Understanding the potential of Sentinel-2 for monitoring methane point emissions

Javier Gorroño[1], Daniel J. Varon[2], Itziar Irakulis-Loitxate[1], and Luis Guanter[1,3]

[1]Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València, València, Spain
[2]School of Engineering and Applied Science, Harvard University, Cambridge, 02138, USA
[3]Environmental Defense Fund, Reguliersgracht 79, 1017 LN Amsterdam, The Netherlands

Correspondence: Javier Gorroño (jagorvie@upv.es)

Abstract. The use of satellite instruments to detect and quantify methane emissions from fossil fuel production activities is highly beneficial to support climate change mitigation. Different hyperspectral and multispectral satellite sensors have recently shown potential to detect and quantify point-source emissions from space. The Sentinel-2 (S2) mission, despite its limited spectral design, supports the detection of large emissions with global coverage and high revisit frequency thanks to coarse
5   spectral coverage of methane absorption lines in the shortwave infrared. Validation of S2 methane retrieval algorithms is instrumental in accelerating the development of a systematic and global monitoring system for methane point sources. Here we develop a benchmarking framework for such validation. We first develop a methodology to generate simulated S2 datasets including methane point-source plumes. These benchmark datasets have been created for scenes in three oil and gas basins (Hassi Messaoud, Algeria; Korpeje, Turkmenistan; Permian Basin, USA) under different scene heterogeneity conditions and
10   for simulated methane plumes with different spatial distributions. We use the simulated methane plumes to validate the retrieval for different flux rate levels and define a minimum detection threshold for each case study. The results suggest that for homogeneous surfaces, the detection limit of the proposed S2 methane retrieval ranges from $1000\,\mathrm{kg\,h^{-1}}$ to $2000\,\mathrm{kg\,h^{-1}}$, whereas for areas with large surface heterogeneity, the retrieval can only detect plumes in excess of $5000\,\mathrm{kg\,h^{-1}}$. The different sources of uncertainty in the flux rate estimates have also been examined. Dominant quantification errors are either wind-related or plume mask-related, depending on the surface type. Uncertainty in wind speed, both in the 10-m wind (U_{10}) and in mapping
15   U_{10} to the effective wind (U_{eff}) driving plume transport, is the dominant source of error for quantifying individual plumes in homogeneous scenes. For heterogeneous scenes, the surface structure underlying the methane plume affects the plume masking and can become a dominant source of uncertainty.

1   Introduction

Methane (CH_4) is the second most important anthropogenic greenhouse gas after carbon dioxide (CO_2). Its global warming potential is 86 times that of CO_2 over 20 years, and it has a relatively short lifetime in the atmosphere (9±1 years) (Etminan

et al., 2016). Fossil fuel production activities represent between 30% and 42% of the total global anthropogenic emissions (Saunois et al., 2020). These emissions come from oil and gas (O&G) production infrastructure (e.g. wells, compressor stations)

25 (Lyon et al., 2016) and coal mines. The facility typically emits a gas plume of highly concentrated methane from a small source area (referred to as "point emissions")(Duren et al., 2019). The mitigation of a large fraction of these emissions is technically feasible and cost-effective and represents a key means to reduce the amount of greenhouse gases in the atmosphere (CCAC, 2021).

Satellites measuring backscattered solar radiation in the short-wave infrared (SWIR) spectral region can detect subtle signal

30 variations linked to methane absorption (Jacob et al., 2022). Area flux mappers such as GOSAT and TROPOMI provide daily mapping of methane concentrations at a large spatial scale (Qu et al., 2021). This provides long-term methane trends at a global and regional scales, the latter being instrumental in identifying hotspot areas of anthropogenic methane emissions. At a finer spatial resolution ($3.7\,\mathrm{m}$ to $100\,\mathrm{m}$), several studies have demonstrated that point-source imager satellites can identify individual emissions with detection thresholds from $100\,\mathrm{kg\,h^{-1}}$ to $10000\,\mathrm{kg\,h^{-1}}$. For example, Varon et al. (2018) described

35 a set of algorithms to detect and quantify methane point sources focusing on the private GHGSat constellation. More recent studies have shown the feasibility of quantifying methane point sources from hyperspectral and multispectral missions such as PRISMA (Guanter et al., 2021), Sentinel-2 (S2) (Varon et al., 2021), Landsat-8 (Ehret et al., 2022) and WorldView-3 (Sánchez-García et al., 2021). These recent advances have revealed the existence of large point emissions from O&G production regions (Irakulis-Loitxate et al., 2021, 2022b), coal carbon mines (Guanter et al., 2021) and O&G offshore platforms (Irakulis-Loitxate

40 et al., 2022a).

Hyperspectral satellite imagery (e.g. PRISMA) is capable of resolving fine methane absorption features in the SWIR and typically provides a detection limit from roughly $100\,\mathrm{kg\,h^{-1}}$ to $500\,\mathrm{kg\,h^{-1}}$(Jacob et al., 2022). However, current hyperspectral missions have a low revisit frequency that limits their usefulness for monitoring on a global scale. By contrast, the S2 and Landsat missions provide a unique combination of global coverage, high-quality calibration, and high temporal revisit (Gascon

45 et al., 2017; ESA, 2022; Micijevic et al., 2022), but the sensitivity of their coarse SWIR bands to methane absorption is limited so that expected detection thresholds can be slightly over $2000\,\mathrm{kg\,h^{-1}}$ as first suggested by Varon et al. (2021) or slightly lower (at $1800\,\mathrm{kg\,h^{-1}}$) under favourable conditions as suggested by Irakulis-Loitxate et al. (2022b). The usefulness of the S2 mission for mapping point-source emissions has been already demonstrated at the Korpeje O&G region in Turkmenistan by Varon et al. (2021) and Irakulis-Loitxate et al. (2022b), and at a set of O&G basins across the globe by Ehret et al. (2022).

50 The development of global and systematic monitoring of methane point sources requires the accurate validation of the retrieval algorithm in different conditions such as solar illumination, surface heterogeneity, or atmospheric composition. These validation exercises can be either based on in-situ controlled emissions or simulated analysis. The former approach provides the most accurate benchmark and real conditions but is limited to a small set of data points due to high logistic costs and the negative impact on the atmosphere (Sherwin et al., 2022). The latter approach, despite its challenging reproduction of real

55 scene conditions, can provide validation scenarios of methane point-source emissions for any satellite at a global level.

These simulated approaches can be based on an end-to-end framework that covers the solar irradiance, atmosphere, surface and satellite characteristics (Jongaramrungruang et al., 2021). This is a convenient method to study the sensitivity of different

satellite characteristics such as sensor noise or spectral resolution. However, the more subtle specifications of the instrument (e.g. per-pixel viewing dependence) or scene characteristics might be difficult to reproduce. Cusworth et al. (2019) introduced

60   a detailed study of the detection and quantification of methane point sources from several hyperspectral missions based on their spectral and noise characteristics. This alternative approach used simulated methane plumes and incorporated their transmittance into simulated satellite acquisitions. By doing so, the satellite radiance observations were not altered while the plume enhancement was integrated into the scene. We adopt this latter methodology here, and refine it in section 3 to account for a compensation term that corrects for the convolution of the methane transmittance signal after its band integration. This error

65   is present for every type of instrument at different levels, but for multispectral missions with a large bandwidth such as S2, a larger impact is expected.

Our simulated datasets validate the retrieval algorithm proposed in section 4 for a set of sites and methane plumes presented in section 2. These plumes represent the temporal evolution of a point source plume with changing spatial distribution. The selected sites are representative of well-known O&G basins with different scene heterogeneity.

70   The validation process presented in subsection 6.2 generates several datasets at different flux levels. This methodology validates the methane flux rate at different levels and, at the same time, defines the detection limit of the instrument for a set of sites and methane plumes. The results highlight several uncertainty contributions associated with the methane flux that we disentangle into different sources in subsection 6.3.

2   Sentinel-2 data and sites of study

75   The first satellite unit (S2A) of the S2 mission was launched in 2015 and, together with the second satellite unit S2B launched in 2017, provides continuous monitoring of terrestrial surfaces and coastal waters at a global scale with better than five-days revisit. Both satellites carry on-board the Multi-Spectral Instrument (MSI) with 13 spectral bands in the visible and near-infrared (VNIR) and the shortwave infrared (SWIR) at spatial resolutions of $10\,\mathrm{m}$, $20\,\mathrm{m}$ or $60\,\mathrm{m}$. Top-of-Atmosphere (TOA) reflectance images are delivered to the public after geometric and radiometric calibration (L1C products) Products are free

80   open-access from the Sentinels Scientific Data Hub in a set of ortho-images (in UTM/WGS84 projection) of about $100\times 100\,\mathrm{km}^2$ Gascon et al. (2017).

The bands in the SWIR region, B11 and B12, measure the TOA radiance at an approximate $20\,\mathrm{m}$ spatial resolution. Although they were not designed to detect and quantify methane point-source emissions, these two bands are sensitive to methane absorption and have shown capability to monitor large emissions at a global scale (Varon et al., 2021; Ehret et al., 2022).

85   Figure 1 illustrates the spectral response of bands B11 and B12 for both satellite units together with the methane transmittance for a surface concentration of $1904\,\mathrm{ppm}$ and air-mass factor (AMF) of 2.

Both of these bands cover a small spectral region of methane absorption. In the case of B12, it is significantly ~~larger~~ than B11. Despite carrying the same MSI instrument, S2A and S2B do not exactly match the same spectral response. Although the differences are within the specifications, it shows that the S2A B12 is slightly more sensitive to methane absorption as

90   compared to S2B satellite unit.
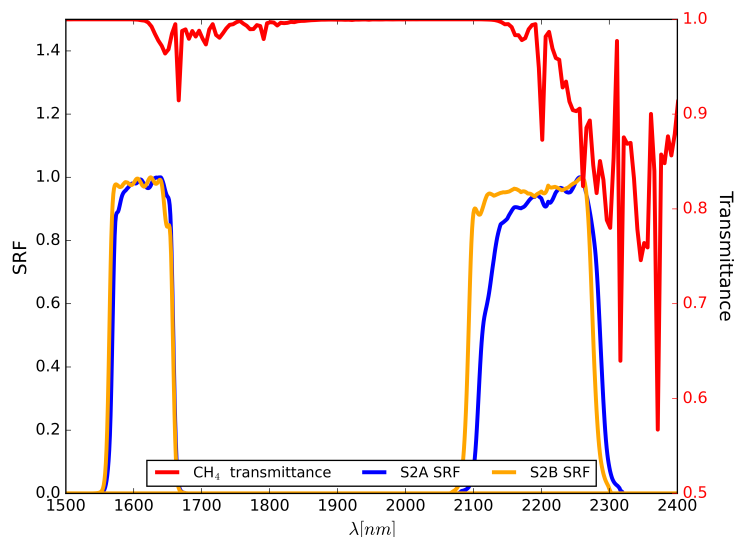
Figure 1. S2A and S2B spectral response functions of bands B11 and B12 ~~combined with the~~ methane transmittance for a surface concentration of $1904\,\mathrm{ppm}$ and AMF of 2.

Figure 2 shows the selected areas of study: Korpeje in Turkmenistan, Hassi Messaoud in Algeria, and Permian Basin in the USA. The specific locations correspond to an area of $1500 \times 1500\,\mathrm{m}^2$ centered at the latitude and longitude 38.493918 °, 54.216550°(Korpeje), 31.82652 °, 6.15011°(Hassi Messaoud) and 31.72070°, -102.28663°(Permian basin).

These areas have been selected (1) for being well-known O&G activity locations, (2) for showing persistent high methane concentrations as observed by TROPOMI for years, and (3) because they are well-studied reference sites in the literature (Varon et al., 2019; 2; Zhang et al., 2020; Varon et al., 2021; Cusworth et al., 2021; Guanter et al., 2021; Irakulis-Loitxate et al., 2022b; Lauvaux et al., 2022). In the case of Korpeje, the selected location is near a compressor station with high emission fluxes and ~~high emission~~ frequency but far enough away so that no real plumes contaminate the test images. In Algeria, the Hassi Messaoud oil field is located in a desert area with extremely uniform spatiotemporal conditions. Finally, the location selected in the Permian basin is characterized by including a high number of wells and pipeline crossings. In addition, the scene might suffer from temporal variations that include atmospheric conditions, vegetation changes or human activities in the area. In the literature, other study areas widely known for high emission rates or high emission potential can be found, such as the Shanxi region in China (Guanter et al., 2021) or the O&G fields in Canada (Chan et al., 2020; Williams et al., 2021; Tyner and Johnson, 2021). However, the high surface heterogeneity and cloud cover throughout the year over these areas limits the application of methane retrieval methods with multispectral instruments such as S2.

The selected S2 acquisitions for the simulation of methane plumes are the S2A tile 40SBH acquired on the 18th of July 2021 for the Korpeje site, the S2A tile 32SKA acquired on the 2nd of July 2021 for Hassi Messaoud, and S2A tile 13SGR on the 4th of July 2021 for Permian basin. A temporal normalisation selection searches for the closest cloud-free overpass with the

(a) Korpeje (Turkmenistan)	(b) Hassi Messaoud (Algeria)	(c) Permian basin (USA)
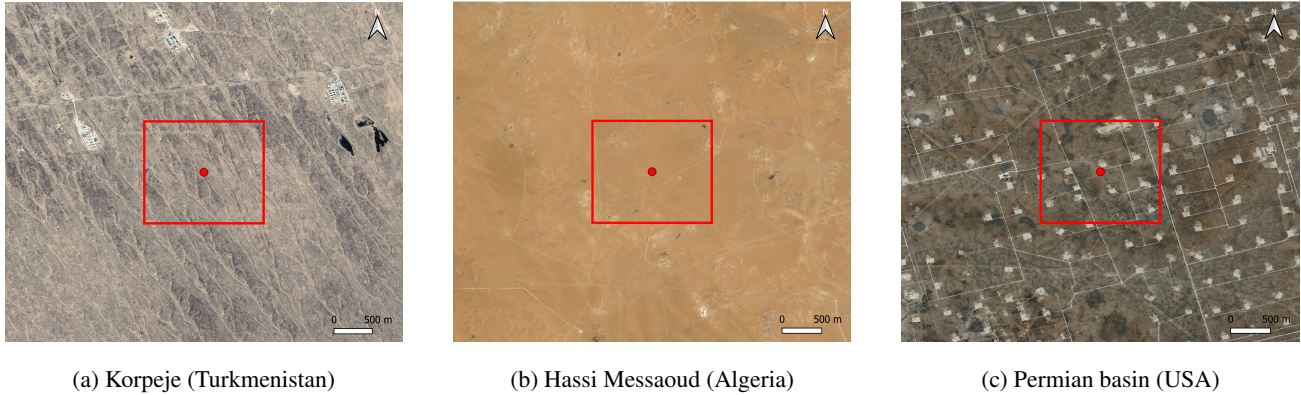
Figure 2. RGB image of the selected areas of 2a Korpeje (Turkmenistan), 2b Hassi Messaoud (Algeria), and 2c Permian basin (USA). The selected area (red box) corresponds to $1500 \times 1500 \text{ m}^2$ centered at the latitude and longitude 38.493918 °, 54.216550°(Korpeje), 31.82652 °, 6.15011°(Hassi Messaoud), and 31.72070°, -102.28663°(Permian basin). Source map from © Google Earth and © Microsoft Bing.

same viewing overpass (see section 4). If the closest cloud-free overpass is selected, this will typically correspond to a different

110 satellite unit (due to orbit phase differences between S2A and S2B satellite units) with slightly different spectral responses. Thus, this corresponds to S2B acquisitions on the 13th of July 2021 (Korpeje), 27th of June 2021 (Hassi Messaoud), and the 9th of July 2021 (Permian basin).

3 Simulating Sentinel-2 observations of methane plumes

We used the Weather and Research Forecasting Model in large-eddy simulation model (WRF-LES) to generate realistic

115 methane plumes (three-dimensional mass distributions) in the atmospheric boundary layer over time, following the approach of Varon et al. (2018). The simulation was conducted at $30 \times 30 \text{ m}^2$ spatial resolution with an initial geostrophic wind field of 3.5 m s^{-1} and a sensible heat flux $H = 300 \text{ W m}^{-2}$.

The original resolution of the simulation has been resampled to the 20-m spatial resolution of S2. This resampling is performed at each layer of the vertical grid to minimize interpolation error. The simulation is performed under a reference flux

120 that is then scaled linearly to the desired level following Guanter et al. (2021) and Sánchez-García et al. (2021).

The resulting 3-D methane distributions can be vertically integrated in order to obtain a 2D mass field of column mass enhancement $\Delta\Omega$ in kg m^{-2}. These maps are further converted into column-averaged dry molar mixing ratio ΔX_{CH4}(unitless) maps following the relationship:

$$\Delta\Omega = \frac{M_{CH_4}}{M_a}\Omega_a\Delta X_{\text{CH4}} \tag{1}$$

125 where M_a and M_{CH_4} refer to the molar mass of dry air and methane respectively [kg mol^{-1}], and Ω_a refers to the column of dry air [kg m^{-2}].

(a) Plume 1   (b) Plume 2   (c) Plume 3

(d) Plume 4   (e) Plume 5
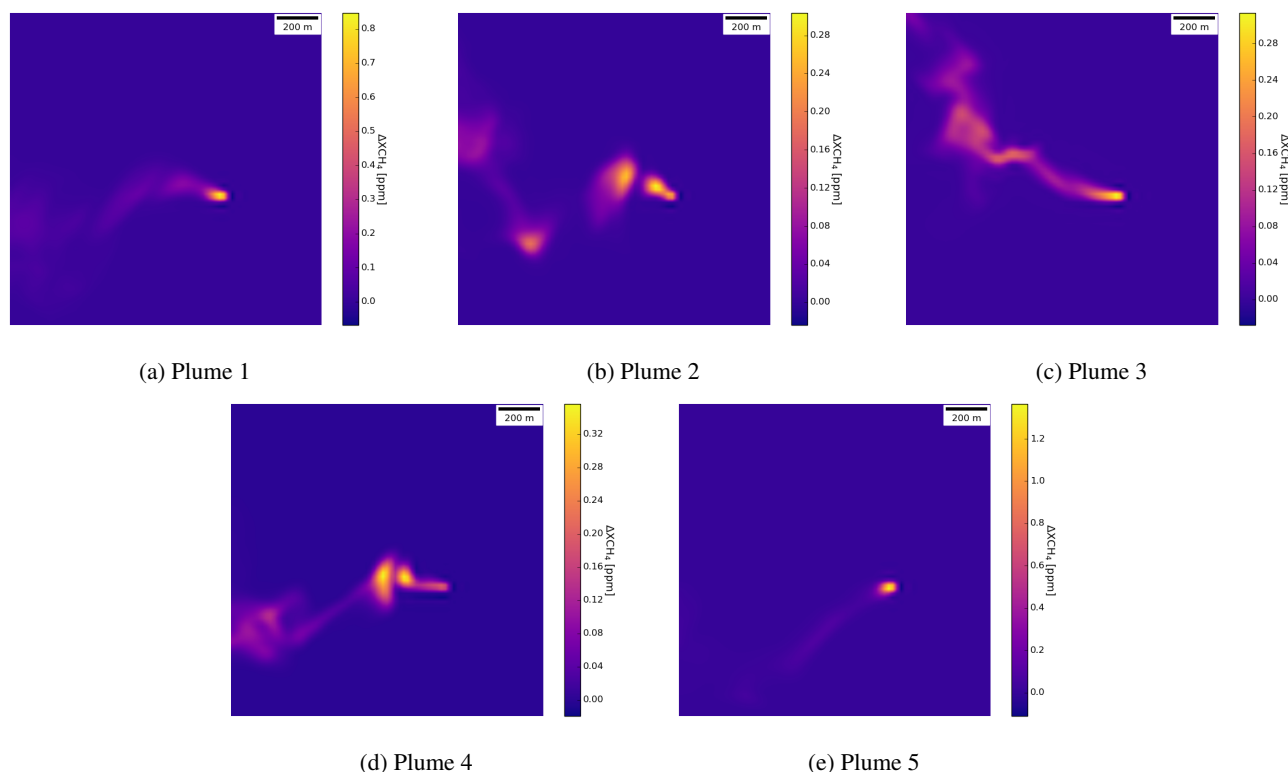
Figure 3. Methane enhancement map ΔX_{CH4} of the 5 plumes considered for the simulations scaled at a flux rate $Q = 1000\,\mathrm{kg\,h^{-1}}$ after resampling to 20-m resolution.

From the WRF-LES simulations, we have selected 5 different snapshots that have been used as representative shapes of methane plumes for the simulations (see Figure 3).

The methane plume transmittance map as a function of wavelength, $T_{\mathrm{plume}}(\lambda)$, can be modelled at a fine spectral resolution.

130 Similarly to Cusworth et al. (2019), the spectral optical depth $\tau(\lambda)$ of each atmospheric layer can be described as the multiplication of the HITRAN absorption cross-sections (σ_{H}; Kochanov et al., 2016), the vertical column density of dry air, and the methane volume mixing ratio enhancement (ΔVMR) from the LES.

On the other hand, the resulting map of ΔX_{CH4} can be directly converted into $T_{\mathrm{plume}}(\lambda)$ by using a predefined look-up table (LUT) that establishes the relationship between the two quantities at different AMF values with MODTRAN (Berk et al.,

135 2014). This alternative approach has been proven to produce similar results with a much lower computational cost. Thus, this is the method proposed when a large number of simulations are required and implemented for the datasets in Gorroño et al. (2022).

Once the transmittance of the methane plumes has been calculated, it is convolved with the S2 spectral response to obtain a methane plume transmittance map for each spectral band. This can be directly multiplied with the original S2 TOA radiance

140 maps. This approach has the largest advantage since it reproduces the exact same observation conditions as the original obser-

vation. The resulting simulation incorporates a real scene from S2, meaning that effects such as instrument noise or angular pixel dependencies are already included in the simulation. The drawback of this approach is that it becomes mathematically impossible to convolve a signal that corresponds to an already band-integrated quantity.

Recent work by Guanter et al. (2021) and Sánchez-García et al. (2021) included a compensation term in the calculated total
145 atmospheric transmittance capable of partially correcting for the bias introduced by this a-posteriori convolution of the methane transmittance for PRISMA and WV-3 simulations respectively. In those two cases, a simplified approach was selected by considering the total transmittance of the atmosphere $T_{atm}(\lambda)$ rather than reconstructing the full TOA radiance. This approach can be sufficient for an instrument that measures in narrow bands where it can be assumed that the surface reflectance and downwelling irradiance are not rapidly changing across them. However, this assumption might not hold for highly heterogeneous
150 surfaces and spectral bands that span through a larger portion of the SWIR spectrum such as S2 B12 (bandwidth of $180\,\mathrm{nm}$). In that case, it is advisable to include a reconstructed top-of-atmosphere radiance L_{TOA} in the correction. Here we will test the implications and their correction by including a term that not only includes the total transmittance but also a reconstructed L_{TOA}. The corrected L_{TOA} for the B12 band can be written as:

$$L_{B12}^{plume} = L_{B12} \left(\int_{B12} T_{plume}(\lambda)d\lambda \right) \epsilon_{convolution} \tag{2}$$

155 where T_{plume} indicates the transmittance of the methane plume and $\epsilon_{convolution}$ refers to the bias introduced by the a-posteriori convolution. Here, we model this correction as:

$$\epsilon_{convolution} = \frac{\int_{B12} L_{TOA}(\lambda)T_{plume}(\lambda)d\lambda}{\left(\int_{B12} L_{TOA}(\lambda)d\lambda\right)\left(\int_{B12} T_{plume}(\lambda)d\lambda\right)} \tag{3}$$

with the $L_{TOA}(\lambda)$ modelled as:

$$L_{TOA}(\lambda) = T_{atm}(\lambda)\rho_{surface}(\lambda)E_g(\lambda) \tag{4}$$

160 T_{atm} refers to the total transmittance through the atmosphere, $\rho_{surface}$ represents the surface reflectance, and E_g is the downwelling irradiance at the ground level. The path radiance and spherical albedo are not considered in this case since their impact in the SWIR region can be considered negligible in the correction.

Combining Eqs. 2 and 3, the term $\int_{B12} T_{plume}(\lambda)d\lambda$ is cancelled and we can also reconsider the equation as a simulated plume transmittance convolved with a simulated hyperspectral TOA radiance.

165 This mathematical framework has been implemented in a software capable of producing S2 datasets with simulated methane plumes (Gorroño et al., 2022). Its flow diagram is presented in Figure 4.

The lower part of the diagram describes the calculation of the methane plume transmittance. The software implementation simplifies the process by only considering the S2 B12 band. The impact over the rest of the bands can be considered negligible
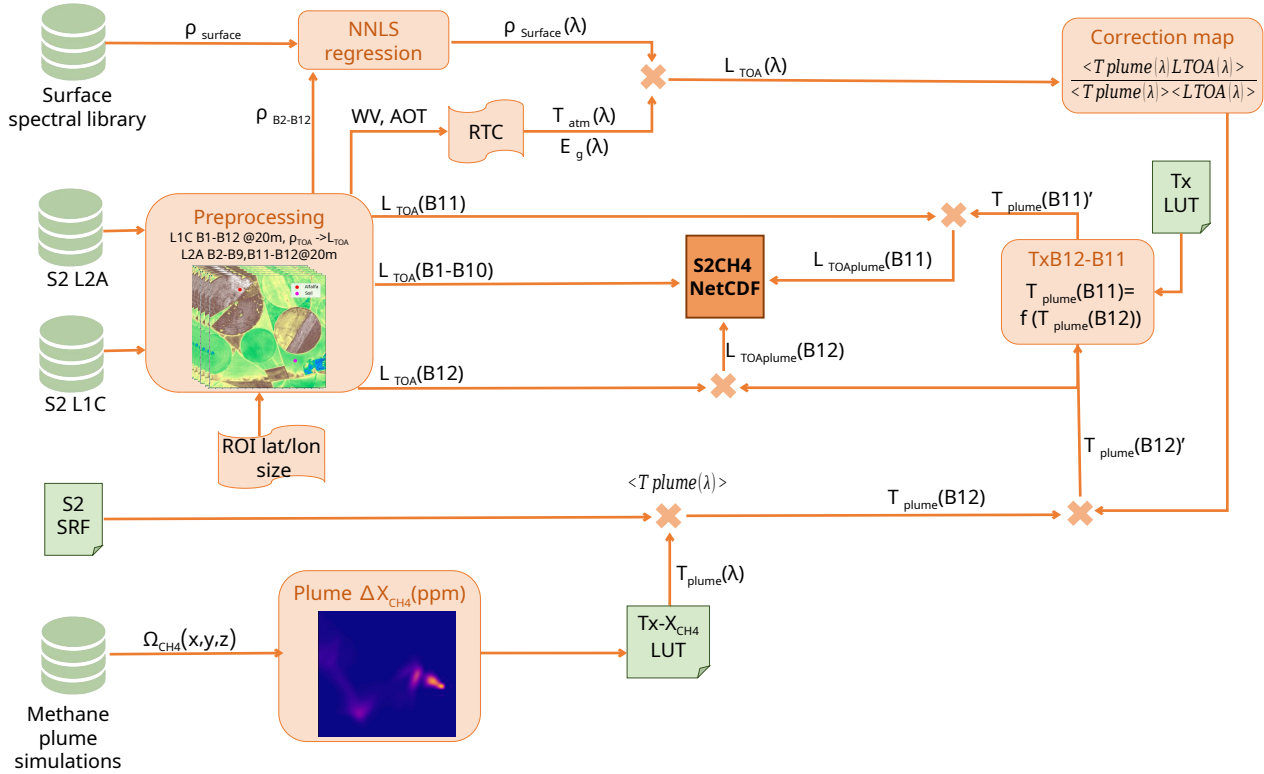
Figure 4. Flow diagram describing the process to generate S2 datasets with embedded methane plume simulations. The diagram contains Copernicus Sentinel-2 images provided by ESA.

except for the B11. At this spectral region, there is a smaller area of methane absorption that must be correctly captured in

170 the simulation. Rather than a direct calculation, it was found that a pre-compiled relationship between the methane plume transmittance of B12 and B11 was more efficient. Using the radiative transfer software MODTRAN (Berk et al., 2014), we generated different methane transmittances at different concentration values. Further convolution of the transmittance by the B11 and B12 spectral response established the mathematical dependency.

The upper part of Figure 4 captures the calculation process of L_{TOA}. This requires the processing of the S2 L2A product and

175 obtains the values for surface reflectance, water vapour, and AOT. The latter two are used to parameterize a radiative transfer model based on Libradtran (Emde et al., 2016; Mayer and Kylling, 2005) that results in an estimated total transmittance through the atmosphere T_{atm} and downwelling irradiance E_{g}. Although the simulation can be performed at a pixel-level to reduce the computational cost, the radiative transfer is only run for a mean value over the area of interest under the assumption that atmospheric variations are small in the vicinity area. The surface reflectance for each S2 band is combined with a set

180 of spectral signatures from the USGS speclib (Kokaly et al., 2017). A non-negative linear regression (NNLS) estimates the hyperspectral surface reflectance.

4 Retrieval methodology

The estimation of methane concentration enhancement is typically based on the fitting of high-spectral-resolution observations in the SWIR spectral region against a modelled radiance spectrum in a single overpass (Thorpe et al., 2014; Jacob et al., 2016).

For multispectral instruments, the SWIR spectral region cannot be resolved in detail, but it is still possible to retrieve methane concentration enhancements from an estimated methane transmittance image. This methane transmittance is calculated as the ratio between the radiance at a spectral channel affected by methane absorption and a methane-free reference band. Mathematically the plume transmittance can be described as:

$$T_{\text{plume}}(\lambda) \sim \frac{L}{L_{\text{ref}}} = e^{-\text{AMF} \cdot \sigma_{\text{CH4}} \cdot \Delta X_{\text{CH4}}} \qquad (5)$$

where L and L_{ref} represent the radiance of the methane-sensitive band and the methane-free reference band, respectively. AMF refers to the air mass factor. The methane concentration enhancement ΔX_{CH4} is the quantity of interest and expresses the increment produced by the plume from the background methane present in the atmospheric column.

Thus, the proposed rationale to detect methane plumes for the S2 mission is based on an estimation of the plume transmission defined by the ratio:

$$\frac{\rho}{\rho_{\text{ref}}} = \frac{\rho_{\text{B12plume}} / \rho_{\text{B11plume}}}{\rho_{\text{B12plumefree}} / \rho_{\text{B11plumefree}}} \qquad (6)$$

where the subscript *plume* and *plumefree* refer to the overpasses with methane plume emissions and without it, respectively, and ρ refers to the TOA reflectance.

The methane plume quantification is obtained from the isolation of the term ΔX_{CH4} in Eq. 5. Then, the solution becomes a function of the AMF and the radiance ratio proposed in Eq. 6. A LUT with the relationship between these two quantities and the methane enhancement is established. This LUT incorporates the small but non-negligible effect of methane absorption in B11. Thus, the transmittance in Eq. 6 has been calculated as the ratio of the transmittance between B12 and B11 bands due to different methane concentrations.

In order to test the validity of this approach, the retrieval method has been tested on the datasets generated in section 3 (Gorroño et al., 2022). The test has set the temporal normalization against the same observation without plume emission. This cancels the noise of the sensor, and an ideal scenario can be tested. Figure 5 shows the test of such an ideal scenario with plume 2 (see Figure 3b) and Q of $5000 \, \text{kg} \, \text{h}^{-1}$. In the absence of noise, only the linear regression between retrieved and reference methane enhancement pixels is shown. The figure includes a simulation with no correction (i.e. $\epsilon_{\text{convolution}} = 1$), another with the correction of the $T_{\text{atm}}(\lambda)$, and a last one with the $L_{\text{TOA}}(\lambda)$ correction as explained above.

This plot shows that the methodology proposed in Figure 4 results in simulated datasets with a large agreement. The uncertainty on the methane retrieval algorithm, and especially the flux rate quantification, is expected to be considerably higher. Therefore, the potential uncertainty on the simulated reference (here better than 3%) will have a negligible impact on the validation. The results also show how the correction methodology largely improves the simulation as compared with no correction
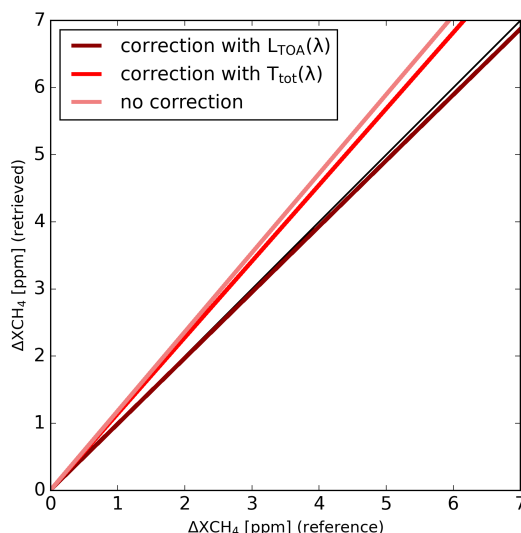
Figure 5. Linear regression of retrieved pixel values versus reference methane enhancement for the Korpeje site (see Figure 2a), plume 2 (see Figure 3b) and Q of 5000kg/h. The reference acquistion is the same as the target one with no excess methane to simulate an ideal scenario.

applied. From the three elements considered in the TOA model in Eq. 4 the term E_g plays a major role here. The S2 B12 spans from $2100\,\mathrm{nm}$ to $2300\,\mathrm{nm}$, and the downwelling irradiance in this region rapidly decreases with higher wavelengths. Thus,

215 the direct multiplication of the methane transmittance shows an overestimation if no weighting is applied (the S2 B12 band is more sensitive to methane in the $2200\,\mathrm{nm}$ to $2300\,\mathrm{nm}$ region). The term $T_{\mathrm{atm}}(\lambda)$ improves the correction but at a minor scale similarly to ρ_{surface}. However, this latter term is scene dependent, and its impact could be considerable for other type of surfaces such as snow at different granule levels.

Once the methane enhancement map has been generated, a median filter with a 3-pixel window is applied for the masking

220 process. The standard deviation σ of a region not affected by the methane plume is calculated and the 2σ (95% confidence interval) is applied as the masking threshold. The selected region for the noise threshold calculation of the plumes in Figure 3 corresponds to the right-hand area of $300 \times 1500\,\mathrm{m}^2$ out of the plume emission. This masking process is slightly different to the one proposed by Varon et al. (2018) since it applies a median filter before masking rather than masking with the background noise and applying a median filter over the Boolean result. Although this masking process alters the original enhancement map,

225 it is helpful to reduce the significant noise levels present in the S2 methane retrievals, and the resulting mask is applied to the original concentration map. A median filter (or other high-frequency filtering process) is expected to have a low impact on the plume signal since this is distributed in the atmosphere as a low-frequency signal. Nonetheless, this process could remove the signal for very narrow plumes (just one or two pixels' width).

The result of the process is a masked image that flags the potential pixels with a plume emission. The actual plume detection

230 is achieved with an automated cluster detection algorithm in the scikit-learn Python package (Pedregosa et al., 2011), which searches for clusters of at least 40 masked pixels. This is a slightly conservative approach (hereafter referred to as "conservative

masking") since the plume can be detected with a lower threshold if a supervised approach or other techniques are used. Thus, we proposed a second scenario (hereafter referred to "supervised masking") where the masking threshold and normalisation reference observation can be changed and optimised to improve this detection limit. In this second scenario the number of

235 pixels is reduced to 20 and the reference observation can be the average of the two closest acquitions separated by 5 days. The same reference observation is maintained for both Korpeje site (see Figure 2a) and the Permian basin (see Figure 2c). In Korpeje the average of two products resulted into almost and identical result (the retrieval noise improvement across the scene was lower than 5 ppm) whereas in the Permian one of the closest images was contaminated by clouds. For Hassi Messaoud (see Figure 2b), the reference observation is the average of the 27th of June 2021 and the 7th of July 2021 overpasses. The

240 use of these two acquisitions resulted in a scene noise improvement over 20 ppm. By reducing the number of pixels to 20, the masking process might include retrieval artifacts in addition to true methane plumes. By examining the retrieved methane enhancement maps ΔX_{CH4}, we can manually discard those areas that cannot be attributed to emissions. For example, this was the case for the roads in the Permian basin (see Figure 2c) or the O&G facilities close to border of the selected region in Hassi Messaoud (see Figure 2b).

245 The method to convert the methane enhancement maps ΔX_{CH4} into methane flux rates Q, is the integrated mass enhancement (IME) model which refers to the measure of the total excess mass of observed methane (Frankenberg et al., 2016; Varon et al., 2018). The ΔX_{CH4} values are converted from ppm units to kg (Thompson et al., 2016; Duren et al., 2019) by considering the 20 m S2 pixel resolution, the Avogadro's law where 1 mole of gas occupies 22.4 L, and M_{CH_4} (see eq. 1). Then, the IME is computed as the sum of the ΔX_{CH4} pixels defined by the plume mask. Finally the emission flux rate is calculated following

250 this expression:

$$Q = \frac{U_{\mathrm{eff}} \cdot \mathrm{IME} \cdot 3600}{L} \tag{7}$$

where L is a plume length scale in m (square root of the entire area covered by the plume pixels), and U_{eff} is the effective wind speed derived from WRF-LES simulations by Varon et al. (2021) for S2 observations of Korpeje and Hassi Messaoud oil fields in Algeria.

255 5   Real case results

The methodology described in section 4 has been applied to methane point sources of known O&G activity in the ~~same~~ areas ~~as~~ described in section 2. The location of each source is 38.493966°, 54.197664°in Korpeje, 31.805489°, 6.154881°in Hassi Messaoud, and 31.731557°, -102.042006°in the Permian basin. The overpass for the emission detection is the 23th of June 2022, the 31st of August 2021, and the 11th of July 2020 respectively.

260 Figure 6 presents the ΔXCH_4 map for each one of the three real emissions here described.

The maps show clear methane emissions at each one of the three sites combined with some outliers coming from different facilities and roads in the area. Figure 6f includes a granularity pattern which might be the result of inter-band and/or

(a) ΔXCH_4 (Korpeje) (b) ΔXCH_4 (Hassi Messaoud) (c) ΔXCH_4 (Permian)

(d) ΔXCH_4 mask (Korpeje) (e) ΔXCH_4 mask (Hassi Messaoud) (f) ΔXCH_4 mask (Permian)
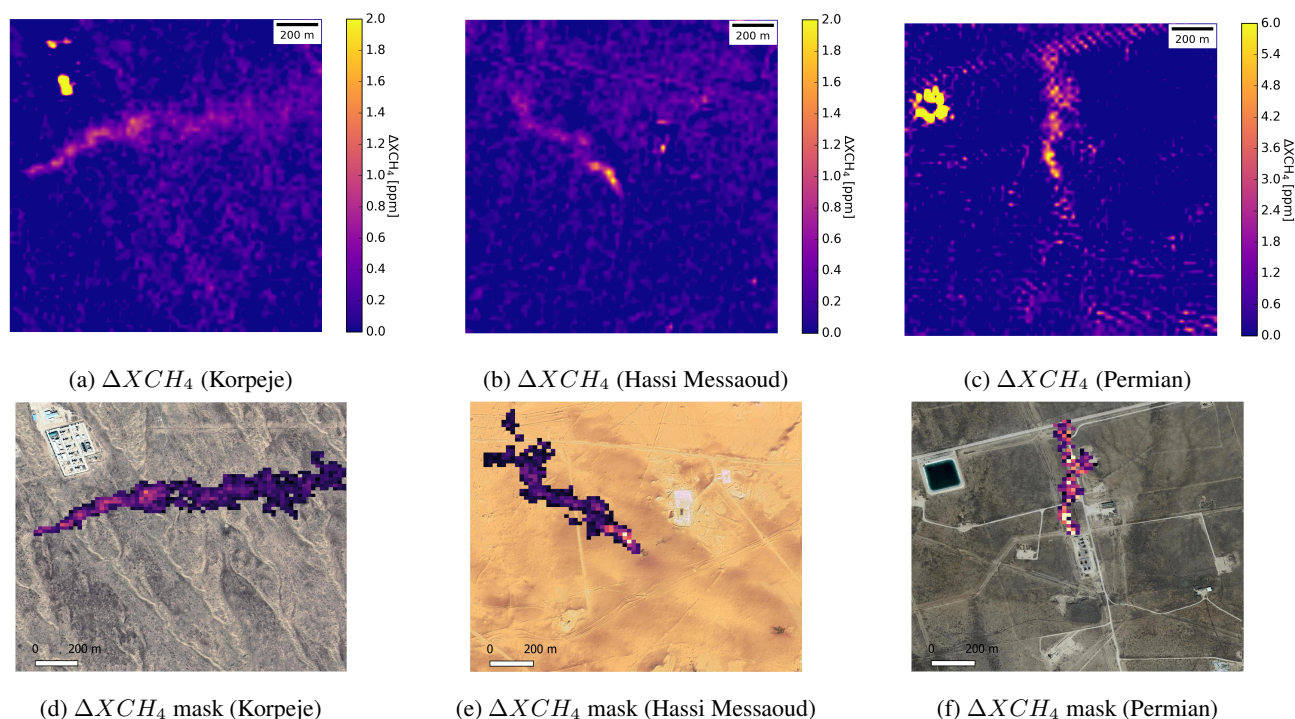
Figure 6. ΔXCH_4 maps for the known methane point sources in Korpeje (Fig. 6a), Hassi Messaoud (Fig. 6b), and Permian basin (Fig. 6c) O&G production regions. Figures 6d, 6e and 6f background map from © Google Earth and © Microsoft Bing

product registration. These enhancement maps have been converted into Q estimates as presented in section 4. For the estimation of U_{eff} the average 10 m wind U_{10} is required(Varon et al., 2021). We downloaded data from the NASA GEOS-FP meteorological reanalysis product at $0.25° \times 0.3125°$ resolution Molod et al. (2012) with values of $8.53174\,\mathrm{m\,s^{-1}}$ for Korpeje, $5.589153\,\mathrm{m\,s^{-1}}$ for Hassi Messaoud, and $4.46345\,\mathrm{m\,s^{-1}}$ for the Permian basin. The flux rates Q estimated in each case were $12\,091 \pm 5220\,\mathrm{kg\,h^{-1}}$, $5453 \pm 2200\,\mathrm{kg\,h^{-1}}$, and $19\,698 \pm 7585\,\mathrm{kg\,h^{-1}}$ respectively.

Both the enhancement maps and the flux rates here presented indicate that the algorithm is capable of detecting methane and provide a quantification estimate within the expected range. However, this qualitative assessment cannot provide an accurate assessment of the algorithm performance. The comparison against an accurate reference is recommended to understand the level of flux rate errors, the retrieved methane enhancement at a pixel level, and highlight potential algorithm improvements.

6 Simulation results

6.1 Retrieval results

The simulations for the three sites described in subsection 3 have been used to test the methane retrieval methodology described in Section 4.

Figure 7 illustrates an example of the map of ΔXCH_4 values, and the detection masks (based on either a conservative or supervised criteria as defined in section 4) against the validation reference obtained for a simulated dataset. This includes the methane plume 2 (see Figure 3b) with a flux rate $Q = 5000\,\mathrm{kg\,h^{-1}}$ for Korpeje and Hassi Messaoud sites. A value of $Q = 20\,000\,\mathrm{kg\,h^{-1}}$ has been used for the map of ΔXCH_4 values in the Permian basin due to its higher surface heterogeneity.

280   For the Korpeje site, the methane plume with a flux rate $Q = 10\,000\,\mathrm{kg\,h^{-1}}$ is clearly visible in the enhancement map in Figure 7a. The major structure visible in the image are the dunes that follow a North-west to South-East direction. Similarly, Figure 7b shows a visible plume in Hassi Messaoud over a similar dune structure but with slighly lower intensity. On the contrary, the enhancement map in the Permian basin (Figure 7c) presents a visible methane plume, but this is mixed with outliers despite its larger flux rate ($Q = 20\,000\,\mathrm{kg\,h^{-1}}$).

285   The masking process shows positive results for the Korpeje and Hassi Messaoud sites (see Figures 7d and 7g for Korpeje and Figures 7e and 7h for Hassi Messaoud respectively). Only the effect of the dune seems to have a small effect on the masking process with a slightly better detection for Hassi Messaoud with respect to Korpeje. The mask in Figure 7d is smaller compared to Figure 7g. By reducing the threshold of pixels required for detection changes between criteria (see Section 4) more pixels can be detected further along the plume. In the case of Hassi Messaoud, the average of two observations results in a smoother mask, and the plume shape is better defined.

290   When the process is repeated for the simulated plume over the Permian basin, the masking process is still successful; however, it is highly affected by outliers in the surrounding area (e.g. oil rigs and roads in Figure 2c). Figure 7f shows an example of this masking process where a significant area of the plume is detected. However, it does not include the pixels near the point source. By reducing the required number of pixels threshold to 20, those pixels near the source are detected (see Figure 7i). Nevertheless, by setting this criteria several parts of the road crossing the entire image from north to south were also detected (see Figure 2c). They were removed by masking the pixels close to that road.

295   The retrieval noise over the $300 \times 1500\ \mathrm{m}^2$ to the right of the plume (see section 4) was for a conservative masking criteria of 251.7 ppb, 151.5 ppb, and 1488.2 ppb, respectively for the Korpeje, Hassi Messaoud, and Permian basin sites. These are the noise values that define the threshold in the masking process (see section 4). The difference in the noise levels reported for each site closely correlates with the results presented in Figure 7. When a supervised masking criteria is applied, the noise levels do not vary in Korpeje and Permian basin sites since the same normalisation reference is used. However, the selection of the two closest S2 overpasses as a normalisation for Hassi Messaoud reduces the noise down to 139.1 ppb.

The retrieval noise in the Korpeje and Hassi Messaoud sites is lower than the 21%-49% of an 1875 ppb methane background reported by Varon et al. (2021) (assuming an 1800ppb background level, those noise levels correspond to 378 - 882 ppb).

305   The lower value obtained here is likely the result of differences in retrieval methodology, location, scene size, outliers, and temporal references. Nevertheless, the retrieval noise for the Permian basin scene is much higher because that scene is much more heterogeneous.

In addition to the retrieval maps and masked areas, Figure 8 presents the scatter plot that compares the retrieved results against the methane enhancement reference at a pixel level:

(a) ΔXCH_4 Q=10 t h^{-1} (Korpeje)

(b) ΔXCH_4 Q=10 t h^{-1} (Hassi Messaoud)

(c) ΔXCH_4 Q=20 t h^{-1} (Permian)

(d) Conservative mask
Q=10 t h^{-1} (Korpeje)

(e) Conservative mask
Q=10 t h^{-1} (Hassi Messaoud)

(f) Conservative mask
Q=20 t h^{-1} (Permian)

(g) Supervised mask
Q=10 t h^{-1} (Korpeje)

(h) Supervised mask
Q=10 t h^{-1} (Hassi Messaoud)

(i) Supervised mask
Q=20 t h^{-1} (Permian)
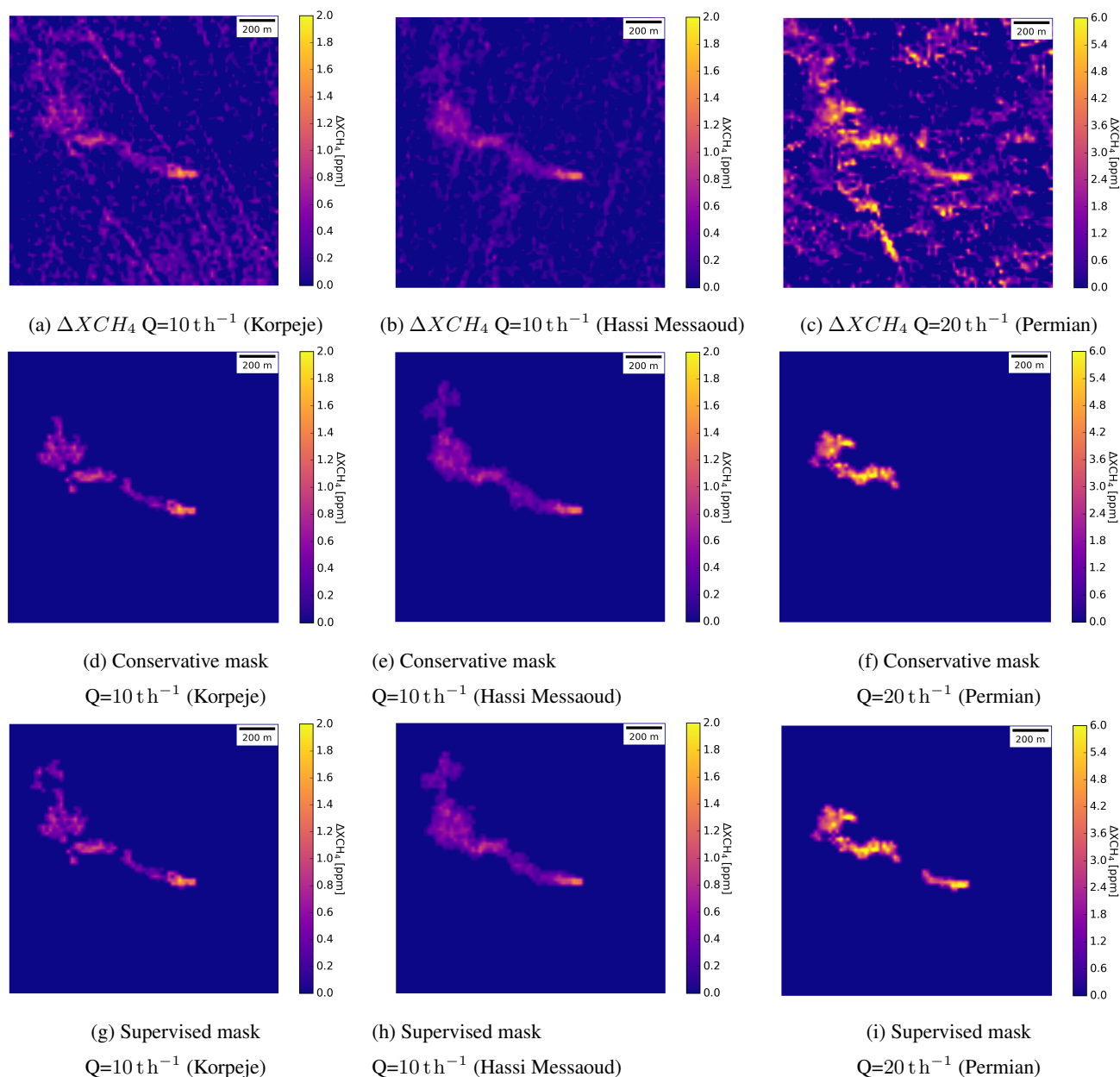
Figure 7. Results of the retrieval for the simulated datasets that include the map of ΔXCH_4 values (7a, 7b, and 7c), the mask obtained under a conservative masking criteria (7d, 7e and 7f) and a supervised masking criteria (7g, 7h, and 7i). The selected simulations include the methane plume 2 (see Figure 3b) with a flux rate $Q = 5000$ kg h^{-1} for Korpeje and Hassi Messaoud and $Q = 20\,000$ kg h^{-1} for the Permian basin. Note the different colorbar bounds for the Permian scene than for the Korpeje and Hassi Messaoud scenes.

(a) Scatter Q=$10 \, \mathrm{t \, h^{-1}}$ (Korpeje) (b) Scatter Q=$10 \, \mathrm{t \, h^{-1}}$ (Hassi Messaoud) (c) Scatter Q=$10 \, \mathrm{t \, h^{-1}}$ (Permian)
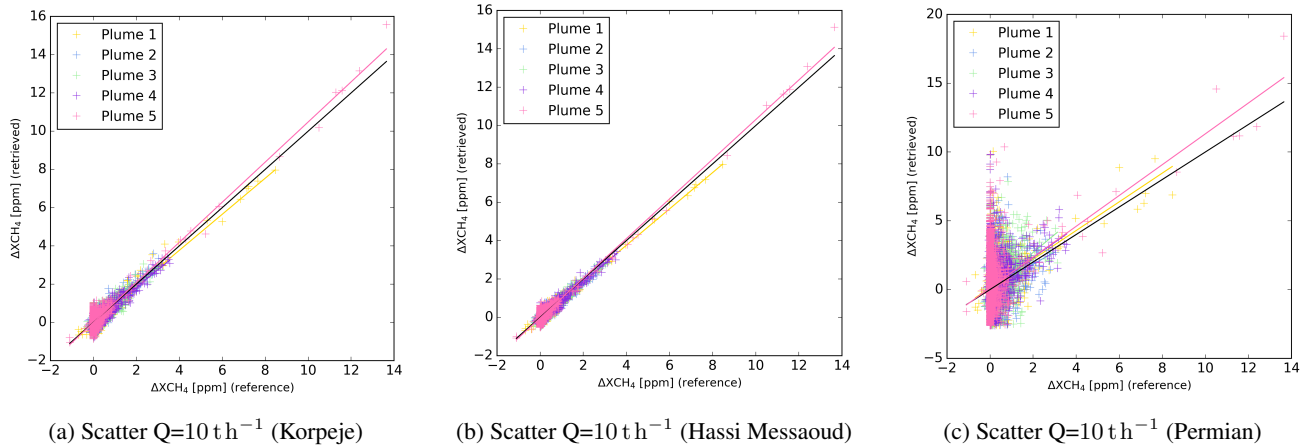
Figure 8. Scatter plots of the retrieval results against the validation reference at a pixel level for Korpeje (8a), Hassi Messaoud (8b) and Permian basin (8c) for a simulation that includes the methane plume 2 (see Figure 3b) scaled to a flux rate $Q =10\,000 \, \mathrm{kg \, h^{-1}}$

310   Similarly to the ideal case shown in Figure 5, the scatter plot of the retrieved ΔXCH_4 values against the reference enhancement map presents a good agreement for Figures 8a, 8b, and 8c. The data points are more sparse in this case as a consequence of the considerable noise levels of the S2 methane retrievals.

6.2 Flux rate validation

As previously explained in Section 3, the methane plumes can be scaled at different flux rates and cover a wide range of possible
315   emission levels. For both Korpeje and Hassi Messaoud scenes, the Gorroño et al. (2022) datasets have been scaled to different flux rates from $0 \, \mathrm{kg \, h^{-1}}$ to $20\,000 \, \mathrm{kg \, h^{-1}}$. A step of $100 \, \mathrm{kg \, h^{-1}}$ is set from $600 \, \mathrm{kg \, h^{-1}}$ to $3000 \, \mathrm{kg \, h^{-1}}$ so that the expected range of detection limits in the area can be accurately defined (Irakulis-Loitxate et al., 2022b). Larger steps of $500 \, \mathrm{kg \, h^{-1}}$ and $1000 \, \mathrm{kg \, h^{-1}}$ define the rest of the covered range. For the Permian scene, the covered range goes from $0 \, \mathrm{kg \, h^{-1}}$ to $50\,000 \, \mathrm{kg \, h^{-1}}$. Due to the larger potential range of detection limits, a step of $200 \, \mathrm{kg \, h^{-1}}$ is set from $5000 \, \mathrm{kg \, h^{-1}}$ to $15\,000 \, \mathrm{kg \, h^{-1}}$. The step
320   above $15\,000 \, \mathrm{kg \, h^{-1}}$ is relaxed to $1000 \, \mathrm{kg \, h^{-1}}$.

For each one of the flux rates at each site, we have obtained the Q value following the methodology in Section 4. Comparing the retrieved flux level against the simulation flux rate reference at each step defines a curve as a function of the flux rate that indicates the performance of the retrieval against our benchmark.

Figure 9 shows the result of this procedure for the Korpeje, Hassi Messaoud, and Permian basin sites and the methane plume
325   snapshots illustrated in Figure 3. Here, a conservative detection masking criteria has been applied as detailed in section 4.

The flux rate validation in Figures 9a and 9b evidence a better than 20% agreement in general. The bias is consistent across the different methane flux rates with a positive peak in some plumes near the detection limit. This effect is mainly visible in both plume 1 and plume 5 (see Figures 3a and 3e), representing the most concentrated plumes.

(a) Q validation (Korpeje) (b) Q validation (Hassi Messaoud) (c) Q validation (Permian)
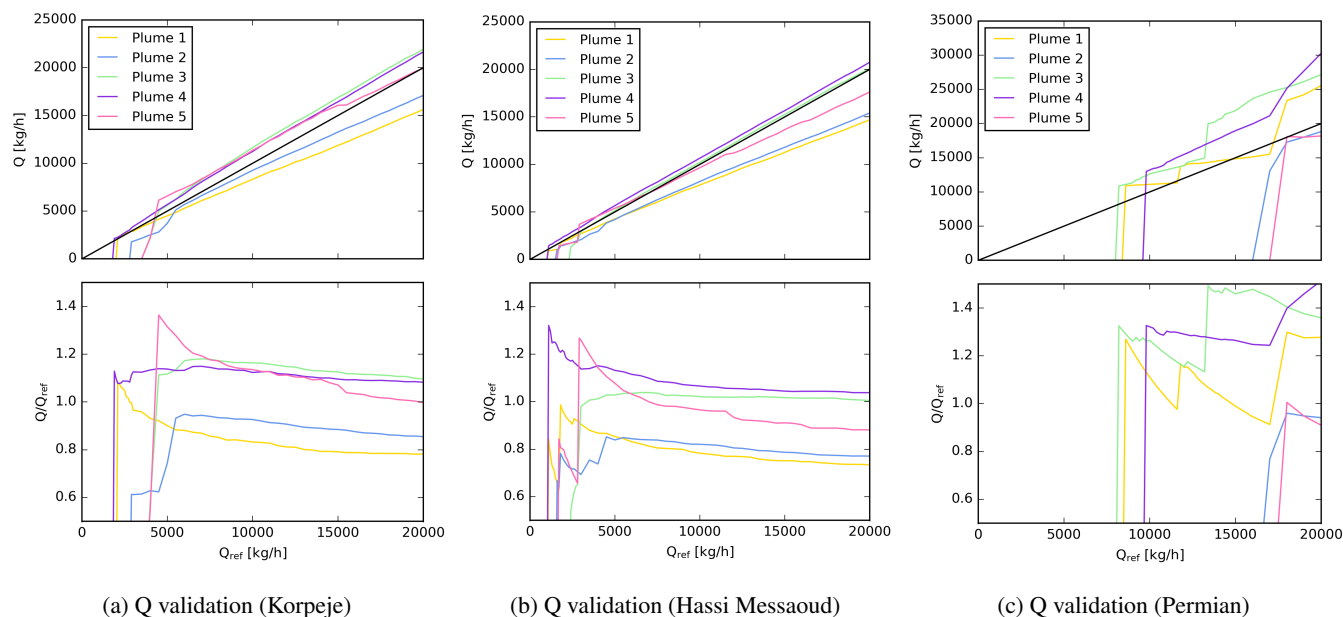
Figure 9. Q validation against reference values for a variety of plume shapes with a conservative detection masking criteria as detailed in section 4.

Figure 7f presents the results for the Permian basin. Due to its larger heterogeneity, the detection limit is considerably
330 higher. Moreover, the curves present larger disagreement and a non-stable behaviour. This can be explained by a large number
of outliers in the image (see blob-like shapes in Figure 7c) that interfere with the plume mask.

Figure 10 shows the analogous results to Figure 9 but with a supervised masking criteria as detailed in section 4.

The results in Korpeje (see Figure 10a) are highly similar to those in Figure 9a. The lower threshold in the number of pixels
also reduces the detection threshold leading to an increase in the peak levels in the relative error. Nonetheless, the opposite
335 occurs in Hassi Messaoud (Figure 10b) where the detection threshold is reduced together with the peaks close to it. In this case
the noise is lower (see subsection 6.1) and suggests that these peaks are generated by the background noise below the masked
area. Indeed, the comparison of the masked areas in the two cases showed that the lower noise was translated into a smoother
background retrieval pattern and eventually into a better plume masking.

For the Permian site, the results significantly decrease the detection limit but at the expense of larger disagreements and non-
340 stable trends (some of the trends are above the 50% error limit). The first sampling of the flux rate jumps from 0 to $5000\,\mathrm{kg\,h^{-1}}$.
At this level, the plume 3 can be already partially detected (the jump in the graph is the consequence of this sampling and not
produced by an outlier detection). These results suggest that the significant disagreements are the consequence of an important
effect of the background. For relatively small plume lengths, the underlying background is not only composed of noise but
of a structure with blob-like features mixing with the methane signal. Thus, quantifying methane plumes becomes highly
345 challenging for these heterogeneous sites, and specific management of these background features might be needed.

(a) Q validation (Korpeje)   (b) Q validation (Hassi Messaoud)   (c) Q validation (Permian)
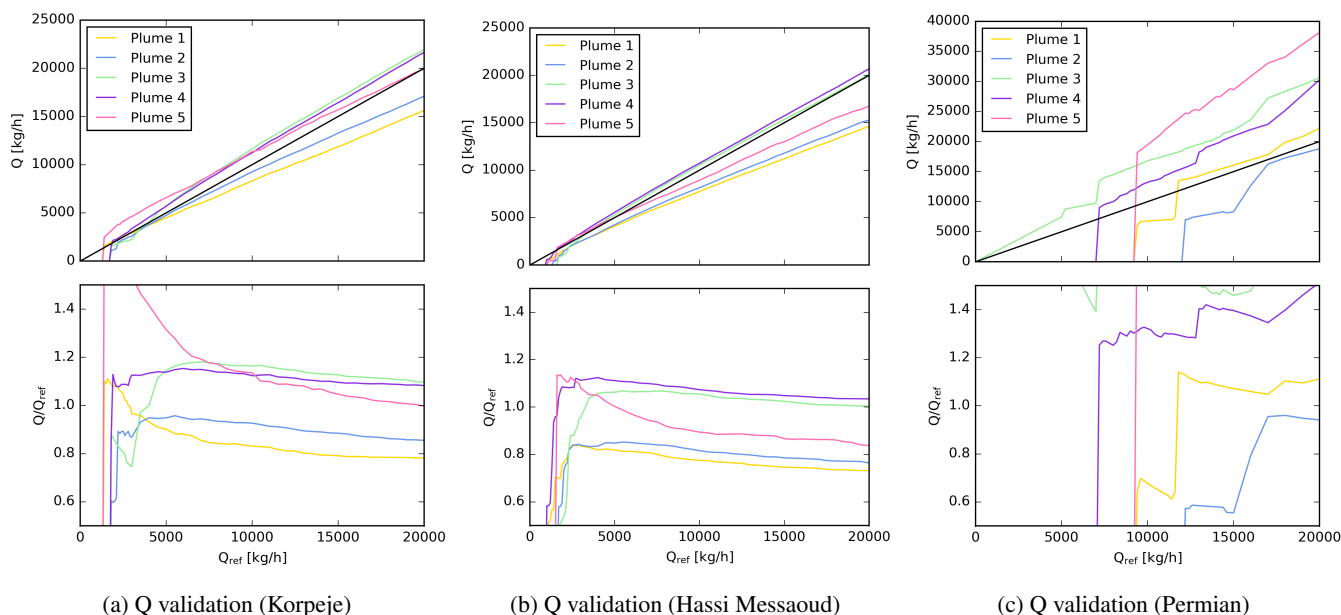
Figure 10. Q validation against reference values for a variety of plume shapes with a supervised detection masking criteria as detailed in section 4.

Interestingly, this validation is also helpful to provide an estimation of the detection limit for each site and plume. From Figures 9 and 10 we can identify the first flux values at which the plume can be captured.

Table 1 includes the detection limit for both scenarios and three sites considered in Section 2.

Table 1. Methane flux rate detection limit for the selected sites at Korpeje (Turkmenistan) (see Figure 2a), Hassi Messaoud (Algeria) (see Figure 2b) and Permian basin (USA) (see Figure 2c) under two different detection masking criteria (conservative and supervised) detailed in Section 4.

Plume	Korpeje (Turkmenistan)		Hassi Messaoud (Algeria)		Permian basin (USA)	
	Conservative $[\mathrm{kg\,h^{-1}}]$	Supervised $[\mathrm{kg\,h^{-1}}]$	Conservative $[\mathrm{kg\,h^{-1}}]$	Supervised $[\mathrm{kg\,h^{-1}}]$	Conservative $[\mathrm{kg\,h^{-1}}]$	Supervised $[\mathrm{kg\,h^{-1}}]$
1	2100	1400	1100	1100	8600	9400
2	2900	1800	1600	1400	17000	12200
3	4000	1800	2400	1700	8200	5000
4	1900	1800	1100	1000	9800	7200
5	4000	1400	1700	1100	18000	9400

The validation for Korpeje in Figure 9a shows a detection limit that ranges from $1900 \, \mathrm{kg\,h^{-1}}$ to $4000 \, \mathrm{kg\,h^{-1}}$ with a conser-
350   vative masking criteria. They are significantly reduced to levels from $1400 \, \mathrm{kg\,h^{-1}}$ to $1800 \, \mathrm{kg\,h^{-1}}$. Similarly the results in Hassi
Messaoud span from $1100 \, \mathrm{kg\,h^{-1}}$ to $2400 \, \mathrm{kg\,h^{-1}}$ with conservative masking criteria and from $1100 \, \mathrm{kg\,h^{-1}}$ to $1700 \, \mathrm{kg\,h^{-1}}$
under supervised masking criteria.

These results are coherent with the thresholds reported by Varon et al. (2021) ($2600 \, \mathrm{kg\,h^{-1}}$ for an approximately 500 ppm re-
trieval noise). In the specific case of Korpeje, the reported value of $1800 \, \mathrm{kg\,h^{-1}}$ in Irakulis-Loitxate et al. (2022b) is in the same
355   ranges as those presented in Table 1. For the Permian basin, the results suggest that the detection limit highly varies depending
on the plume shape (in combination with the underlying structure). The values are typically in the order of $10\,000 \, \mathrm{kg\,h^{-1}}$ down
to $5000 \, \mathrm{kg\,h^{-1}}$ under the best conditions. These results suggest that only very large methane emissions in heterogeneous areas
such as the Permian basin can be detected with the S2 mission with current retrieval methodologies.

6.3   Understanding the flux rate uncertainty sources

360   The results in Figures 9 and 10 show generally good agreement between retrieved and ground-truth emission rates. The dis-
agreement can arise from error generated during either the retrieval of the ΔXCH_4 map, its masking process or the IME to
Q conversion. The first of them might be generated by the limited accuracy of the LUT (see subsection 6.1) or the underlying
structure and noise below the masked area (see subsection 6.2).

The validation results in Korpeje and Hassi Messaoud (see Figures 10a and 10b) show that the error for each individual
365   plume lies approximately between $\pm 20\%$ of the reference flux rate. This is a correlated error among the different methane flux
rate levels but apparently uncorrelated among each individual plume. In this case, neither the ΔXCH_4 retrieval map nor the
underlying background features are expected to constitute a major source of error. Thus, these systematic errors are expected
to arise from the U_{eff} associated with the IME-to-Q conversion.

Here, the entire WRF-LES has been explored and a specific U_{eff} has been generated based on the dataset considered for the
370   simulations (see subsection 3). Each plume has been scaled to $10\,000 \, \mathrm{kg\,h^{-1}}$, and the relationship between the U_{eff} and U_{10} has
been established with a Huber regressor (Huber and Ronchetti, 2011). Figure 11 reports the specific U_{eff} regression curve and
compares it to the U_{eff} calibration reported in Varon et al. (2021).

The results indicate that the uncertainty on the U_{eff} calibration depends not only on that of U_{10} but also on the fitting knowl-
edge. This fitting knowledge can be further separated into a systematic component and a random one. The former depends on
375   the WRF-LES simulation and its specific parameterisation: emission altitude, surface roughness, turbulence... These simulation
datasets and the retrieval methodology will generate different relationships between the U_{eff} and U_{10} relationship. This error is
relatively small for this specific example (0-9% error for the 5 plumes considered) with higher disagreement the higher U_{10} is
for a linear fitting. However, it is a critical source of uncertainty for the assessment of emissions over O&G areas since these
errors will largely remain correlated under similar atmospheric conditions and, consequently, largely remain in the uncertainty
380   budget for the studied area.

The latter effect is the consequence of differences in the U_{eff} value between snapshots, and it results in an uncorrelated error
between methane plume snapshots. The root mean squared error between the U_{eff} calibration and the individual snapshots in
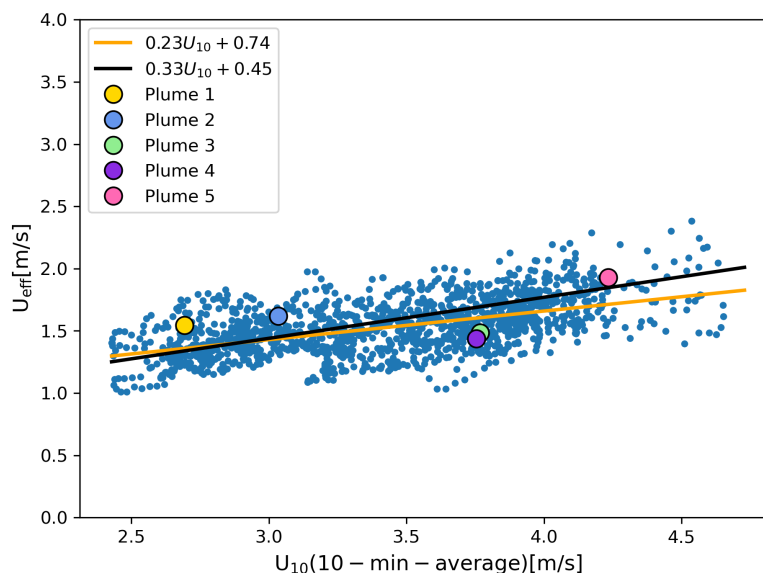
Figure 11. U_{eff} calibration for the specific simulation dataset (see subsection 3) at a flux rate of $10\,000\,\mathrm{kg\,h^{-1}}$. The yellow line fits a Huber regressor to the data. The black line shows the U_{eff} fit of Varon et al. (2021) based on a set of 25-m LES.

Figure 11 is 0.20. This represents a substantial error in relative terms that ranges from 10 to 20 %. This uncertainty source generally dominates for a single plume emission estimation especially at lower U_{10} values. However, its uncorrelated nature

385 with other methane plume emissions means its impact will decrease when an O&G area of several point-source emissions is studied. Large levels of turbulence will impact on the dispersion of the U_{eff} calibration values due to larger dispersion of the ΔXCH_4 snapshots, and the averaged U_{10} values. This error would ideally be zero if the dissipation of the plume occurred by uniform transport downwind of the source. The oscillation error between each plume in Figures 9 and 10 is inversely correlated to the specific plume deviations in Figure 11.

390 Thus, the major sources of uncertainty for the reported flux rates are the U_{10} uncertainty (~50% when using a meteorological reanalysis product (Varon et al., 2018)), a systematic uncertainty of U_{eff} in the range 0-9% and a random fitting uncertainty between 10 and 20%, a ΔXCH_4 retrieval accuracy (typically below 10%), and the impact of background features under the detection mask mainly applicable to the Permian basin case.

7 Conclusions and further work

395 This work presents a methodology and implementation to generate Sentinel-2 benchmark products with simulated methane plumes. The method keeps the original S2 observations and adds the simulated methane plume transmittance over the reflectance map. It also includes a compensation term in the calculated total atmospheric transmittance capable of correcting,

with a negligible residual, for the bias introduced by convolving a signal that corresponds to an already band-integrated quantity.

400   The proposed algorithm for the generation of a Sentinel-2 ΔXCH_4 map retrieves methane concentration enhancements from an estimated methane transmittance image. This methane transmittance is calculated as the ratio between the reflectance at a spectral channel affected by methane absorption and a methane-free reference band. This algorithm has been validated against the Sentinel-2 benchmark products in a real context where it was also possible to define a minimum emission detection threshold for specific sites and methane plumes. The scaling of these products and comparison against reference source magni-

405 tudes results in the validation of the retrieval at different flux rate levels and conditions. We find that for homogeneous surface types, a detection limit of $1000\,\mathrm{kg\,h^{-1}}$ to $2000\,\mathrm{kg\,h^{-1}}$ is feasible for S2. For these types of surface the validation resulted in an excellent agreement (within $\pm\,20\%$) largely depending on the plume shape. For areas with a large surface heterogeneity, the detection limit rises considerably over $5000\,\mathrm{kg\,h^{-1}}$. In these cases, the masking process becomes very sensitive to small variations due to the surface structure underlying the methane point sources. That results in a larger spread of disagreements

410 depending on the plume shape and masked area.

    The study was also capable of separating the different uncertainty sources associated to the flux rate estimates. The effective wind speed (U_{eff}) calibration has been tuned with the specific LES simulation and, in addition to the uncertainty associated to the 10-m wind speed (U_{10}), it describes both a systematic and random fitting error. The former effect largely depends on the WRF-LES simulation dataset and fitting method. The disagreement between fitting curves here was relatively small for this spe-

415 cific example (0-9% error for the 5 plumes considered) with higher disagreement the higher U_{10} is. The latter effect represents the departures from the U_{eff} calibration between snapshots, for which we found a root mean squared error of 0.20, representing a substantial error in relative terms that ranges from 10 to 20 %. However, when considering its effect over multiple plumes across an O&G area its impact will largely decrease due to the uncorrelated nature of the error between individual methane plume snapshots. In contrast, the systematic component might be highly correlated under similar atmospheric conditions and,

420 consequently, largely remain in the uncertainty budget for the studied O&G area. Finally, the study of heterogeneous scenes indicates the large impact of background features under the detection mask. In these cases, it is possible that this uncertainty contribution becomes relevant, or even dominates the uncertainty budget.

    The study here presented can be rapidly applied to any other mission, including Landsat, CHIME, or EnMAP. For each specific mission, a set of simplifications might reduce the computational costs, such as the assumption of a spectrally constant

425 surface reflectance ρ_{surface}.

    Further evolutions of the methodology might seek to refine the link between the modeled WRF-LES plumes and the reference ΔXCH_4 map. Among different improvements we can include: the consideration of the scene altitude to provide a better estimation of the column of dry air Ω_a (see Eq. 1), the selection of WRF-LES plumes with finer spatial resolution or the possibility to consider the point spread function of the instrument to generate a more realistic total transmittance map.

430   The accuracy of the retrieved ΔXCH_4 map largely depends on the selected observations for normalisation. In the case of Hassi Messaoud, we have seen how the noise was reduced from 151.5 ppb down to 139.1 ppb when the temporal normalization selects the two closest S2 overpasses. Temporal normalisation requires a complex understanding of different effects

that involve the consideration of different orbit viewing overpasses, satellite spectral, and geometric differences as well as the site temporal variation itself. The development of these simulations is expected to be helpful in testing novel approaches for

435 upcoming temporal normalisation methodologies such as in Ehret et al. (2022). Furthermore, the generation of realistic and accurate simulated datasets might be helpful to train upcoming machine learning methodologies such as the one presented in Jongaramrungruang et al. (2022).

The use of complementary methodologies and the continuous validation of the methane emissions is key for the development of robust estimates that contribute to the development of a global and systematic monitoring of methane emission from the

440 O&G sector. For example, the results here presented are consistent with those reported by Sherwin et al. (2022) based on an in-situ campaign in Arizona where controlled methane releases validated several missions and retrieval algorithms.

Data availability. These resulting maps of ΔX_{CH4} are included in a dataset by Gorroño et al. (2022). These dataset contain simulated S2 methane enhancement maps that can be used as a reference for calibration/validation purposes.

Author contributions. J.G designed the methods and implemented model code, D.J.V performed WRF-LES simulations and supported the

445 analysis of results, I.I. guided the selection of sites and scenarios, L.G. supported the study conceptualisation and provided general guidance. J.G. prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. J.G. is funded by the ESA Living Planet fellowship (ESA contract no. 4000130980/20/I-NS). This project has been funded by the ESA AO/1-10468/20/I-FvO-FUTURE EO-1 EO SCIENCE FOR SOCIETY PERMANENTLY OPEN CALL FOR PROPOS-

450 ALS. Project web: https://hiresch4.upv.es/

References

Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., and van den Bosch, J.: MODTRAN6: a major upgrade of the MODTRAN radiative transfer code, in: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX, edited by Velez-Reyes, M. and Kruse, F. A., vol. 9088, pp. 113 – 119, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2050433, 2014.

CCAC, U.: United nations environment programme and climate and clean air coalition, Global Methane Assessment: Benefits and Costs of Mitigating Methane Emissions. Nairobi: United Nations Environment Programme, 2021.

Chan, E., Worthy, D. E. J., Chan, D., Ishizawa, M., Moran, M. D., Delcloo, A., and Vogel, F.: Eight-Year Estimates of Methane Emissions from Oil and Gas Operations in Western Canada Are Nearly Twice Those Reported in Inventories, Environmental Science & Technology, 54, 14 899–14 909, https://doi.org/10.1021/acs.est.0c04117, pMID: 33169990, 2020.

Cusworth, D. H., Jacob, D. J., Varon, D. J., Chan Miller, C., Liu, X., Chance, K., Thorpe, A. K., Duren, R. M., Miller, C. E., Thompson, D. R., Frankenberg, C., Guanter, L., and Randles, C. A.: Potential of next-generation imaging spectrometers to detect and quantify methane point sources from space, Atmospheric Measurement Techniques, 12, 5655–5668, https://doi.org/10.5194/amt-12-5655-2019, 2019.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Olson-Duvall, W., Heckler, J., Chapman, J. W., Eastwood, M. L., Helmlinger, M. C., Green, R. O., Asner, G. P., Dennison, P. E., and Miller, C. E.: Intermittency of Large Methane Emitters in the Permian Basin, Environmental Science & Technology Letters, 8, 567–573, https://doi.org/10.1021/acs.estlett.1c00173, 2021.

Duren, R., Thorpe, A., Foster, K., Rafiq, T., Hopkins, F., Yadav, V., Bue, B., Thompson, D., Conley, S., Colombi, N., Frankenberg, C., McCubbin, I., Eastwood, M., Falk, M., Herner, J., Croes, B., Green, R., and Miller, C.: California's methane super-emitters, Nature, 575, 180–184, https://doi.org/10.1038/s41586-019-1720-3, 2019.

Ehret, T., De Truchis, A., Mazzolini, M., Morel, J.-M., d'Aspremont, A., Lauvaux, T., Duren, R., Cusworth, D., and Facciolo, G.: Global Tracking and Quantification of Oil and Gas Methane Emissions from Recurrent Sentinel-2 Imagery, https://doi.org/10.1021/acs.est.1c08575, pMID: 35797726, 2022.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L.: The libRadtran software package for radiative transfer calculations (version 2.0.1), Geoscientific Model Development, 9, 1647–1672, https://doi.org/10.5194/gmd-9-1647-2016, 2016.

ESA: Sentinel-2 MSI Level-1C data quality report, https://sentinel.esa.int/web/sentinel/data-product-quality-reports, [Online; accessed 23/06/2022], 2022.

Etminan, M., Myhre, G., Highwood, E. J., and Shine, K. P.: Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing, Geophysical Research Letters, 43, 12,614–12,623, https://doi.org/https://doi.org/10.1002/2016GL071930, 2016.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S., and Green, R. O.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, Proceedings of the National Academy of Sciences, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, 2016.

Gascon, F., Bouzinac, C., Thepaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Tremas, T., Cadou, E., De Bonis, R., Isola,

C., Martimort, P., and Fernandez, V.: Copernicus Sentinel-2A Calibration and Products Validation Status, Remote Sensing, 9, 1–81, https://doi.org/10.3390/rs9060584, 2017.

Gorroño, J., Varon, D., Irakulis-Loitxate, I., and Guanter, L.: Sentinel 2 L1C products with simulated methane plumes (S2CH4),
490   https://doi.org/10.7910/DVN/KRNPEH, 2022.

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J., Cogliati, S., and Colombo, R.: Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer, Remote Sensing of Environment, 265, 112 671, https://doi.org/https://doi.org/10.1016/j.rse.2021.112671, 2021.

Huber, P. J. and Ronchetti, E. M.: Robust statistics, Wiley, 2011.

495   Irakulis-Loitxate, I., Guanter, L., Liu, Y.-N., Varon, D. J., Maasakkers, J. D., Zhang, Y., Chulakadabba, A., Wofsy, S. C., Thorpe, A. K., Duren, R. M., Frankenberg, C., Lyon, D. R., Hmiel, B., Cusworth, D. H., Zhang, Y., Segl, K., Gorroño, J., Sánchez-García, E., Sulprizio, M. P., Cao, K., Zhu, H., Liang, J., Li, X., Aben, I., and Jacob, D. J.: Satellite-based survey of extreme methane emissions in the Permian basin, Science Advances, 7, eabf4507, https://doi.org/10.1126/sciadv.abf4507, 2021.

Irakulis-Loitxate, I., Gorroño, J., Zavala-Araiza, D., and Guanter, L.: Satellites Detect a Methane Ultra-emission Event from an Offshore
500   Platform in the Gulf of Mexico, Environmental Science & Technology Letters, 9, 520–525, https://doi.org/10.1021/acs.estlett.2c00225, 2022a.

Irakulis-Loitxate, I., Guanter, L., Maasakkers, J. D., Zavala-Araiza, D., and Aben, I.: Satellites Detect Abatable Super-Emissions in One of the World's Largest Methane Hotspot Regions, Environmental Science & Technology, 56, 2143–2152, https://doi.org/10.1021/acs.est.1c04873, pMID: 35102741, 2022b.

505   Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., and Frankenberg, C.: Satellite observations of atmospheric methane and their value for quantifying methane emissions, Atmospheric Chemistry and Physics, 16, 14 371–14 396, https://doi.org/10.5194/acp-16-14371-2016, 2016.

Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. K., Worden, J. R., and Duren, R. M.: Quantifying methane emissions from the global scale down
510   to point sources using satellite observations of atmospheric methane, Atmospheric Chemistry and Physics Discussions, 2022, 1–44, https://doi.org/10.5194/acp-2022-246, 2022.

Jongaramrungruang, S., Matheou, G., Thorpe, A. K., Zeng, Z.-C., and Frankenberg, C.: Remote sensing of methane plumes: instrument tradeoff analysis for detecting and quantifying local sources at global scale, Atmospheric Measurement Techniques, 14, 7999–8017, https://doi.org/10.5194/amt-14-7999-2021, 2021.

515   Jongaramrungruang, S., Thorpe, A. K., Matheou, G., and Frankenberg, C.: MethaNet – An AI-driven approach to quantifying methane point-source emission from high-resolution 2-D plume imagery, Remote Sensing of Environment, 269, 112 809, https://doi.org/https://doi.org/10.1016/j.rse.2021.112809, 2022.

Kokaly, R., Clark, R., Swayze, G., Livo, K., Hoefen, T., Pearson, N., Wise, R., Benzel, W., Lowers, H., and Driscoll, R.: USGS Spectral Library Version 7, https://doi.org/10.3133/ds1035, 2017.

520   Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., Shindell, D., and Ciais, P.: Global assessment of oil and gas methane ultra-emitters, Science, 375, 557–561, https://doi.org/10.1126/science.abj4351, 2022.

Lyon, D. R., Alvarez, R. A., Zavala-Araiza, D., Brandt, A. R., Jackson, R. B., and Hamburg, S. P.: Aerial Surveys of Elevated Hydrocarbon Emissions from Oil and Gas Production Sites, Environmental Science & Technology, 50, 4877–4886, https://doi.org/10.1021/acs.est.6b00705, pMID: 27045743, 2016.

525 Mayer, B. and Kylling, A.: Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use, Atmospheric Chemistry and Physics, 5, 1855–1877, https://doi.org/10.5194/acp-5-1855-2005, 2005.

Micijevic, E., Rengarajan, R., Haque, M. O., Lubke, M., Tuli, F. T. Z., Shaw, J. L., Hasan, N., Denevan, A., Franks, S., Choate, M. J., et al.: ECCOE Landsat quarterly Calibration and Validation report—Quarter 3, 2021, Tech. rep., US Geological Survey, https://doi.org/https://doi.org/10.3133/ofr20221025, 2022.

530 Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., and Eichmann, A.: The GEOS-5 atmospheric general circulation model: Mean climate and development from MERRA to Fortuna, Tech. rep., 2012.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

535 Qu, Z., Jacob, D. J., Shen, L., Lu, X., Zhang, Y., Scarpelli, T. R., Nesser, H., Sulprizio, M. P., Maasakkers, J. D., Bloom, A. A., Worden, J. R., Parker, R. J., and Delgado, A. L.: Global distribution of methane emissions: a comparative inverse analysis of observations from the TROPOMI and GOSAT satellite instruments, Atmospheric Chemistry and Physics, 21, 14 159–14 175, https://doi.org/10.5194/acp-21-14159-2021, 2021.

Sánchez-García, E., Gorroño, J., Irakulis-Loitxate, I., Varon, D. J., and Guanter, L.: Mapping methane plumes at very high spatial resolution
540 with the WorldView-3 satellite, Atmospheric Measurement Techniques Discussions, 2021, 1–26, https://doi.org/10.5194/amt-2021-238, 2021.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin,
545 M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton, J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J., Peng, C., Peng, S., Peters, G. P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf,
550 G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: The Global Methane Budget 2000–2017, Earth System Science Data, 12, 1561–1623, https://doi.org/10.5194/essd-12-1561-2020, 2020.

Sherwin, E., Rutherford, J., Chen, Y., Aminfard, S., Kort, E., Jackson, R., and Brandt, A.: Single-blind validation of space-based point-source methane emissions detection and quantification, https://doi.org/10.31223/X5DH09, 2022.

555 Thompson, D. R., Thorpe, A. K., Frankenberg, C., Green, R. O., Duren, R., Guanter, L., Hollstein, A., Middleton, E., Ong, L., and Ungar, S.: Space-based remote imaging spectroscopy of the Aliso Canyon CH4 superemitter, Geophysical Research Letters, 43, 6571–6578, https://doi.org/https://doi.org/10.1002/2016GL069079, 2016.

Thorpe, A. K., Frankenberg, C., and Roberts, D. A.: Retrieval techniques for airborne imaging of methane concentrations using high spatial and moderate spectral resolution: application to AVIRIS, Atmospheric Measurement Techniques, 7, 491–506, https://doi.org/10.5194/amt-
560 7-491-2014, 2014.

Tyner, D. R. and Johnson, M. R.: Where the Methane Is—Insights from Novel Airborne LiDAR Measurements Combined with Ground Survey Data, Environmental Science & Technology, 55, 9773–9783, https://doi.org/10.1021/acs.est.1c01572, pMID: 34251207, 2021.

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, Atmospheric Measurement Techniques, 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.

Varon, D. J., McKeever, J., Jervis, D., Maasakkers, J. D., Pandey, S., Houweling, S., Aben, I., Scarpelli, T., and Jacob, D. J.: Satellite Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production, Geophysical Research Letters, 46, 13 507–13 516, https://doi.org/https://doi.org/10.1029/2019GL083798, 2019.

Varon, D. J., Jervis, D., McKeever, J., Spence, I., Gains, D., and Jacob, D. J.: High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations, Atmospheric Measurement Techniques, 14, 2771–2785, https://doi.org/10.5194/amt-14-2771-2021, 2021.

Williams, J. P., Regehr, A., and Kang, M.: Methane Emissions from Abandoned Oil and Gas Wells in Canada and the United States, Environmental Science & Technology, 55, 563–570, https://doi.org/10.1021/acs.est.0c04265, pMID: 33322902, 2021.

Zhang, Y., Gautam, R., Pandey, S., Omara, M., Maasakkers, J. D., Sadavarte, P., Lyon, D., Nesser, H., Sulprizio, M. P., Varon, D. J., Zhang, R., Houweling, S., Zavala-Araiza, D., Alvarez, R. A., Lorente, A., Hamburg, S. P., Aben, I., and Jacob, D. J.: Quantifying methane emissions from the largest oil-producing basin in the United States from space, Science Advances, 6, eaaz5120, https://doi.org/10.1126/sciadv.aaz5120, 2020.