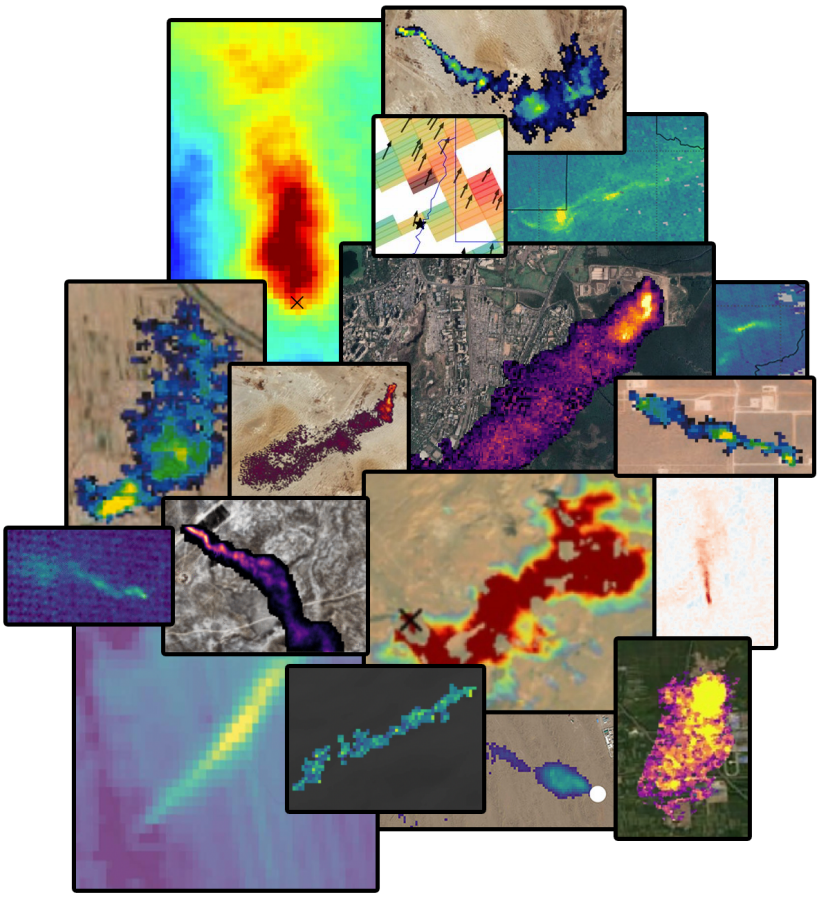
Framework for uncertainty characterization and prognostic estimation including its application to several different instruments/retrievals.

WP420 and WP430 - Deliverable D4.1 & D4.2



ESA Contract No. 4000143908/24/I-LR

Authors

Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València (UPV)

- Javier Gorroño (WP400 & document lead)
- Luis Guanter
- Javier Roger
- Shanyu Zhou

SRON Netherlands Institute for Space Research

- Joannes D (Bram) Maasakkers
- Matthieu Dogniaux
- Xin Zhang

Approved by

Date: 2026-02-24 Signature: 	Date: Signature: Simon Pinnock Digitally signed by Simon Pinnock Date: 2026.03.30 16:07:48 +01'00'
Ilse Aben (SRON) MEDUSA Science Lead	Simon Pinnock (ESA) Technical Officer

Change log

Version	Date	Status	Authors	Reason for change
1.0	4-April-2025	Draft	see Authors list	Version submitted to ESA
1.1	19-May-2025	Resubmitted to ESA	see Authors list	Revised based on RIDs received

Contents

1 Preface

This document corresponds to the first version of two for Deliverable 4.1 (“Framework for uncertainty characterisation and prognostic estimation”) and 4.2 (“Application of the framework to several different instruments/retrievals and its evaluation”) from the Work Package (WP) 420, “A framework for methane flux rate uncertainty” and WP 430, “ Application of the community framework”.

It describes an uncertainty framework for plume imagers and flux mappers covering the measurement equations in Section 2 and uncertainty sources in Section 3. It also contains a framework for the probability of detection of methane plumes in Section 5

The uncertainty analysis will focus on Sentinel-2 (as representative for band imagers), EnMAP (as representative for hyperspectral imagers), and TROPOMI (as representative for area mappers). For all, we will work through the IME quantification Varon et al. (2018). It is expected that the examples and general framework for uncertainty analysis are readily-applicable to other missions. Because of the different spatial scales and level-2 product availability, not everything discussed here will apply to all three instrument types and hence not all analyses will be done for all three. For instance, TROPOMI uncertainty analysis will take L2 product (i.e. methane concentration map) as the starting point, since this is a well calibrated and validated operational product.

We have agreed to merge deliverables D4.1 and D4.2 because the methodology and implementation of uncertainty contributions are difficult to separate.

2 Uncertainty tree for methane flux rates

The uncertainty analysis requires as a first step the definition of a measurand and a measurement function. Our measurand is the methane flux rate Q with a set of measurement functions defined by the IME methodology described by Varon et al. (2018) as follows:

$$Q = \frac{U_{\text{eff}} \cdot \text{IME}}{L}, \tag{2.1}$$

where U_{eff} is an effective wind speed, IME refers to the total excess mass of methane contained in the plume and L the plume length scale in m, which is typically approximated by the square root of the plume mask area.

Each one of these quantities can be further detailed with the IME in kg units defined as:

$$\text{IME} = k \sum_{i=1}^{n_p} \Delta XCH4(i), \tag{2.2}$$

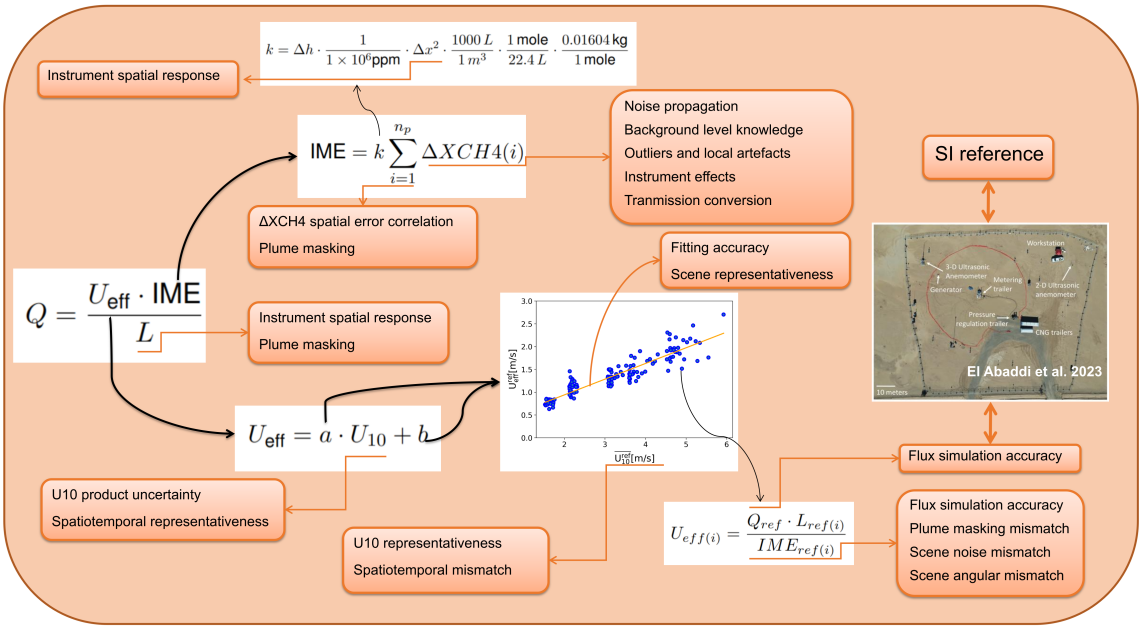


Figure 1: Uncertainty tree diagram for methane flux rate quantification of plume imagers using the IME method.

where n_p is the number of pixels in the plume and k is a scaling factor that converts the total of the pixel-wise methane concentration values $\Delta XCH4$ in ppb to kg by assuming Avogadro's law and taking into account the pixel size as follows:

$$k = \Delta h \cdot \frac{1}{1 \times 10^6 \text{ppm}} \cdot \Delta x^2 \cdot \frac{1000\,L}{1\,m^3} \cdot \frac{1\,\text{mole}}{22.4\,L} \cdot \frac{0.01604\,\text{kg}}{1\,\text{mole}} \tag{2.3}$$

where Δh refers to the height of the methane column in metres and Δx refers to the pixel resolution (note that the equation assumes a squared pixel but can be separated into two different spatial resolutions).

Finally, U_{eff} accounts for turbulent diffusion of the methane flux, and also for the fraction of the plume that is not detected due to retrieval noise (Varon et al., 2018). U_{eff} can be expressed as a linear or logarithmic function of the measurable 10-m wind speed U_{10}. This function can be defined with plume simulations that include the corresponding spatial sampling and retrieval noise of each instrument.

The definition of these functions and their dependency is key to setting a traceability chain that, ideally, should be referred to a reference standard. This can be synthetised and graphically displayed in an uncertainty tree Mittaz et al. (2019). This helps interconnect the different measurement equations and, from each one of their parameters, produce different uncertainty contributions. The result of this work for the IME method is illustrated in Figure 1.

The uncertainty tree departs from the original IME equation (see Eq. 2.1).

The upper part of the graph defines the total mass IME uncertainty (see Eq. 2.1). This uncertainty is defined by contributions from the enhancement methane map $\Delta XCH4$. This is affected by instrument noise and other effects (e.g. stripping), but also by artefacts and offsets that are generated during the retrieval itself. The retrieval process generally obtains an estimated transmission of the methane plume that must be further converted into a methane concentration with the associated conversion uncertainty. Masking of the plume introduces an uncertainty itself, but also reduces the impact of the pixel-level $\Delta XCH4$ uncertainty due to error correlation. Finally, conversion from ppm units into kg introduces uncertainty due to the instrument spatial response. The term Δh does not introduce an uncertainty, as it is cancelled during the U_{eff} calibration if a consistent value is selected. Neither can the rest of conversion terms be considered as a source of uncertainty given their deterministic nature.

The middle part of the graph shows the contributions associated with the plume length L. This term is closely related to the IME since their main contributions (*spatial response* and *plume masking*) affect both of them. Therefore, rather than studying them separately, we define their impact at the IME/L ratio (i.e. implicitly accounting for the error correlation).

The lower part of the uncertainty tree defines the U_{eff} uncertainty. The definition sets a regression curve dependent on U_{10}. We have an uncertainty associated with U_{10} that is associated with the wind product itself and its spatio-temporal representativeness. Furthermore, there is an associated uncertainty with the regression curve that depends on the selected fitting and set of references. In this particular case, we assume that the references are simulated emissions. The selection of curve fitting and simulated scenes are important sources of uncertainty. Equally important is any mismatch between the simulated retrieval methodology and the operational one. This includes simulation of the noise, the angular configuration, or the masking methodology. In calibration jargon, this is usually referred as “test as you fly and fly as you test”.

Finally, the LES model can reproduce the methane plume turbulences under some assumptions and limitations that do not perfectly match the real conditions. Thus, the link between IME_{ref} and Q_{ref} might differ from a real scenario and it is reflected in an associated uncertainty with the simulated flux rate. This can be obtained by comparing to controlled methane releases that, in turn, are calibrated against SI standards from metrology laboratories.

3 Uncertainty contributions assessment

The uncertainty tree in Figure 1 and its discussion in Section 2 set a first step in our uncertainty analysis for methane flux rates. This defines the different uncertainty sources and their interrelation based on the measurement equation.

In the following subsections, we will describe them. In some cases, the uncertainty sources have been merged into a single subsection due to a similar assessment.

3.1 $\Delta XCH4$: Noise propagation

This contribution is the result of propagating the instrument noise through the retrieval.

The resulting errors are typically decorrelated in the spatial and temporal dimensions leading to a minimisation of its effect with large plume areas and with global budgets over areas and/or periods of time. This is discussed separately in subsection 3.5.

There is the possibility that the *noise* includes correlated patterns from straylight or detector striping. However, this is generally taken as a separate effect and discussed in Subsection 3.2.

This uncertainty can be measured directly on the resulting methane enhancement map Varon et al. (2021); Guanter et al. (2021); Sánchez-García et al. (2022). However, we can define an analytical propagation of the noise based on the measurement equation as in Fahlen et al. (2024).

The former procedure is straightforward, but has the main drawback that the resulting error distribution accounts for both methane enhancement noise and scene and methane plume variations. Although efforts can be placed to minimise this impact (e.g. select an area with low artefacts), it will result in an overestimation of the expected noise. The latter method provides a bias-removed statistical uncertainty per pixel but requires an explicit mathematical formulation that correctly captures the mathematical relationship, error correlations, and expected distribution. This is hard to model for some missions and retrieval methodologies where the mathematical relationship might not be fully explicit and assumptions such as nonlinear response or Gaussian distribution are included.

Based on the previous considerations, we have defined a noise propagation scheme that defines its domain of validity. The MonteCarlo (MC) is a suitable reference, as it does not require an explicit mathematical model and fully propagates any input distribution to the output considering internal dependencies (JCGM, 2008b; BIPM et al., 2011).

The first tests suggest a normal distribution for values below 2ppm and large noise (over 200ppm). Over these values, the nonlinear relationship is translated into a nonnormal distribution and the analytical (and standard deviation metrics) might not be representative.

At TROPOMI scale, the $\Delta XCH4$ noise can be directly obtained from the TROPOMI L2 files that provide the retrieval precision for XCH4. Currently, Schuit et al. (2023) use the averaged L2 precision of TROPOMI XCH4 in the image multiplied by a factor 2 as suggested in the TROPOMI L2 documentation as $\Delta XCH4$ uncertainty estimate, which is used to perturb the background value globally (conservative fully-correlated perturbation, within $\pm\, 2\, \Delta XCH4$). As the evaluation of TROPOMI hotspot products in MEDUSA starts from Level 2 upwards, this noise will not be further analyzed in the project.

3.2 $\Delta XCH4$: Instrument effects

Because most of the retrievals are based on spectral and/or temporal differences, absolute radiometry is not a major concern. Instead, local artefacts and spectral definition are considered to have a considerable impact.

For example, striping (in the across or along track directions) can have an important impact on the retrieval. It originates from radiometric differences at detector level, but also from microvibrations or straylight patterns.

The definition of an uncertainty budget suggests that, where possible, these known systematic effects are compensated for or, in the absence of a correction, are added linearly to the budget JCGM (2008a). This correction is very specific for each instrument, but for hyperspectral sensors this is generally minimised by considering the specific spectral response and matched-filter statistics at the column level Guanter et al. (2009, 2021).

The case of missions such as Copernicus Sentinel 2 or Landsat becomes more challenging. For those instruments, effects such as stripping and spectral uniformity are not minimised or accounted for in the uncertainty budget. Its correction becomes more challenging because current products include an orthorectification and the connection to the focal plane is lost. Thus, upcoming studies should consider images at the focal plane (e.g. this is Level-1B products for Copernicus Sentinel 2).

The TROPOMI product used in Schuit et al. (2023) was destriped following the approach introduced by Borsdorff et al. (2019) for Carbon Monoxide. In the most recent operational product, a destriped methane field has been included. At TROPOMI scale, we can perform a best-effort sensitivity analysis by exploring the impact of various destriping kernel parameters, and of taking destriping into account or not on quantification results.

3.3 $\Delta XCH4$: Background knowledge and outliers

The definition of $\Delta XCH4$ is based on the fact that a reference methane concentration must be set to define a background level. We can set this reference over homogeneous and a priori methane-free areas, or alternatively, select a large area and assume sparsity.

Defining an uncertainty associated with this effect could explore the sensitivity to define this background with different methods and thresholds. These events can occur at a synoptic scale (e.g. dust storms), but more challenging are those local artefacts (or outliers) caused by scene spectral responses that interfere with the retrieval process. In those cases, the IME and the plume selection itself will be biased. Defining an uncertainty value might not be appropriate because it is a more specific effect. However, in Fahlen et al. (2024) some metrics have been proposed that could be used as an auxiliary indicator.

Because TROPOMI provides high accuracy estimates of absolute XCH4 values, the back-

ground XCH4 value can be obtained through the median of the pixels outside the plume mask, as currently being done by Schuit et al. (2023). Contributions of this background estimation to TROPOMI-based quantification uncertainties are discussed in Sect 3.5.

3.4 $\Delta XCH4$ Transmission conversion

The original inputs to retrieve the $\Delta XCH4$ map are the different radiance bands of the sensor. Typically, a ratio or band fitting will be required to estimate the transmission as a result of the methane plume. The conversion of this transmission into a methane concentration requires the use of a radiative transfer or the Beer-Lambert law.

The different assumptions (e.g., the air-mass factor) should be considered to estimate the impact of this contribution. The methane enhancement units can be expressed in absolute terms (mol/cm^2), (kg/m^2) or a relative value (ppm). For the second, we use the following conversion from kg/m^2 to ppm units:

$$\Delta XCH4 = \Delta\Omega \cdot \frac{M_a}{M_{CH4}} \cdot \frac{1}{\Omega_a} \tag{3.1}$$

where the molar mass of dry air $M_a = 0.0289644\ kg/mol$ and the one of methane is $M_{CH4} = 0.01604246\ kg/mol$. The term $\Omega_a = 10332\ kg/m^2$ refers to the column of dry air at sea level.

The latter term is the main source of uncertainty when we convert these units or model in a radiative transfer. In both cases, the dry air assumption and the surface level introduce a systematic error that must be compensated for and/or included in the uncertainty budget. For example, a transmittance conversion including different altitudes (or surface pressure) and water vapour should compensate or provide a sensitivity assessment of this effect.

At TROPOMI scale, the conversion coefficient is determined from pixel-wise surface pressure values included in the L2 TROPOMI XCH4 product. Currently, Schuit et al. (2023) do not account for this contribution in their uncertainty ensemble. Standard surface pressure precision values (encompassing reanalysis precision as well as spatio-temporal representativity) can be determined by comparing the surface pressure included in TROPOMI L2 product with local in-situ surface pressure measurements (e.g. at airports), or use the scatter of reanalysis ensemble values.

3.5 IME: ΔXCH4 spatial error correlation

The uncertainty sources previously commented on in $\Delta XCH4$ refer to a single pixel. For the IME method, the calculation requires the addition of several pixels. In the case of plume imagers (with 60 m or below spatial resolution), it is likely that the number of pixels is considerably large. In that case, we can assume as a rule of thumb that

- noise propagation (see subsection 3.1) becomes negligible due to the uncorrelated nature of the noise,
- instrument effects (see subsection 3.2) are partially compensated depending on the instrument effect, orientation of the plume or methodology,
- background knowledge and outliers (see subsection 3.3 are not significantly reduced because of the expected spatially correlated background error in the plume area and the non-symmetric distribution of errors due to outliers and,
- transmission conversion (see subsection 3.4) is not significantly reduced because water vapour and altitude are not expected to largely vary in the plume area.

At TROPOMI scale, using the scatter of XCH4 values in non-plume pixels, this contribution can be evaluated by perturbing background values under two extreme hypotheses: (1) fully uncorrelated perturbations at pixel level; and (2) fully correlated perturbations (global shift of background). Currently, Schuit et al. (2023) include the impact of the XCH4 background value in their uncertainty ensemble by varying the background globally by ± 2 standard deviations of the scene-averaged L2 precision of TROPOMI XCH4 (by 0.4 steps). Whether this uncertainty estimate is representative can be explored using the proposed data-driven scatter of non-plume XCH4 background values.

3.6 IME/L: instrument spatial response and plume masking

There is a *limited knowledge of the spatial resolution* of the pixel that is typically defined by the point spread function (PSF) or the modulation transfer function (MTF)(Markham, 1985). This limited knowledge can be attributed to preflight measurement uncertainty, but also to changes across the focal plane or thermolastic effects. For plume imagers, this effect is not currently considered. It is not expected to be a major source of uncertainty but, for very short plume lengths, this might have a certain impact.

The plume masking is an important source of uncertainty. However, it must be carefully assessed since it cannot be considered directly as a source of error but in combination with other effects. At this stage, it is proposed to study its impact in combination with the image background and outliers in $\Delta XCH4$ as well as the mismatch criteria in the calibration of U_{eff} (see subsections 3.3 and 3.13)

At TROPOMI scale, the contribution of the PSF is not fully accounted for in most standard approaches exploiting simulations to interpret TROPOMI data: the simulations are generally resampled using a cookie-cutting approach (area-weighted average of simulation pixels intersecting TROPOMI pixels), regardless of any actual PSF function. To study the impact of this assumption, a best-effort test case can be set up to provide an example quantification comparing results with and without accounting of the TROPOMI PSF.

Currently, Schuit et al. (2023) operationally include plume masking in their uncertainty ensemble by exploring different values of masking thresholds with which to apply their dilation-based masking approach. The uncertainty caused by the assumption of neglecting the PSF in the total plume mass may be added to the contribution of the masking threshold based on overall results of the best-effort test case that has been mentioned in the previous paragraph (with the risk of including similar error contribution twice).

3.7 U_{10}: product uncertainty

The inclusion of uncertainty as part of the reanalysis U_{10} products is the preferred approach, as it streamlines the production of methane products. This is already (partially) occurring for ERA5 Soci et al. (2024) and ERA5 Land products Muñoz Sabater et al. (2021), which include a 10-member ensemble spread. However, the perturbations are typically an underestimate with mostly random components and not systematic ones. In addition, it does not consider the physics of the surface model MacLeod et al. (2016).

The alternative is to provide an uncertainty based on a validation exercise or dataset statistics. Varon et al. (2018) reported 2m/s comparing local wind speed values measured at different airports with the GEOS-FP values. A similar value of 2 m/s, for values equal to or greater than 4 m/s, can be inferred from the validation of MERRA-2 products in Carvalho (2019). Finally, Ayasse et al. (2024) estimates the wind uncertainty based on the standard deviation from a 9 km window around the source location and from a 3 hour window around the plume acquisition.

The main drawback of an approach based on validation and/or product statistics is that it requires a lot of resources (including well-calibrated wind references), might not be representative of the plume area and the different scenarios, and does not separate spatiotemporal representativeness from U_{10} product uncertainty.

For plume imagers, and given the limited uncertainty information of reanalysis products, the validation exercises are taken as a reference. However, further efforts are directed towards a separation of the reanalysis product knowledge against the spatio-temporal representativenes for the plume area. This is further discussed in Subsection 3.8.

At TROPOMI scale, the uncertainty on wind speed values can be assessed similarly as Varon et al. (2018) by comparing sampled reanalysis wind speed values with in-situ wind speed measurements (e.g. at airports). Such an uncertainty estimate aggregates product uncertainty with spatio-temporal representativeness uncertainty. An additional complementary way of estimating wind speed product uncertainty can be to compare different meteorological reanalysis results.

Currently, Schuit et al. (2023) conservatively account for this contribution by including an ensemble of three different wind products in their uncertainty ensemble, combined with a 50% uncertainty assumption on product-wise wind speed values. Refinements can be explored by carrying out a more comprehensive comparison of these wind speed products against in-situ

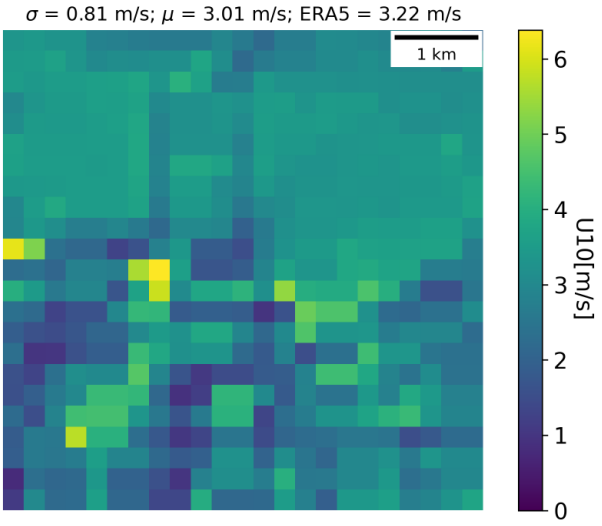


Figure 2: Global Atlas wind example over the North of Spain. The size area is approximately 4× 4 km with a mean of 3.01 m/s and spread of 0.81 m/s. *ERA5* refers to the mean wind speed on the same Global Atlas if we took the area of 9× 9 km based on the footprint of the closest ERA5 pixel.

data for example, and evaluating if the spread in the different wind products that are used is representative for the uncertainty on the individual products. The uncertainty in the wind speed is the largest contributor to the overall uncertainty and therefore important to evaluate.

3.8 U_{10}: spatiotemporal representativeness

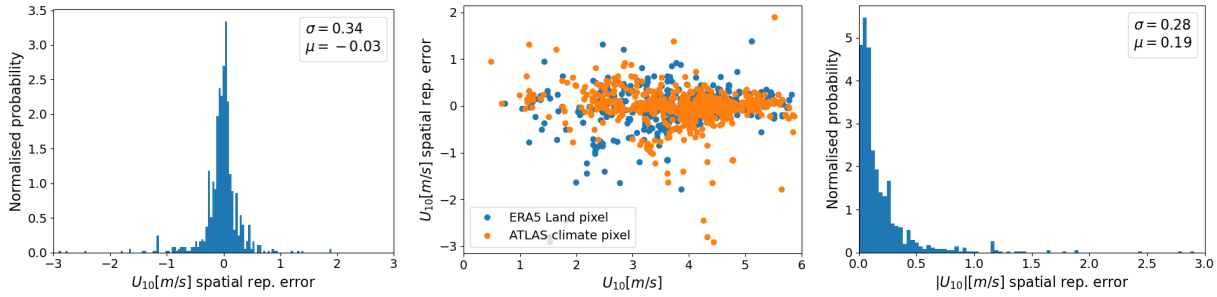
The spatial representativeness of the wind can be studied using (with caveats) a high-resolution wind climatology as a reference. We selected the Global Wind Atlas with 250-m spatial resolution from Davis et al. (2023). An example of the product in the north of Spain is displayed in Figure 2.

Selecting different regions of interest with different sizes indicates the potential errors introduced by the spatial resolution of the U_{10} products. In Figure 2, the size area is approximately 4× 4 km with a mean of 3.01 m/s and spread of 0.81 m/s. If we select the area of 9× 9 km based on the footprint of the closest ERA5 pixel, it would be 3.22 m/s. The spatial representativeness error is defined as

$$\varepsilon_{\text{rep}} = U10_{\text{plume\_clim}} - U10_{\text{ERA5\_clim}} \tag{3.2}$$

where $U10_{\text{plume\_clim}}$ is defined as the wind atlas pixel average inside the selected wind reanalysis product pixel (e.g., ERA5 or GEOS-FP) and $U10_{\text{ERA5\_clim}}$ is defined as the wind atlas pixel value. The latter can be assumed as a representative value of the plume although a plume mask can also be considered.

We have studied this effect at a global scale by selecting a list of 1000 locations from the



(a) U_{10} error dist. (b) U_{10} scatter (c) $|U_{10}|$ error dist.

Figure 3: Spatial representativeness errors for 1000 locations of UNEP's IMEO database as 3a U_{10} error distribution, 3b scatter plot of $U10_{\text{plume\_clim}}$ against $U10_{\text{ERA5\_clim}}$ and 3c absolute error distribution.

UNEP's IMEO database. Figure 3 summarises the main results obtained for the spatial representativeness error.

As a general figure, we can provide a value of spatial representativeness error around 0.3-0.4 m/s. The nearly symmetric distribution in Fig. 3a indicates that this uncertainty is reduced with the spatial integration of emissions. The scatter result in Fig. 3b also indicates that the uncertainty is highly independent of the U10 value (in absolute terms), which means that its impact varies in relative terms when grouped over different U10 ranges. This range of relative errors moves between 10 and 30%.

Further analysis of these contributions will seek to repeat the process over all +7000 UNEP's IMEO points and refine by clustering (e.g. O& G basins).

We will also need to complement the investigation of the spatiotemporal representativeness with simulations (sensitivity through time and space). A database of simulations is currently under development (see Subsection 5.2) which should also be expanded with simulations at the product scale (several km). Local representativeness (or mismatch to the U_{10} product during the U_{eff} calibration) is discussed separately in subsection 3.11

Given the scale of TROPOMI plumes, the observed methane enhancement maps are the results of not just the instantaneous wind speed, but also of the wind speed history. Proper accounting of this wind history in mass balance estimates would require transport-accounting approaches (e.g. Lagrangian simulations) that are out of the scope of this MEDUSA effort. At TROPOMI scale, we suggest to provide aggregate uncertainty contributions that combine wind product uncertainty and spatio-temporal representativeness, as described in Sect. 3.7, and evaluate the use of different wind speed averaging (spatial and temporal) based on model simulations (as described in the following sections).

Currently, Schuit et al. (2023) sample the U_{10} wind speed at the reanalysis grid point closest in space and time to the automatically-estimated plume origin.

3.9 U_{eff} calibration: fitting accuracy

The fitting accuracy is the result of describing the relationship between U_{eff} and U_{10} through a polynomial. The selection of a polynomial has an impact on the resulting fitting, and thus a sensitivity analysis is expected to be carried on. At the time of writing, a large database of simulations is being populated (see Subsection 5.2). This will be the core input for testing potential fittings. Furthermore, we will tentatively explore the main reasons for this spread of values and potential improvements of the fitting (e.g., a fitting based on shape similarity).

At TROPOMI scale, the U_{eff} calibration fitting accuracy can be evaluated as the spread (e.g. standard deviation) of the residuals between the fitted calibration and the ensemble of synthetic situations that were used to derive the calibration. Currently, Schuit et al. (2023) operationally assume a 5% uncertainty of effective wind speed calibration parameters, which are perturbed together in one dimension of the uncertainty quantification ensemble. This can be refined by quantifying uncertainties based on the ensemble of U_{eff} rather than the fitted parameters themselves. Efforts are currently underway at SRON to improve effective wind calibration for the IME method applied to TROPOMI data.

3.10 U_{eff} calibration: scene representativeness

One of the most challenging areas when generating a U_{eff} calibration is compiling a representative set of scenes that describe the potential scenarios in which a plume may originate.

These sets of scenes may comprise different wind speeds, but also latent flux heat, temperature profiles, or even topographic reliefs.

This contribution can be explored by sensitivity analysis of the fit against different scene sets. At the time of writing, a large database of simulations is being populated (see Subsection 5.2) which will be the main input to select different scene sets.

At TROPOMI scale, the U_{eff} scene representativeness contribution can be quantified based on a N-fold cross-validation approach where 1 subset (grouping all scenes from a given region) is withheld to evaluate a calibration fitted on N-1 subsets (thus excluding this given region), assuming that the set of scenes is a fair representation of all regions globally (desert, vegetation, urban, etc., all continents, etc.). The scene representativeness uncertainty contribution can then be calculated as the scatter of the average emission rate error on all the N individually-withheld subset cases. Currently, Schuit et al. (2023) do not operationally account for this contribution. Refinements of the uncertainty calculation can be explored by applying the above-suggested approach.

3.11 U_{eff} calibration: U_{10} representativeness and spatiotemporal mismatch

A theoretical approach to estimate U_{eff} would be to calculate the effective wind speed that corresponds to the mean transport speed of the plume. The effective wind speed is the vertically averaged wind speed weighted by the vertical profile of the trace gas concentration. It can be calculated as

$$u(x,y) = \frac{\int_0^{z^T} \rho_e(x,y,z) u(x,y,z)\,\mathrm{d}z}{\int_0^{z^T} \rho_e(x,y,z)\,\mathrm{d}z}, \tag{3.3}$$

By default, U_{eff} can be trained using large-eddy simulations considering this equation. However, in real observations, this is not possible because we do not have access to a 3D distribution of the wind.

A common practical approach is to infer U_{eff} from U_{10} using LES simulations such as from Varon et al. (2018). However, in subsections 3.7 and 3.11 we discuss how this relationship can be limited by knowledge of U_{10} and its representativeness. Here, we discuss this spatiotemporal representativeness between the product and the simulations (i.e. mismatch).

When calibrating U_{eff}, the selection of a large area of U_{10} closely represents the selected pixel in the reanalysis product although not the physical meaning of the flux rate.

For the spatial dimension, we propose to train the U_{eff} at three different levels:

- Whole simulation area (this is 1.2×1.8 km for the simulation database presented in Subsection 5.2)
- A weighted average of U_{10} with the plume concentration (similar to Eq. 3.3 but not considering the vertical dimension)
- training against the exact source location.

Furthermore, it is possible to explore alternative definitions of U_{eff} such as the wind in the boundary layer U_{PBL} currently used in the quantification of TROPOMI (see (Gorroño et al., 2025)). These alternatives can be considered if better spatiotemporal representativeness or fitting is achieved.

At TROPOMI scale, the U_{10} representativeness can be explored by sampling values from WRF-based calibration simulations U_{10} values for different locations within the plume (average plume mask, plume source, etc.) and at different timestamps (from TROPOMI overpass up to a few hours before, if relevant). Given the scale of TROPOMI plumes and the scale of transport involved, using winds from other altitudes could also be a dimension to explore. Currently, Schuit et al. (2023) quantifications rely on a set of calibration synthetic plumes from which U_{10} was sampled as the enhancement weighted average of the pixels included in the plume mask. Refinements can be explored by testing several of the U_{10} sampling options described above. This contribution is currently not included in the uncertainty ensemble.

3.12 U_{eff} calibration: flux rate simulation accuracy

The U_{eff} calibration requires a reference to which a regression curve can be set. The reference is a simulated emission that reproduces, as much as possible, realistic conditions of gas transport in the atmosphere, partially solving the Navier-Stokes equations.

The main challenge is that, at the moment, there is no benchmarking of these simulations, and consequently, an accuracy associated to them. This is out of scope of this project, but we can define a potential methodology. The proposed methodology is based on the comparison of the validation results between those of real satellite products and simulated products. In simple mathematical terms, this difference ε_{sim} can be described as:

$$\varepsilon_{\text{rep}} = \beta_{\text{simulations}} - \beta_{\text{satellite}} \tag{3.4}$$

where $\beta_{\text{simulations}}$ and $\beta_{\text{satellite}}$ represent the slope of the validation curve for the simulated products and satellite products, respectively.

The detailed methodology will be refined once a larger simulated database is available (see Subsection 5.2) and more controlled releases are performed. For example, simulations could be placed in the vicinity areas or at the same emission source location for days with a satellite overpass and no controlled release emission (before, during, or after the experiment).

Because of the reliance on controlled releases, this evaluation will not be extended to the TROPOMI scale.

3.13 U_{eff} calibration: scene modelling and plume masking mismatch

The generation of calibration coefficients for U_{eff} requires that the simulated IME and $\Delta XCH4$ maps are not only realistic but also match the same conditions as their operational use.

In the case of plume masking, it is relatively straightforward, since we need to ensure that the same methodology is followed during the simulation and operational retrieval. However, noise patterns or angular considerations remain a challenging area in the U_{eff} calibration.

The instrument noise can be modelled by reproducing samples out of a normal distribution with a specific spread. However, the instrument effects (see Subsection 3.2) are more difficult to include in the simulation. Efforts should be made to include effects such as stripping or the spatial response of the pixels in the simulation. However, it requires a knowledge of the PSF and its application over a finer spatial resolution (e.g. at the 1-2 m spatial resolution).

One of the areas that we have investigated in detail in this first stage is that of considering the angular configuration when calculating the $\Delta XCH4$ maps.

The large eddy simulation recreates the entire 3D plume distribution over the atmosphere.

In order to obtain a $\Delta XCH4$ map, the vertical dimension must be collapsed. The most simple approach is to sum all the vertical layers per pixel. However, this does not recreate a real scenario in which the projection of this plume depends on both the solar and the viewing angles.

Because the plume rises above the surface, the projection on the ground is shifted from the nadir view. This apparent displacement of the plume (i.e parallax) results in a total shift Δd at the voxel i, j, z that can be expressed as:

$$\Delta d_{i,j,z} = h_{i,j,z} \cdot tan(ZA) \tag{3.5}$$

where $h_{i,j,z}$ expresses the altitude of the voxel that here is calculated based on the layer index and vertical spatial resolution. ZA refers to the viewing or sun zenith angle in radians.

We can further decompose this shift into the spatial components x and y as:

$$\Delta d_x = \Delta d_{i,j,z} \cdot cos(AA) \Delta d_y = \Delta d_{i,j,z} \cdot sin(AA) \tag{3.6}$$

Because we have two different paths (downwelling related to the sun angle and upwelling related to the viewing angle), it results in two different $\Delta XCH4$ maps (ΔX_{CH4up} and $\Delta X_{\text{CH4down}}$) with their corresponding spatial shifts at a pixel level. Furthermore, we only consider the direct path because diffuse components are small in the SWIR.

Following the Beer-Lambert law, we can describe the plume transmission as the combination of the upward and downward transmission ($T_{\text{up}}(\lambda)$ and $T_{\text{down}}(\lambda)$) and the air mass factors (AMF_{up} and AMF_{down}):

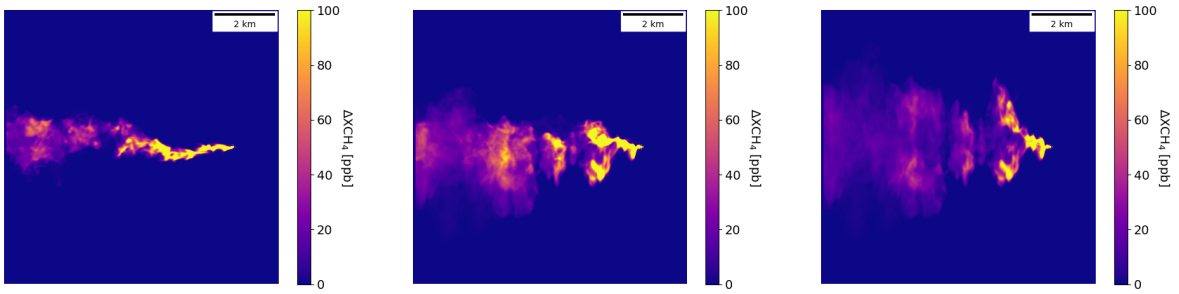
$$T_{\text{plume}}(\lambda) = T_{\text{up}}(\lambda) \cdot T_{\text{down}}(\lambda) = e^{-(\text{AMF}_{\text{up}} + \text{AMF}_{\text{down}}) \cdot \sigma_{\text{CH4}} \cdot \Delta X_{\text{CH4}}} \tag{3.7}$$

Given this equation, we can establish a direct link between each of the enhancement maps and obtain the combined methane enhancement $\Delta XCH4$ as follows:

$$\text{AMF}_{\text{up}} \cdot \Delta X_{\text{CH4up}} + \text{AMF}_{\text{down}} \cdot \Delta X_{\text{CH4down}} = (\text{AMF}_{\text{up}} + \text{AMF}_{\text{down}}) \cdot \Delta X_{\text{CH4}} \tag{3.8}$$

We tested this methodology with a simulation that has a spatial resolution of 25 m and a vertical resolution of 15 m. The original spatial resolution (X and Y dimensions) are resampled to 20 m, which matches that of the Copernicus Sentinel 2 mission. The modelled noise is set to 150 ppb, typically found over homogeneous scenes, and the plume is scaled to 5t/h. The simulation has a geostrophic wind direction of 90°azimuth angle. Figure 4 illustrates $\Delta XCH4$ maps with viewing and solar azimuth angles of 0 and 180°, respectively (i.e. an angular configuration across the plume direction).

The plume in Fig. 4a represents a nadir configuration as currently modelled. The plume



(a) SZA 0°, VZA 0° (b) SZA 60°, VZA 10° (c) SZA 60°, VZA 60°

Figure 4: ΔX_{CH4} map for a plume snapshot with zenith angles of (4a) 0° for both zenith angles,(4b) 60° for solar and 10° for viewing, and (4c) 60° for both zenith angles. Viewing and solar azimuth angles orthogonal to the plume direction and separated by 180 °.

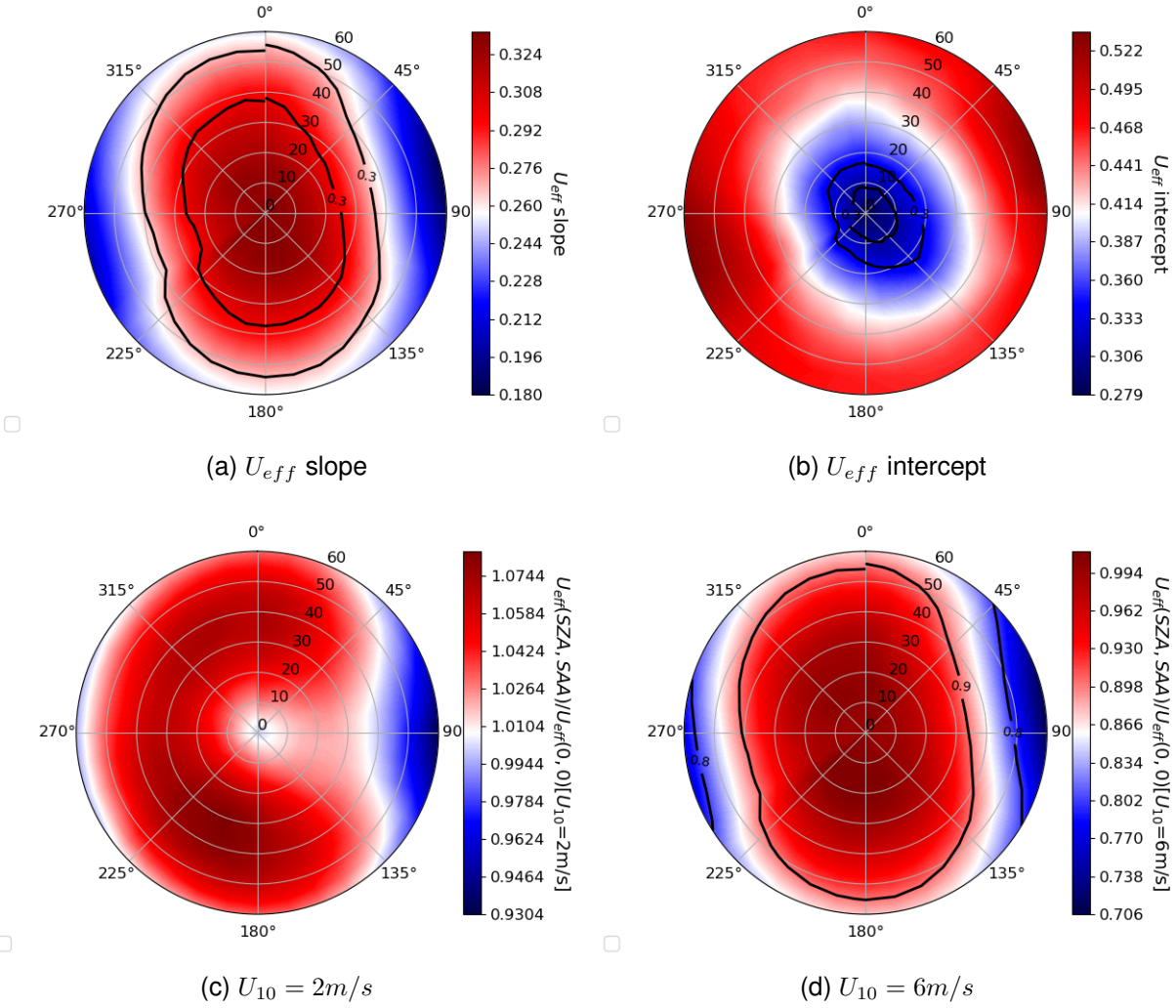
on Fig. 4c displays a clear double plume but unrealistic scenario, whereas the middle image scenario in Fig. 4b is more realistic and still shows a close to double plume and a structure very different from a nadir view modelling.

The angular consideration in the simulated methane enhancement map has a direct impact on the U_{eff} calibration. Figure 5 includes the slope and intercept of U_{eff} coefficients assuming a linear regression and the nadir view. It also includes the relative error from nadir U_{eff} at a fixed value of U_{10} equal to 2 and 6 m/s. The polar plot represents the solar angle dependence of these coefficients.

These examples show that above a sun zenith angle of 60°, errors are 10% at 2 m/s and above 20% for higher winds. At these large angles, U_{eff} is generally underestimated, and a correction of this effect rather than its uncertainty would be more appropriate.

This is included as part of the uncertainty analysis, but in general, we propose a different U_{eff} calibration based on a more realistic scenario for different cases. Furthermore, the generation of these realistic simulations implies an improved PoD assessment or training dataset that will need to be included in future updates of the validation methodology (see Section 5).

At TROPOMI scale, the VZA influence on the effective wind speed calibration can be explored by studying the impact of TROPOMI pixel area on the obtained calibration results. Efforts are underway at SRON to study the impact of resampling WRF outputs to TROPOMI footprints (also accounting for averaging kernels) on effective wind calibration results for IME.Currently, Schuit et al. (2023) quantifications rely on a calibration that was obtained without resampling the 4 $\times$ 4 km^2 regular WRF grid to TROPOMI pixel geometry. Refinements can be explored based on the efforts described above. This contribution to the uncertainty is currently not included in the uncertainty ensemble.



(a) U_{eff} slope

(b) U_{eff} intercept

(c) $U_{10} = 2m/s$

(d) $U_{10} = 6m/s$

Figure 5: Polar plot of (5a) U_{eff} slope, (5b) U_{eff} intercept, (5c) U_{eff} at 2m/s and (5d) U_{eff} at 6m/s as a function of solar angles and considering nadir view. Figs 5c and 5d have been normalised to nadir solar angle. The black lines correspond to a 10% and 20% relative difference.

4 Uncertainty combination

The different contributions described in Section 3 are interrelated according to the uncertainty tree described in Figure 1.

The combination of them results in a final Q uncertainty. The combination can be based mainly on the law of propagation of uncertainty JCGM (2008a) or a Monte Carlo combination JCGM (2008b). The latter has the benefit of propagating the full distribution through the measurement equation, which implicitly accounts for the correlation of errors. Its main disadvantage is its high computational cost. Because we are not calculating the uncertainty at a pixel level but for each emission, the computational burden will be minimum, and Monte Carlo is preferred here.

At TROPOMI scale, Schuit et al. (2023) quantifications rely on an ensemble exploring five different uncertainty dimensions along step-wise linear increments over realistic intervals, for a total of 43923 element members: (1) wind speed product; (2) plume masking threshold; (3) background XCH4 values; (4) wind speed values; and (5) effective wind calibration. This grid-ensemble-based approach to quantification and uncertainty evaluation can be discussed against using an actual Monte Carlo approach, relying on probability distributions that could be evaluated following some of the ideas presented in the previous Section. The number and nature of dimensions to be explored in such a Monte Carlo approach can also be discussed.

5 A probability of detection framework for methane plume imagers

5.1 Probability of detection methodology

The definition of a probability of detection (PoD) at a global level is desired. This is key to understanding the percentage of methane emissions that can be detected for different satellites, scenes, or seasons.

The strategy at this initial stage is to define an optimised process that is capable of retrieving a flux rate validation and PoD with the minimum computational cost. The intention is to apply as an automatic process during nominal retrieval. That is, rather than a global picture of PoD, we seek specific results for each reported emission. Notification and potential mitigation of methane emissions require specific verification. This might be of relevant importance for current and upcoming legislative frameworks.

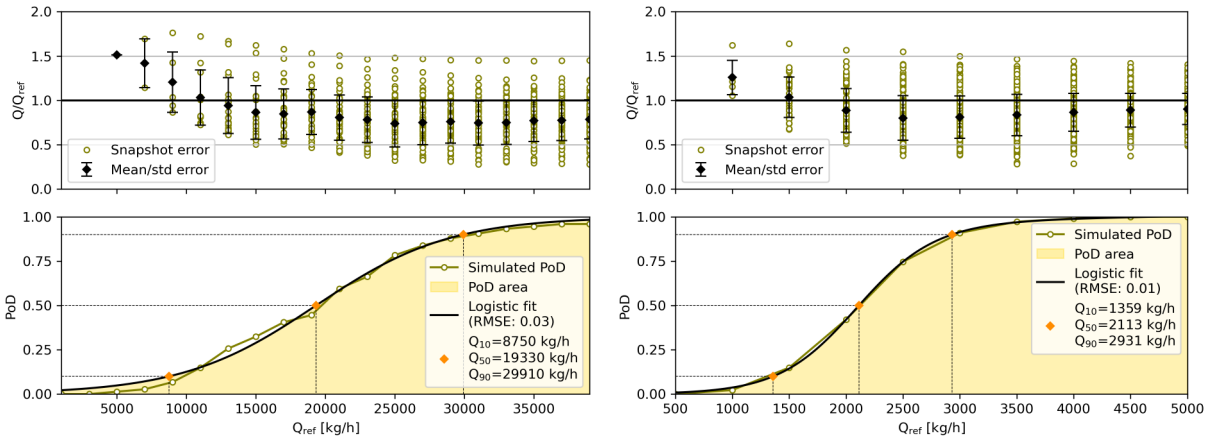
The algorithm is defined in several filtering steps to optimise processing time. It starts with a small subset of only 14 $\Delta XCH4$ simulated maps at a specific flux rate. The selected flux rate is set as very large (e.g., for ocean, we typically set Q above 40 t/h, for land, 10 t/hr). The process is discarded if less than 2 plumes are detected at this flux rate. Note that the same detection and quantification methodology is used in the simulated maps as in nominal retrieval.

The next stage repeats the same process (with 14 $\Delta XCH4$ simulated maps) but at several flux rates defined by the user based on the a priori expected broad range. If we can define a complete curve with PoD at 10,50 and 90%, we define a fine validation based on the fitted curve for 1% $<$ PoD$<$ 99% and a larger number of $\Delta XCH4$ simulated maps.

Figure 6 presents the results applied to offshore and onshore sites.

The example shows the validation of the flux rates with expected underestimation in both cases and an overestimation for low values of PoD. Thus, the combined validation and the PoD analysis are important for the interpretation of the results.

The PoD for the two sites shows different values as expected between sites with strong and low radiances, respectively. We have defined a value of Q_{10}, Q_{50}, and Q_{90} that represent the



(a) Offshore (b) Onshore

Figure 6: Flux rate validation and PoD curve for (6b) an onshore and (6a) offshore sites.

different PoD probabilities and synthesise the information for the user.

This is useful not only for monitoring a specific site and event, but also to verify the evolution over a long temporal period. It can be helpful to understand changes in the area, verify mitigation actions, or study cyclic periodicity.

Figure 7 refers to the annual cycle of PoD that can be found usually in offshore areas for the Sentinel 2 mission.

The expected PoD of Sentinel 2 over offshore area is expected to be large (typically over 5t/hr for Q_{10}). Sentinel 2 overpasses the selected location with a consistent viewing angle of approximately 7 °and 110 °zenith and azimuth, respectively. The solar pattern throughout the year produces that the highest radiance is close the summer because the angular configuration gets closer to the sunglint. For example, for Q_{50} the values can be below 10t/hr during the summer but above 20 t/hr during the winter.

Improvements will consist in improving the methodology (e.g., considering the angles as described in Subsection 3.13) and expanding the database of simulations (see Subsection 5.2) so that the selected simulations for validation are close to the real scenario.

5.2 Generating a large simulation database

Access to a large and realistic simulation database is key to providing validation and PoD assessment over multiple scenarios. Moreover, some of the uncertainty contributions in Section 3 require access to these simulations.

At this first stage, we have focused on setting up a robust simulation set-up. We are currently running the Weather Research and Forecasting Model (WRF) Version 4.2.2 on an HP DL380P

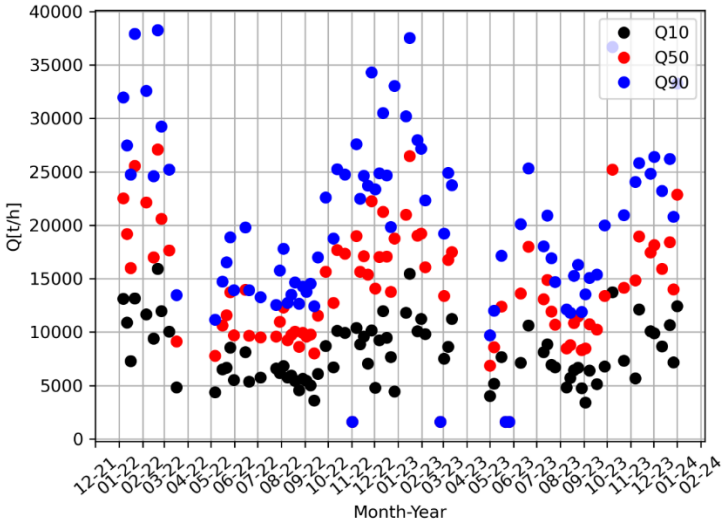
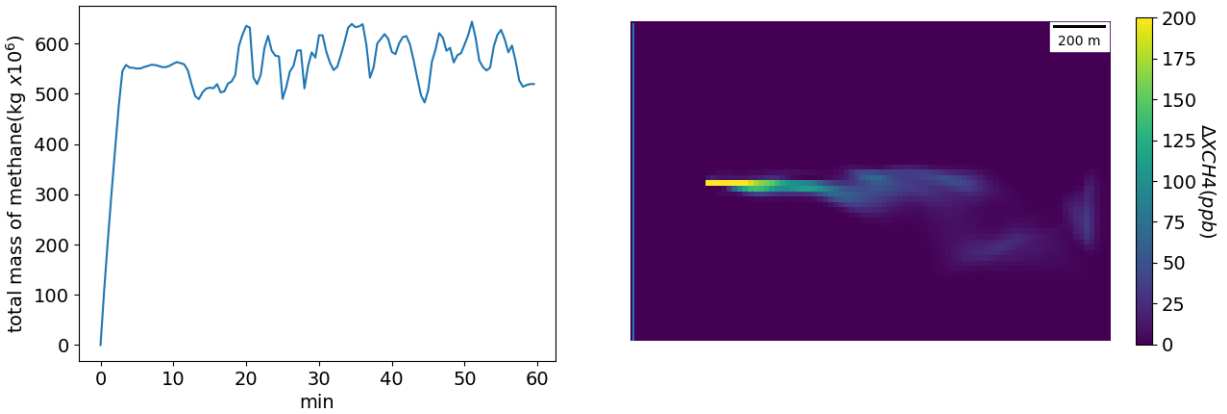


Figure 7: Temporal example of Q_{10}, Q_{50}, and Q_{90} for Sentinel 2 observations over a two year period in an offshore site at lat,lon: 9.2011°, 50.7129°.



(a) Total mass (b) ΔXCH_4

Figure 8: Total mass in the domain (8a) and (8b) ΔXCH_4 snapshot (approx 27min from start) for simulation at 9 m/s, scaled to 1000 kgh/h.

G9 server with 2 x Intel Xeon E5-2690 V4 (14 cores) and 256GB RAM. This dedicated resource (only internally used by the LARS group) ensures the ability to run of simulations on-demand.

The setup currently sets a 2h simulation with snapshots every 30s. We define two domains with a spatial resolution of 60 and 20m, respectively. The first domain has a dimension of 7.2×7.2km and the second domain 1.2×1.8km. The vertical dimension consists of 75 layers up to 2km (i.e. resolution of 20m).

Figure 8 illustrates an example of the evolution of the total mass in the domain and ΔXCH_4 map.

The figure shows a rapid increase in mass, followed by a stabilisation period after approximately 15 minutes. In this first period, Q can be obtained from the slope of the IME.

At the time of writing, we are building this database. We have selected different geosthrophic winds ranging from 1 to 9 m/s. Thus, it results in an approximate range of 0.5 to 6 m/s for U_{10}. The simulation database can be scaled to the desired flux rate and the angular conditions for both solar and viewing angle can be adapted as explained in subsection 3.13.

Further improvements involve the generation of more realistic enhancements, such as:

- Additional temperature profiles, latent heat flux, etc.
- Diffuse simulations: based on multiple sources over the same domain.

6 Summary

This document aims to describe an uncertainty framework for plume imagers and flux mappers. Section 2 begins with the generation of an uncertainty tree. This connects the measurement equation with the different uncertainty sources in an illustrative manner.

Section 3 describes in detail each one of these contributions and proposes a methodology for their assessment. At this point, we have also initiated the implementation, including final results, of some of these contributions (e.g., 3.13).

Section 5 outlines a first approach for estimating the probability of detection of methane plumes. The ultimate goal is to be able to define the PoD in an efficient manner and on a global scale. In this first stage, we outline the first framework proposal and describe the simulation database currently being generated.

This document combines deliverables D4.1 and D4.2 so that both the methodology and implementation of uncertainty contributions are developing in parallel.

7 Towards a community consensus framework and guidelines for methane flux rate uncertainty

This is a first version (out of two), and several sections are expected to be modified in the final version. The main work will be based on the independent assessment of the different uncertainty contributions presented in this first document. In addition, we will undergo a refinement of the PoD framework, an improved database of simulations, and its application in a real context.

The result of this work must be the basis for a community consensus framework and guidelines for the assessment of methane flux rate uncertainty. At this initial point of the study, we

recommend the definition of an uncertainty tree for methane flux rates as defined in Section 2. Importantly, this uncertainty tree needs to go back to the SI reference. The current consensus is that traceability should be linked to methane-controlled releases. Section 3.12 describes an initial approach to estimate this link between simulations and reference emisisons. However, more work is needed to define this link and should be one of the main emphasis during the next stage of the study.

For each of the contributions, this initial document describes some methodologies and implementations to assess them. The next phase of the study should focus on a detailed assessment of these contributions with a particular focus on those that affect the parameter U_{eff} (and U_{10}). The study is not expected to focus on the uncertainty associated with X_{CH4}, but on ad hoc cases where the retrieval results in outliers. Thus, in terms of a consensus framework, it might be appropriate to take this latter case as a separate metric (e.g. definition of uncertainty and outliers as two separate metrics).

8 References

Ayasse, A., Cusworth, D., Howell, K., O'Neill, K., Conrad, B., Johnson, M., Asner, G., Duren, R., 07 2024. Probability of detection and multi-sensor persistence of methane emissions from coincident airborne and satellite observations. EarthArxiv.

BIPM, I., IFCC, I., ISO, I., IUPAP, O., 2011. Evaluation of measurement data—supplement 2 to the 'guide to the expression of uncertainty in measurement'—extension to any number of output quantities. JCGM 102, 2011.

Borsdorff, T., aan de Brugh, J., Schneider, A., Lorente, A., Birk, M., Wagner, G., Kivi, R., Hase, F., Feist, D. G., Sussmann, R., Rettinger, M., Wunch, D., Warneke, T., Landgraf, J., 2019. Improving the tropomi co data product: update of the spectroscopic database and destriping of single orbits. Atmospheric Measurement Techniques 12 (10), 5443–5455.
URL https://amt.copernicus.org/articles/12/5443/2019/

Carvalho, D., 2019. An assessment of nasa's gmao merra-2 reanalysis surface winds. Journal of Climate 32 (23), 8261 – 8281.
URL https://journals.ametsoc.org/view/journals/clim/32/23/jcli-d-19-0199.1.xml

Davis, N. N., Badger, J., Hahmann, A. N., Hansen, B. O., Mortensen, N. G., Kelly, M., Larsén, X. G., Olsen, B. T., Floors, R., Lizcano, G., Casso, P., Lacave, O., Bosch, A., Bauwens, I., Knight, O. J., van Loon, A. P., Fox, R., Parvanyan, T., Hansen, S. B. K., Heathfield, D., Onninen, M., Drummond, R., 2023. The global wind atlas: A high-resolution dataset of climatologies and associated web-based application. Bulletin of the American Meteorological Society 104 (8), E1507 – E1525.
URL https://journals.ametsoc.org/view/journals/bams/104/8/BAMS-D-21-0075.1.xml

Fahlen, J. E., Brodrick, P. G., Coleman, R. W., Elder, C. D., Thompson, D. R., Thorpe, A. K., Green, R. O., Green, J. J., Lopez, A. M., Xiang, C., 2024. Sensitivity and uncertainty in matched-filter-based gas detection with imaging spectroscopy. IEEE Transactions on Geoscience and Remote Sensing 62, 1–10.

Gorroño, J., Guanter, L., Roger, Javier amd Zhou, S., Maasakkers, J. D., Dogniaux, M., Zhang, X., Akani-Guery, J., Groshenry, A., Ramier, A., Buchwitz, M., 2025. Report on uncertainties as used in data products wp410 - deliverable d4.5. Tech. rep., ESA project MEDUSA, Version 1.0, SRON-ESG-RP-2025-006.

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J., Cogliati, S., Colombo, R., 2021. Mapping methane point emissions with the prisma spaceborne

imaging spectrometer. Remote Sensing of Environment 265, 112671.
URL https://www.sciencedirect.com/science/article/pii/S0034425721003916

Guanter, L., Segl, K., Sang, B., Alonso, L., Kaufmann, H., Moreno, J., Jul 2009. Scene-based spectral calibration assessment of high spectral resolution imaging spectrometers. Opt. Express 17 (14), 11594–11606.
URL http://www.opticsexpress.org/abstract.cfm?URI=oe-17-14-11594

JCGM, 2008a. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. Tech. Rep. 100.

JCGM, 2008b. Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method. Tech. Rep. 101.

MacLeod, D. A., Cloke, H. L., Pappenberger, F., Weisheimer, A., 2016. Improved seasonal prediction of the hot summer of 2003 over europe through better representation of uncertainty in the land surface. Quarterly Journal of the Royal Meteorological Society 142 (694), 79–90.
URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2631

Markham, B. L., 1985. The landsat sensors' spatial responses. IEEE Transactions on Geoscience and Remote Sensing GE-23 (6), 864–875.

Mittaz, J., Merchant, C. J., Woolliams, E. R., may 2019. Applying principles of metrology to historical earth observations from satellites. Metrologia 56 (3), 032002.
URL https://dx.doi.org/10.1088/1681-7575/ab1705

Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., Thépaut, J.-N., 2021. Era5-land: a state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13 (9), 4349–4383.
URL https://essd.copernicus.org/articles/13/4349/2021/

Sánchez-García, E., Gorroño, J., Irakulis-Loitxate, I., Varon, D. J., Guanter, L., 2022. Mapping methane plumes at very high spatial resolution with the worldview-3 satellite. Atmospheric Measurement Techniques 15 (6), 1657–1674.
URL https://amt.copernicus.org/articles/15/1657/2022/

Schuit, B. J., Maasakkers, J. D., Bijl, P., Mahapatra, G., van den Berg, A.-W., Pandey, S., Lorente, A., Borsdorff, T., Houweling, S., Varon, D. J., McKeever, J., Jervis, D., Girard, M., Irakulis-Loitxate, I., Gorroño, J., Guanter, L., Cusworth, D. H., Aben, I., 2023. Automated detection and monitoring of methane super-emitters using satellite data. Atmospheric Chemistry and Physics 23 (16), 9071–9098.
URL https://acp.copernicus.org/articles/23/9071/2023/

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., Thépaut, J.-N., 2024. The era5 global reanalysis from 1940 to 2022. Quarterly Journal of the Royal Meteorological Society n/a (n/a).
URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4803

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., Huang, Y., 2018. Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes. Atmospheric Measurement Techniques 11 (10), 5673–5686.
URL https://amt.copernicus.org/articles/11/5673/2018/

Varon, D. J., Jervis, D., McKeever, J., Spence, I., Gains, D., Jacob, D. J., 2021. High-frequency monitoring of anomalous methane point sources with multispectral sentinel-2 satellite observations. Atmospheric Measurement Techniques 14 (4), 2771–2785.
URL https://amt.copernicus.org/articles/14/2771/2021/