Some discussion on whether there could be any reason for systematic under reading by the satellite or over reading by the MS approach would be informative. It is also worth noting that the NAME model approach gave lower results (in Table 1) for both the satellite and MS data on both the May and June results, with the mobile data again resulting in higher emission rates. This either supports the suggestion that the real emission rate is varying, or implies some systematic effect impacting both the NAME model and the other method retrievals. Some more detailed investigation/discussion on this would be informative.

A response on the systematic biases in observations and flux methods can be found below.

The paper would benefit from a more detailed review of the potential influence factors that might affect the methods, in particular any effects that might systematically affect all the methods. The sparse and temporally non-overlapping measurements does make it hard to draw firm conclusions on the performance of the satellite.

The wind data is provided in the Supplement. GHGSat have carried out a number of blind validation tests and found negligible systematic bias in their data (see line 76). The Licor-7810 instrument has a precision of ~1ppb so it is unlikely that there is any systematic bias in both the ground based and satellite measurements. Any systematic biases in NAME are likely to be negligible in comparison to the uncertainty from wind speed estimates and unresolved processes.

We have added a line 549: "Also any systematic biases in the measurements or flux estimation methods are likely to be negligible compared to the uncertainties from wind speed and model uncertainties described in Section 2."

Minor comments

Check consistency of form of units e.g.. in Line 117 both kg h^{-1} and m/s are used.

We have checked the consistency of our units throughout the paper.

Typo Line 462 "We assessed the frequency of pollution events during our both NAME_spring and NAME_long simulations and found a low number of 'leak pollution events'. " Check word order – should it be …during both our …

Thank you for noticing this, we have edited the line to say "during both our".

For the data from the 22nd May (438 ± 215 kg h -1) and 26th May (998 ± 377 kg h-1) the uncertainties do overlap – however, in Figure 2 the error bars do not appear to overlap for these two results – please check.

We have updated Figure 2 to include the fluxes derived from the surface-based observations and Monte-Carlo simulation described in Reviewer 1, Main Comment 2.

For the NAME results for the satellite data on the 22/05/2023 the reported result 384 is not within the bounds provided [173, 292], for all other results the model bounds are above and below the reported number – is this correct?

We have double-checked our NAME fluxes, including the result for 22/05/2023. We have added an explanation to the text on line 522: "On 22nd May the main flux estimate is larger than the estimation bounds (408 [169, 286] kg h^{-1}) and this is due to the plume selection criteria on this day. When we remove values less than 5% of the maximum value, the modelled plume length remains larger than the observed plume and as a result the scaling factor is smaller and the estimated flux is smaller than the main estimate. This is not the case for the other NAME-derived fluxes which use satellite observations."