Detection of Methane Emissions Using Pattern Recognition

E. Ouerghi, T. Ehret, G. Facciolo, E. Meinhardt, J.-M. Morel, C. de Franchis,
T. Lauvaux

DETECTION OF METHANE EMISSIONS USING PATTERN RECOGNITION

E. Ouerghi, T. Ehret, G. Facciolo
E. Meinhardt, J.-M. Morel

Université Paris-Saclay, CNRS
ENS Paris-Saclay, Centre Borelli, France

C. de Franchis[†]
T. Lauvaux[‡]

[†]Kayrros SAS, France
[‡]LSCE, CEA, CNRS, UVSQ/IPSL

ABSTRACT

Reducing methane emissions is essential to tackle climate change. Here, we address the problem of detecting large methane leaks by using hyperspectral data from the satellite Sentinel-5P. By sampling Sentinel-5P spectral data at fine scale, we detect methane absorption features in the shortwave infrared wavelength range (SWIR). Our method involves two separate steps: i) background subtraction and ii) detection of local maxima in the negative logarithmic spectrum of each pixel. In the first step, we remove the impact of the albedo using albedo maps and the impact of the atmosphere by using a principal component analysis (PCA) over a time series of past observations. In the second step, we count for each pixel the number of local maxima that correspond to a subset of local maxima in the methane absorption spectrum. This counting method allows us to set up a statistical *a contrario* test that controls the false alarm rate of our detections.

Index Terms— methane, hyperspectral, pattern recognition, a contrario modeling, time series

1. INTRODUCTION

The detection of large methane (CH_4) leaks from oil and gas production is currently a major stake in order to reduce Greenhouse Gas (GHG) emissions. In a time lapse of 20 years, a CH_4 molecule has a global warming potential 80 times larger than carbon dioxide (CO_2) [1]. A large part of the (CH_4) emissions could be controlled or avoided, as they come from oil rigs and other oil and gas infrastructures.

In order to detect GHG fossil fuel emissions produced by human activities, several satellites have been placed in orbit around the Earth over the past ten years. Here, we focus on the data provided by Sentinel-5P, launched in 2017 by ESA. Sentinel-5P provides hyper-spectral images in wave bands for which CH_4 has a significant absorption coefficient.

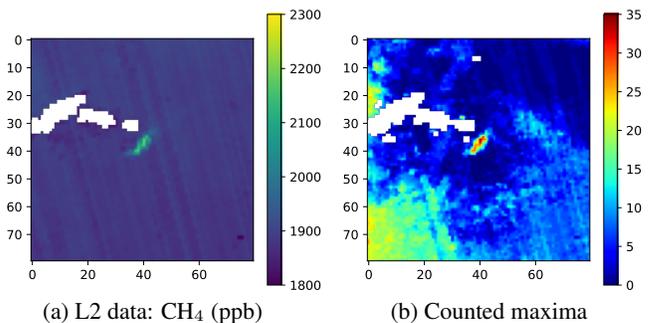(a) L2 data: CH_4 (ppb)   (b) Counted maxima

Fig. 1: Figure (a) shows the concentration of methane in parts per billion (ppb) given by the L2 methane product from Sentinel-5P. We see a methane plume in green. Pixels in white are discarded pixels (cloudy for example). Figure (b) gives the number of counted maxima for each pixel in the original image computed with the proposed method.

Data from Sentinel-5P is publicly available and is already being used by ESA to quantify CH_4 emissions and other greenhouse gases [2]. However, the results provided by ESA on CH_4 detection are rather incomplete. In most Sentinel-5P images, large regions are discarded by the retrieval algorithm. Indeed, this complex algorithm relies on external data sets which are not always available. The approach promoted by ESA involves precise physical modeling of the phenomenon of absorption, reflection and back scattering by atmosphere and ground of the radiation emitted by the sun to quantify X_{CH4}, the dry air column mixing ratio [3, 4]. CH_4 quantification is then performed by inverting the said model.

CH_4 detection can be performed with tools from the field of anomaly detection in hyper-spectral imagery. Usually, detection is performed on infrared wavelengths [5, 6]. Although the absorption features of CH_4 and water vapor (H_2O) sometimes overlap in the infrared spectrum, their separation can be addressed by a wise selection of wavelengths [6].

One classic anomaly detection strategy consists in performing a background subtraction; then locally modeling the remaining residual image as following a Gaussian model. This probabilistic model then enables Neyman-Pearson tests to be carried out on the pixels to detect anomalies [7, 8].

In [5], a CH$_4$-specific technique is developed where the background is removed by precise atmosphere modeling. Then, CH$_4$ is detected by a matched filter using the CH$_4$ absorption spectrum. Another matched filter is used for CH$_4$ detection in the Cluster-Tuned Matched Filter (CTMF), originally dedicated to sulfur dioxide [9], and then applied to CH$_4$ detection in [10]. Spatial clustering allows to obtain results over very large areas without the risk of confusing anomalies due to CH$_4$ with those due to other gases.

Another well-established technique in remote sensing is the band ratio technique. This method is quite usual for glacier monitoring [11] but is also used for hydrocarbons detection [12], and is highly-applicable to our goals. Band ratios are used to enhance the spectral signature of hydrocarbons.

Our objective is to introduce a flexible CH$_4$ emission detection method using the raw L1b data provided by Sentinel-5P. We do so by using the fine spectral sampling of Sentinel-5P data. We detect local maxima in the negative logarithmic spectra of pixels that correspond to maxima in CH$_4$ absorption spectrum. Observing many such maxima in the same pixel should be the consequence of CH$_4$ emissions. Results of this method will then be compared with those obtained by the ESA in the current L2 product.

2. MATERIALS

We use hyper-spectral images from Sentinel-5P. This satellite provides a dense spectrum (nearly 4,000 samples) for each pixel and covers the entire Earth once a day. Spectral samples are organized in eight wavelength bands. Here, we use the SWIR bands 7 and 8 which cover the 2,300-2,389 nm range where the main absorption feature of CH$_4$ is located. These images are part of the level 1 (L1) product.

We also use the level 2 (L2) fully-processed data including cloud maps, albedo and X_{CH4} column mixing ratios. Cloud maps are necessary because their presence precludes any CH$_4$ detection. X_{CH4} images will be used to identify plumes of CH$_4$ and use them as ground truth.

Lastly, we use a detailed CH$_4$ absorption spectrum taken from the HITRAN spectral database [13]. CH$_4$ spectrum varies depending on pressure and temperature. However, here we are only interested in CH$_4$ spectrum within the bottom layers of the atmosphere (below 1,500 meters above ground), because we want to detect CH$_4$ shortly after being emitted, before it rises and dilutes.

Spectrum variations are small in the near-surface atmospheric layers. Moreover, here we use the CH$_4$ absorption spectrum only for detecting local maxima, and as we can see from Fig. 2, its profile practically does not change for the near-surface conditions. Here, we selected the CH$_4$ spectrum at $15°$C and 1 atm to represent near-surface atmospheric conditions.
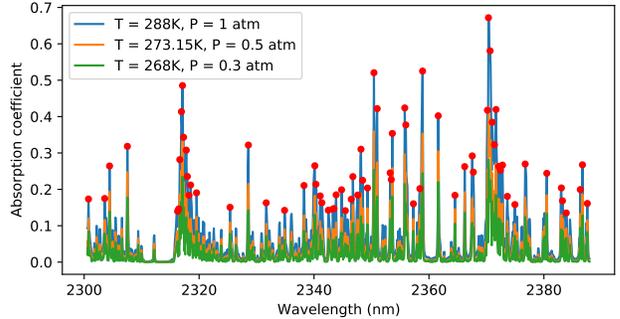

Fig. 2: Comparison of methane absorption spectra under several pressure/temperature conditions. Red dots highlight the 70 maxima we use for detecting the plumes.

3. MODELING

We shall use a simplified atmospheric absorption model to explain the value of each pixel in each channel of an image. Let us consider a pixel $\mathbf{P}$ in a Sentinel-5P image. Such a pixel is a vector in $\mathbb{R}^d$ where d is the number of channels in the image (for our case 960 channels from bands 7 and 8). Each pixel component $\mathbf{P}_i$ corresponds to a wavelength λ_i.

Our model takes into account: the effect of sun irradiance $F_I(\lambda_i)$, the albedo A, the absorption coefficients of the dry atmosphere $K_{atm}(\lambda_i)$, water vapor $K_{H_2O}(\lambda_i)$ and methane $K_{CH_4}(\lambda_i)$. We denote by e_{gas} the thickness of gas crossed by the radiation before reaching the sensor. Making implicit the dependence on λ_i we can write the absorption model for the whole vector $\mathbf{P}$ as

$$\mathbf{P} = A \exp(F_I - K_{atm}e_{atm} - K_{H_2O}e_{H_2O} - K_{CH_4}e_{CH_4}). \quad (1)$$

Similarly to other works [14], we assume here that the albedo is roughly constant over the part of the infrared spectrum we use, as it is extremely regular near 2000nm [15]. We take into account the absorption by the dry atmosphere, as a single gas whose absorption spectrum is well known. This spectrum includes absorption from methane that is always present in the atmosphere. The term $K_{CH_4}e_{CH_4}$ represents the excess of emitted methane over the one already present in the dry atmosphere.

From now on, we shall work with $-\log(\mathbf{P})$ instead of $\mathbf{P}$. This allows us to have a linear model where excesses of methane comes out positively. We denote by P_0 the new pixel value.

4. METHODS

Our methane plume detection method consists of two steps: first a background subtraction to remove the contribution of albedo and atmosphere, and then a counting step where we count for each pixel the number of local maxima that correspond to a shortlist of local maxima in methane absorption spectrum.

4.1. Background subtraction

Our observations are dominated by the absorption of the atmosphere and the albedo. In order to detect potential local excess of CH_4, we start by performing a background subtraction. This background subtraction has two advantages. First, it removes the contribution of albedo and atmosphere from the spectrum of the current pixel. Second, it sets the mean CH_4 concentration to zero. Indeed, there is a nearly-constant concentration of CH_4 in the atmosphere and CH_4 plumes rarely exceed 3% of this concentration. So, we must make sure that this mean CH_4 concentration is completely removed during background subtraction. To do so background subtraction applies three steps.

1. Albedo removal. First, albedo values from the L2 data are used to remove the albedo component from each pixel. Given a pixel P_0 from a pre-processed hyperspectral image I and the albedo A corresponding to this pixel, we compute the albedo-corrected pixel $P_1 = P_0 + \log(A)$. The albedo is assumed to be identical for each channel, as variations of albedo are minor in the infrared spectrum [14, 15]. After this first subtraction, P_1 contains only contributions from irradiance, atmosphere, water vapor and CH_4.

2. Atmosphere removal. For removing the contribution of the atmosphere we assume the irradiance and the dry atmosphere's absorption spectrum to be roughly constant over a short period of time (in practice this analysis is performed over two weeks or less). So, we can estimate those two components using a time series. For each pixel P_1, we gather observations $X_1, ... X_n$ of the same area at earlier dates and without clouds. The background is then modeled as the principal component of $X_1, ..., X_n$, which we denote F. To remove the background of P_1 we then remove its projection on the subspace directed by F, i.e. $P_2 = P_1 - \langle P_1, F \rangle F$.

3. Methane equalization. The last part of the background subtraction is an equalization of the level of CH_4 in the current image. This subtraction works both spatially and spectrally. After the second part of the background subtraction, only CH_4 and water vapor should be left. However, this is not enough to detect CH_4 anomalies. Indeed, there are about 1900 ppb (particles per billion) of CH_4 in the atmosphere, but we want to detect variations in the order of 40 to 80 ppb. When the background is subtracted, a variable fraction of those 1900 ppb are removed, depending on the atmospheric CH_4 concentrations of earlier observations. This difference in background CH_4 can prevent a detection using local maxima. To address this issue we first compute a spatial average of CH_4 concentration M by projecting each pixel on the CH_4 direction

$$M = \frac{1}{|I|} \sum_{P \in I} \left\langle P_2, \frac{K_{CH_4}}{\|K_{CH_4}\|} \right\rangle. \qquad (2)$$

Then, we remove this mean M from each pixel

$$P_3 = P_2 - M \times \frac{K_{CH_4}}{\|K_{CH_4}\|}. \qquad (3)$$

After this last operation each pixel should only display a mix of water vapor and excess CH_4. Thus, CH_4 detection should be possible when the concentration of water vapor is not too high.

4.2. Local maxima detection

Once the background has been removed, the spectrum of a pixel should mostly be composed of excess CH_4, water vapor and sensor noise. CH_4 and water vapor have different spectra in the wavelengths we are interested in. Thus, we expect that the wavelengths where CH_4 absorption is maximal, should also appear as maxima in the negative logarithmic spectrum of a pixel with excess CH_4. Otherwise, there should be only noise on those wavelengths. Therefore, to detect CH_4 plumes, we count for each pixel the number of local maxima coinciding to local maxima in the CH_4 absorption spectrum.

For this, we cannot use all of CH_4 local maxima. Some of them correspond to small absorption coefficients, which could be easily confused with noise. Thus, we can only use the highest maxima in the methane spectrum. We selected the 70 highest maxima in the CH_4 absorption spectrum between 2300nm and 2380nm. We can see them in Figure 2. In addition, we set up two adapted thresholds for the detection of each of these maxima.

The first threshold $\tau_1(P)$ is the median of the spectrum of the pixel P; this prevents low maxima from being detected. Since we use only the 70 highest maxima of the CH_4 spectrum, we should also have high maxima in P. So this threshold will not hinder the detection of CH_4-related maxima.

We then compute a threshold adapted to each of the 70 chosen maxima. With the first threshold some maxima can appear in almost every pixel in the image, just because a specific wavelength usually shows high values. For the wavelength λ we set the threshold $\tau_2(\lambda)$ as the 70% quantile of all the values of the image at that wavelength; i.e for an image I the 70% quantile of $\{P_\lambda \mid P \in I\}$. To summarize, for a maximum at wavelength λ in a pixel P, our detection threshold is $\max(\tau_1(P), \tau_2(\lambda))$.

We need to set up a final decision threshold to tell apart CH_4 plume (excess CH_4) pixels from background pixels based on the counted maxima. In order to do this, we set up an *a contrario* model. We take the *a contrario* assumption that the image has no excess CH_4, and compute the probability of false detection under this assumption [16]. To do so, we index the 70 highest CH_4 spectrum maxima by i, going from 1 to 70, and we denote by p_i the empirical probability that a spectrum maximum occurs at i in a "normal" image. If CH_4 anomalies occur in the image under study, they are generally concentrated on very few pixels. Hence we can estimate p_i from the image itself, and this will lead to a tiny overestimation of this probability if some pixels have excess CH_4. The random variable X_i, which is equal to 1 if the i-th maximum appears and 0 otherwise, follows a Bernoulli distribution with
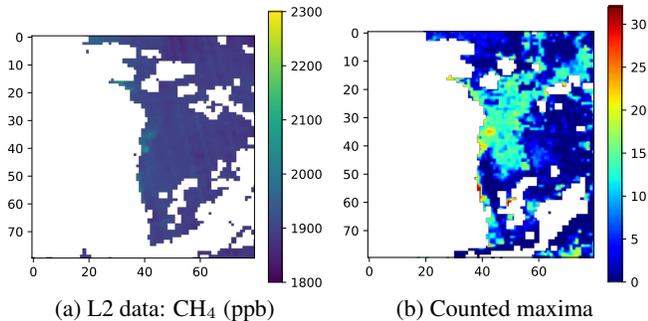
(a) L2 data: CH$_4$ (ppb) (b) Counted maxima

Fig. 3: In the center of (a) we see a methane plume. We detect this plume in our result (b). But we also find two other plumes below the first one that cannot be seen in the L2 product. Note that less pixels are discarded in our image than in the L2 product.

parameter p_i. To complete the *a contrario* model, we assume that $X_1, ..., X_n$ are independent (in the absence of CH$_4$). We denote by $S(P) = X_1 + ... + X_n$ the number of counted maxima on a given pixel P. We then compute a detection threshold τ which guarantees a given false alarm rate p_{fa}

$$p_{fa}(\tau) = \mathbb{P}(S(P) > \tau | P \text{ without excess CH}_4). \quad (4)$$

The value of p_{fa} is set to 10^{-6} in our experiments, which amounts statistically to less than 0.01 false alarm per image.

5. RESULTS AND DISCUSSION

As shown in Fig. 1, the results obtained with our method are consistent with the excess in X$_{CH4}$ detected in L2 products. We detect the major part of the plume seen in L2 data. We can notice that we have a significant gap in the number of maxima counted between the pixels with excess CH$_4$ and the others. This allows us to have few false detections. We tested our algorithm with plumes of several sizes in different locations. Of 12 plumes detected with the algorithm of ESA, 10 were detected as having an NFA lower than 10^{-6}, thus unlikely to appear just by chance.

Conversely, as illustrated in Fig. 3, we can detect potential plumes that are not found in the L2 product. As our method is less dependent on external data, it discards fewer pixels than the L2 product. In short, our method for CH$_4$ detects anomalous pixels backed by a statistical model and can compute a detection threshold for any proposed plume. However, the method does not work in presence of clouds or over water where albedo is very small. Moreover, the model used here is a very simplified model ignoring the Sentinel-5P instrument noise and the influence of other greenhouse gases like NO$_2$. Further atmospheric modeling could improve the background subtraction and the final detection.

6. REFERENCES

[1] "Anthropogenic and Natural Radiative Forcing," in *Climate Change 2013 - The Physical Science Basis*, Intergovernmental Panel on Climate Change, Ed., pp. 659–740. Cambridge University Press, Cambridge, 2013.

[2] S. Pandey et al., "Satellite observations reveal extreme methane leakage from a natural gas well blowout," *PNAS*, vol. 116, no. 52, pp. 26376–26381, Dec. 2019.

[3] A. Butz, A. Galli, O. Hasekamp, J. Landgraf, P.J.J. Tol, and E.A.A. Aben, "Tropomi aboard sentinel-5 precursor: Prospective performance of ch4 retrievals for aerosol and cirrus loaded atmospheres," *Remote Sens Environ*, vol. 120, no. SI, pp. 267–276, 2012.

[4] J. Landgraf, J. Brugh, et al., "Carbon monoxide total column retrievals from tropomi shortwave infrared measurements," *Atmos Meas Tech*, vol. 9, pp. 4955–4975, 10 2016.

[5] R. Scafutto and C. De Souza Filho, "Detection of methane plumes using airborne midwave infrared hyperspectral data," *Remote Sensing*, vol. 10, no. 8, 2018.

[6] Crevoisier C, D. Nobileau, et al., "Tropospheric methane in the tropics–first year from iasi hyperspectral infrared observations," *Atmos Chem Phys Discussions*, vol. 9, 09 2009.

[7] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *Signal Processing Magazine, IEEE*, vol. 19, pp. 29 – 43, 02 2002.

[8] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aero El Sys Mag*, vol. 25, pp. 5 – 28, 08 2010.

[9] C.C. Funk, T. James, D. Roberts, and C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, pp. 1410 – 1420, 08 2001.

[10] A. Thorpe, D. Roberts, E. Bradley, C. Funk, P. Dennison, and I. Leifer, "High resolution mapping of methane emissions from marine and terrestrial sources using a cluster-tuned matched filter technique and imaging spectrometry," *Remote Sens Environ*, vol. 134, pp. 305 – 318, 2013.

[11] M. Citterio, F. Paul, et al., "Remote sensing of glacier change in west greenland: Accounting for the occurrence of surge-type glaciers," *Ann Glaciol*, vol. 50, pp. 70–80, 02 2010.

[12] S. Garain, D. Mitra, and P. Das, "Detection of hydrocarbon microseepage-induced anomalies by spectral enhancements of landsat 7 etm+ images in part of assam–arakan fold belt, india," *J Pet Explor Prod Technol*, vol. 9, 07 2019.

[13] I. E. Gordon, L. S. Rothman, C. Hill, R. V. Kochanov, Y. Tan, et al., "The HITRAN2016 molecular spectroscopic database," *J Quant Spectrosc Ra*, vol. 203, pp. 3 – 69, Dec. 2017.

[14] J. Coakley, "Reflectance and albedo," in *Encyclopedia of the Atmosphere*, pp. 1914–1923. Academic Press, 2002.

[15] B. Montpetit, A. Royer, et al., "New shortwave infrared albedo measurements for snow specific surface area retrieval," *Journal of Glaciology*, vol. 58, pp. 941–952, 09 2012.

[16] A. Desolneux, L. Moisan, and J. Morel, "Maximal meaningful events and applications to image analysis," *The Annals of Statistics*, vol. 31, no. 6, pp. 1822–1851, 2003.