Guided Transformer Network for Detecting Methane Emissions in Sentinel-2 Satellite Imagery

Satish Kumar
University of California Santa Barbara
satishkumar@ucsb.edu

William Kingwill
Orbio Earth
william@orbio.earth

Rozanne Mouton
Orbio Earth
zani@orbio.earth

Wojciech Adamczyk
ETH Zurich
wojtekadamczyk3@gmail.com

Robert Huppertz
Orbio Earth
robert@orbio.earth

Evan Sherwin
Stanford University
evands@stanford.edu

Abstract

Methane (CH_4) is the chief contributor to global climate change and its mitigation is targeted by the EU, US and jurisdictions worldwide [2]. Recent studies have shown that imagery from the multi-spectral instrument on Sentinel-2 satellites is capable of detecting and estimating large methane emissions. However, most of the current methods rely on temporal relations between a ratio of shortwave-infrared spectra and assume relatively constant ground conditions, and availability of ground information on when there was no methane emission on site. To address such limitations we propose a guided query-based transformer neural network architecture, that will detect and quantify methane emissions without dependence on temporal information. The guided query aspect of our architecture is driven by a Sentinel Enhanced Matched Filter ($SEMF$) approach, also discussed in this paper. Our network uses all 12 spectral channels of Sentinel-2 imagery to estimate ground terrain and detect methane emissions. No dependence on temporal data makes it more robust to changing ground and terrain conditions and more computationally efficient as it reduces the need to process historical time-series imagery to compute a single date emissions analysis.

1 Introduction

Methane (CH_4) is estimated to contribute 20% of global warming induced by greenhouse gasses [14] with a Global Warming Potential (GWP) 86 times higher than carbon dioxide (CO_2) in a 20 year period [18]. The longstanding greenhouse gas has a mean atmospheric residence of 7.9 years [17] and its presence in the atmosphere has been increasing since the industrial revolution [1]. The annual growth rate of CH_4 plateaued between 1999 and 2006, but started increasing again in 2007 [20]. Causes of this variable growth rate are highly debated, attributing the sporadic behavior to the natural gas industry, emission from wetlands, and changes in the methane lifetime [19, 26, 27, 25, 33]. The increases in atmospheric CH_4 have prompted governments to enact regulations and action plans such as the 'U.S. Methane Emissions Reduction Action Plan' in 2021 and the 'Global Methane Pledge Energy Pathway' in 2022 to curb CH_4 emission [2, 21]. Accurately identifying and tracking the contribution of various sources to the methane budget will be paramount to enforce these regulations.

Given the strong potential of satellite-based instruments to deliver high-frequency data on global scales and even remote and hard-to-access regions, recent research has depicted the potentials of

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

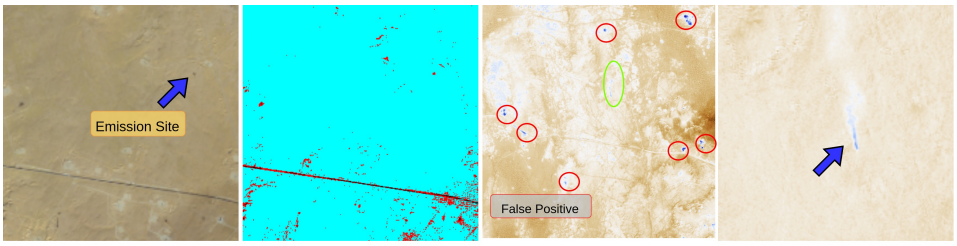

Figure 1: Qualitative visualization of *SEMF* intermediate steps and final estimation.

deploying methane emissions analysis on public, global-mapping, multi-spectral instruments like the ESA Sentinel-2 mission [34, 3, 8, 22, 3]. Most previous Sentinel-2-based methane analysis approaches use similar approaches to the methane column retrieval method in [34], building large parts of signal exploitation on an analysis of temporal deviation between times of excessive methane concentrations in the atmosphere and times without, merged with ratios between methane-sensitive and less-methane-sensitive bands. While this method, and variations of it [5] have revolutionized capabilities of detecting methane emissions with public satellite data, the strong dependency on time-series analysis of spectral reflectance data expose the approach to risky assumptions on (a) knowing when emissions did not exist and (b) temporal albedo stability of the background - that the albedo of a certain area stays constant over time. In return, these assumptions lead to high amounts of false positives, especially in areas with heterogeneous, temporally deviating land cover [34, 36].

To overcome shortcomings of time-series based methane analytics methods, we propose a deeper exploitation of signals from other non-methane-sensitive spectral bands of Sentinel-2 multi-spectral data. Given both, the recent successes of enhance match filters for methane signal processing [31, 6, 16, 32] and progresses in using Machine Learning models for methane emissions analysis [16], we hereby propose a 2-step methodology to generate (1) a novel Sentinel Enhanced Match Filter (SEMF) algorithm and (2) an integration of SEMF into a Transformer-based Convolutional Neural Network architecture [4, 10] as shown in Figure 2. Using the full spectral response captured by Sentinel-2 instruments, we expect the SEMF to support Signal to Noise separation by classifying confusing and hard-to-detect land cover types, artifacts and temporal deviations, such as water bodies, dark green vegetation, calcite, and white painted roofs which are never considered in current band/channel ratio method. Using these classes, SEMF computes statistical properties for each class separately for the whitening of background pixels.

2 Approach

The proposed approach is a transformer [35] based neural network architecture with a $SEMF$ guidance. The input to the network is B1-B12 bands from Sentinel-2 Level 1C data [23]. The output is a segmentation mask that is used with a radiative transfer model for methane emissions analysis. The overall architecture (Figure 2) presents 2 feature extraction blocks (ResNet [9]) as shown in Figure 2, that will extract useful features from both, an input RGB image and a stack of B1-B12 bands from Sentinel-2 Level 1C data [23]. While the singled-out RGB image will provide information about land cover (e.g. Urban areas), the full B1-B12 stack provides additional land cover feature extraction (e.g. water bodies) while also capturing information about methane presence. Extracted features will be projected in a common subspace via a MLP [29] and passed on to the transformer encoder network along with positional information of each pixel in the image as shown in Figure 2. The output attention map [35] from the transformer encoder along with project features are passed onto the transformer decoder. The decoder network will use our $SEMF$ to generate a query of the potential methane emission sites. The $SEMF$ is discussed in more detail Section 2.2.

2.1 Dataset

We will be training and testing the proposed network on a mix of large-eddy based methane plume simulation data (synthetic data) [24] and single-blind release, human-labelled data [28]. The synthetic data includes images that contain simulated methane emissions on different types of background terrain. Each image is be a $10km \times 10km$ tile with 12 channels at different spatial resolutions per pixel. Next to the simulation data, we propose model validation to happen on manually selected

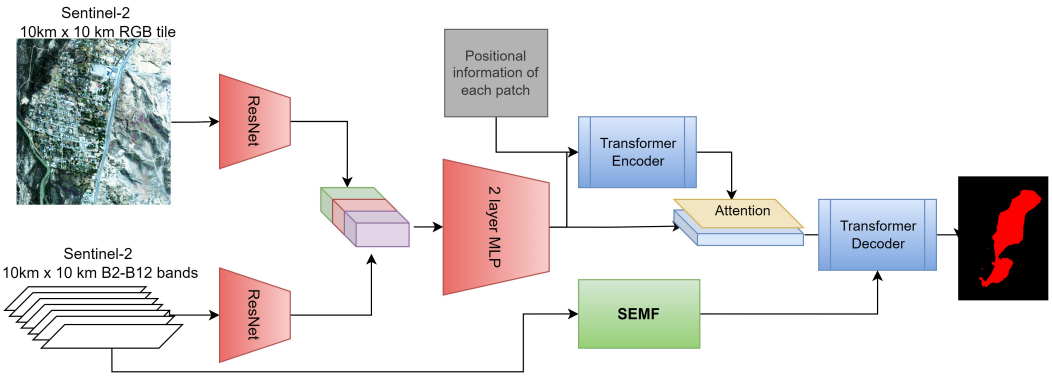

Figure 2: Our proposed transformer based Neural Network architecture, $SEMF$ is one of the most critical block of it presented in current proposal

emissions data from controlled ground releases [28]. The labels used in training corresponding to each multi-spectral image will be a binary segmentation mask (methane, no-methane) with the same spatial dimension as the input image. Along with that, we have a concentration mask, representing the concentration of methane per pixel in the mol/m^2.

2.2 Sentinel Enhanced Matched Filter (SEMF)

$SEMF$ is an essential component of our proposed architecture. The $SEMF$ model is used to generate refined queries for our transformer decoder network, leading to quicker convergence of our model. $SEMF$ is inspired by a deterministic, linear match-filtering approach of finding CH_4 [16, 31]. The linear approach is taking a n-dimensional (number of spectral channels) feature α, and apply as a dot product to each pixel (n-dimension) in the multi-spectral image to generate a scalar output per pixel. The α vector is "matched filter" [16, 31], making the process of finding the best-fitting α critical for signature exploitation in the ground terrain distribution at hand.

In ideal instances when there is no background (i.e. all white ground terrain) and just CH_4 gas present, the α is just the scaled version of the CH_4 signature ($\mathbf{t}$). However in real-world scenarios with spatially varying ground terrain this is not the case. For example, water has strong absorption of solar radiations, therefore the methane on such backgrounds has very weak visibility [11]. On the other hand, bare soil, rocks, etc have lower absorption, and the methane present in such background has strong visibility. An understanding of ground terrain and underlying albedo properties (especially in the methane sensitive spectral ranges) is critical to improve Signal to Noise ratios in our Sentinel-2 data. To account for spatial albedo differences in real-world scenes, we propose to deploy a land cover classification as shown in Figure 1 and use that land cover information to build our $SEMF$ (See Appendix Sec 4.2). The final SEMF used in our architecture is:

$$\hat{\alpha_k}(\mathbf{r}_i) = \frac{(\mathbf{r}_i - \mu_k)^T \mathbf{Cov}_k^{-1} \epsilon \mathbf{t}}{\sqrt{\epsilon \mathbf{t}^T \mathbf{Cov}_k^{-1} \epsilon \mathbf{t}}} \; \forall \; i \in k, \tag{1}$$

$$SEMF(\mathbf{r}_i) = \frac{(\mathbf{r}_i - \mu)^T \mathbf{Cov}^{-1} \mathbf{t}}{\sqrt{\epsilon \mathbf{t}^T \mathbf{Cov}^{-1} \mathbf{t}}} \tag{2}$$

where $\hat{\alpha_k}(\mathbf{r}_i)$ is the estimated methane column enhancement, $\mathbf{r}_i$ is the captured radiance at i^{th} pixel in the multispectral image, μ_k & $\mathbf{Cov}^{-1}$ are the mean and the inverse of covariance matrix for k^{th} class and ϵ represent the chemical properties of CH_4. $SEMF$ is represented by a green block in Figure 2. An example of $\hat{\alpha_k}(\mathbf{r}_i)$ estimations is shown in column-4 of Figure 1. Our approach is simple and effective, it can be implemented with basic python pseudo-code as shown in appendix algorithm 1. Details about $SEMF$ can be found in the Appendix at the end of the paper.

3 Future work

While we have developed and tested the $SEMF$ of CH_4 estimations only on a few samples where some good ground data exists, there is much work to be done to (a) implement it in the proposed

transformer-based neural network architecture, (b) test the whole transformer model with globally simulation-based and release-based emissions data (c) evaluate performance of the model by comparing it with the time-series-based approaches in recent literature.

References

[1] Philippe Ciais, Christopher Sabine, Govindasamy Bala, Laurent Bopp, Victor Brovkin, Josep Canadell, Abha Chhabra, Ruth DeFries, James Galloway, Martin Heimann, et al. Carbon and other biogeochemical cycles. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 465–570. Cambridge University Press, 2014.

[2] Mary Crowell. President biden announces new methane emissions reduction strategy, 2022.

[3] Daniel H Cusworth, Daniel J Jacob, Daniel J Varon, Christopher Chan Miller, Xiong Liu, Kelly Chance, Andrew K Thorpe, Riley M Duren, Charles E Miller, David R Thompson, et al. Potential of next-generation imaging spectrometers to detect and quantify methane point sources from space. *Atmospheric Measurement Techniques*, 12(10):5655–5668, 2019.

[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[5] Thibaud Ehret, Aurélien De Truchis, Matthieu Mazzolini, Jean-Michel Morel, A. d'Aspremont, Thomas Lauvaux, Riley M. Duren, Daniel H. Cusworth, and Gabriele Facciolo. Global tracking and quantification of oil and gas methane emissions from recurrent sentinel-2 imagery. *Environmental science & technology*, 2022.

[6] Christian Frankenberg, Andrew K Thorpe, David R Thompson, Glynn Hulley, Eric Adam Kort, Nick Vance, Jakob Borchardt, Thomas Krings, Konstantin Gerilowski, Colm Sweeney, et al. Airborne methane remote measurements reveal heavy-tail flux distribution in four corners region. *Proceedings of the national academy of sciences*, 113(35):9734–9739, 2016.

[7] Christopher C Funk, James Theiler, Dar A Roberts, and Christoph C Borel. Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery. *IEEE transactions on geoscience and remote sensing*, 39(7):1410–1420, 2001.

[8] Takashi Hamazaki, Yutaka Kaneko, Akihiko Kuze, and Kayoko Kondo. Fourier transform spectrometer for greenhouse gases observing satellite (gosat). In *Enabling sensor and platform technologies for spaceborne remote sensing*, volume 5659, pages 73–80. SPIE, 2005.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] ASM Iftekhar, Satish Kumar, R Austin McEver, Suya You, and BS Manjunath. Gtnet: Guided transformer network for detecting human-object interactions. *arXiv preprint arXiv:2108.00596*, 2021.

[11] Itziar Irakulis-Loitxate, Luis Guanter, Yin-Nian Liu, Daniel J Varon, Joannes D Maasakkers, Yuzhong Zhang, Apisada Chulakadabba, Steven C Wofsy, Andrew K Thorpe, Riley M Duren, et al. Satellite-based survey of extreme methane emissions in the permian basin. *Science Advances*, 7(27):eabf4507, 2021.

[12] Daniel J Jacob, Alexander J Turner, Joannes D Maasakkers, Jianxiong Sheng, Kang Sun, Xiong Liu, Kelly Chance, Ilse Aben, Jason McKeever, and Christian Frankenberg. Satellite observations of atmospheric methane and their value for quantifying methane emissions. *Atmospheric Chemistry and Physics*, 16(22):14371–14396, 2016.

[13] California Institute of Technology Jet Propulsion Laboratory. Airborne visible infrared imaging spectrometer - next generation (aviris-ng) overview, 2009.

[14] Stefanie Kirschke, Philippe Bousquet, Philippe Ciais, Marielle Saunois, Josep G Canadell, Edward J Dlugokencky, Peter Bergamaschi, Daniel Bergmann, Donald R Blake, Lori Bruhwiler, et al. Three decades of global methane sources and sinks. *Nature geoscience*, 6(10):813–823, 2013.

[15] Roman V Kochanov, IE Gordon, LS Rothman, P Wcisło, C Hill, and JS Wilzewski. Hitran application programming interface (hapi): A comprehensive approach to working with spectroscopic data. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 177:15–30, 2016.

[16] Satish Kumar, Carlos Torres, Oytun Ulutan, Alana Ayasse, Dar Roberts, and BS Manjunath. Deep remote sensing methods for methane detection in overhead hyperspectral imagery. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1776–1785, 2020.

[17] JOS Lelieveld, Paul J Crutzen, and Frank J Dentener. Changing concentration, lifetime and climate forcing of atmospheric methane. *Tellus B*, 50(2):128–150, 1998.

[18] Gunnar Myhre, Drew Shindell, and Julia Pongratz. Anthropogenic and natural radiative forcing. 2014.

[19] EG Nisbet, EJ Dlugokencky, MR Manning, D Lowry, RE Fisher, JL France, SE Michel, JB Miller, JWC White, B Vaughn, et al. Rising atmospheric methane: 2007–2014 growth and isotopic shift. *Global Biogeochemical Cycles*, 30(9):1356–1370, 2016.

[20] Euan G Nisbet, Edward J Dlugokencky, and Philippe Bousquet. Methane on the rise—again. *Science*, 343(6170):493–495, 2014.

[21] US Department of State. U.s.-eu joint press release on the global methane pledge energy pathway, 2022.

[22] Sudhanshu Pandey, Ritesh Gautam, Sander Houweling, Hugo Denier Van Der Gon, Pankaj Sadavarte, Tobias Borsdorff, Otto Hasekamp, Jochen Landgraf, Paul Tol, Tim Van Kempen, et al. Satellite observations reveal extreme methane leakage from a natural gas well blowout. *Proceedings of the National Academy of Sciences*, 116(52):26376–26381, 2019.

[23] Darius Phiri, Matamyo Simwanda, Serajis Salekin, Vincent R Nyirenda, Yuji Murayama, and Manjula Ranagalage. Sentinel-2 data for land cover/use mapping: a review. *Remote Sensing*, 12(14):2291, 2020.

[24] Anja Ražnjević, Chiel van Heerwaarden, Bart van Stratum, Arjan Hensen, Ilona Velzeboer, Pim van den Bulk, and Maarten Krol. Interpretation of field observations of point-source methane plume using observation-driven large-eddy simulations. *Atmospheric Chemistry and Physics*, 22(10):6489–6505, 2022.

[25] Matthew Rigby, Stephen A Montzka, Ronald G Prinn, James WC White, Dickon Young, Simon O'doherty, Mark F Lunt, Anita L Ganesan, Alistair J Manning, Peter G Simmonds, et al. Role of atmospheric oxidation in recent methane growth. *Proceedings of the National Academy of Sciences*, 114(21):5373–5377, 2017.

[26] Hinrich Schaefer, Sara E Mikaloff Fletcher, Cordelia Veidt, Keith R Lassey, Gordon W Brailsford, Tony M Bromley, Edward J Dlugokencky, Sylvia E Michel, John B Miller, Ingeborg Levin, et al. A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by 13ch4. *Science*, 352(6281):80–84, 2016.

[27] Stefan Schwietzke, Owen A Sherwood, Lori MP Bruhwiler, John B Miller, Giuseppe Etiope, Edward J Dlugokencky, Sylvia Englund Michel, Victoria A Arling, Bruce H Vaughn, James WC White, et al. Upward revision of global fossil fuel methane emissions based on isotope database. *Nature*, 538(7623):88–91, 2016.

[28] Evan David Sherwin, Jeffrey S Rutherford, Yuanlei Chen, Sam Aminfard, Eric A Kort, Robert B Jackson, and Adam R Brandt. Single-blind validation of space-based point-source methane emissions detection and quantification. 2022.

[29] Hind Taud and JF Mas. Multilayer perceptron (mlp). In *Geomatic approaches for modeling land change scenarios*, pages 451–455. Springer, 2018.

[30] James Theiler, Bernard R Foy, and Andrew M Fraser. Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*, volume 6565, pages 26–37. SPIE, 2007.

[31] DR Thompson, I Leifer, H Bovensmann, M Eastwood, M Fladeland, C Frankenberg, K Gerilowski, RO Green, S Kratwurst, T Krings, et al. Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane. *Atmospheric Measurement Techniques*, 8(10):4383–4397, 2015.

[32] DR Thompson, AK Thorpe, C Frankenberg, RO Green, R Duren, A Hollstein, L Guanter, E Middleton, L Ong, and S Ungar. Orbital measurement of the aliso canyon ch4 super-emitter. *Geophys Res Lett*, 43:6571–6578, 2016.

[33] Alexander J Turner, Christian Frankenberg, Paul O Wennberg, and Daniel J Jacob. Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl. *Proceedings of the National Academy of Sciences*, 114(21):5367–5372, 2017.

[34] Daniel J Varon, Dylan Jervis, Jason McKeever, Ian Spence, David Gains, and Daniel J Jacob. High-frequency monitoring of anomalous methane point sources with multispectral sentinel-2 satellite observations. *Atmospheric Measurement Techniques*, 14(4):2771–2785, 2021.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[36] Zhan Zhang, Evan D Sherwin, Daniel J Varon, and Adam R Brandt. Detecting and quantifying methane emissions from oil and gas production: algorithm development with ground-truth calibration based on sentinel-2 satellite imagery. *EGUsphere*, pages 1–23, 2022.

4 Appendix

4.1 Background & Motivation:

Although important for regulation enforcement,successfully detecting and delineating CH_4 plumes on a regular basis poses a great challenge to researchers and stakeholders given its amorphous characteristics. To tackle the overall lack of reliable and actionable methane emissions information on the vast scales global natural and anthropogenic methane-emitters (e.g. oil & gas sites, landfills, wetlands), there are increasing efforts by the research and industry community to detect and quantify methane plumes with airborne [13] and satellite-based instruments [34, 12, 3, 22] . Given the strong potential of satellite-based instruments to deliver high-frequency data on global scales and even remote and hard-to-access regions, recent research has depicted the potentials of deploying methane emissions analysis on public, global-mapping, multi-spectral instruments like the ESA Sentinel-2 mission [34, 3, 8, 22, 3]. With two polar-orbiting, sun-synchronous satellites, the Multispectral Instrument (MSI) onboard the Sentinel-2 satellites measures the reflected radiance from Earth in multiple bands covering various areas of the electromagnetic spectrum [23]. Among these bands, band 11 ($\sim 1500nm - 1660nm$) and band 12 ($\sim 2090nm - 2290nm$) are able to capture methane's SWIR absorption features at a spatial resolution of $20m^2$, leading to a large breadth of work and studies on using Sentinel-2 data to detect and quantify methane emissions [34, 36, 12, 5].

4.2 Enhanced Matched Filter:

The passive multispectral image is $H \times W \times n \in \mathbb{R}$ where H and W are height and width of the image respectively and n ($n \sim 10$) spectral channels. In this data, we are looking for a very weak signature (CH_4) of interest hidden in background (variable ground terrain). The most common linear approach for finding CH_4 is taking a n-dimensional (number of spectral channels) feature α, and apply as a dot product to each pixel (n-dimension) in the hyperspectral image to generate a scalar

output per pixel. This operation is supposed to remove the ground terrain and amplifies CH_4 signature. The α vector is "matched filter" [16, 31]. Therefore computing right α is very critical for finding the signature of interest. It is dependent on desired signature and on the distribution of the ground terrain. To model α, let $\mathbf{r}_i \in \mathbb{R}^n$ be a i^{th} radiance vector from the hyperspectral image representing the ground terrain pixel and sensor noise, and $\mathbf{t}$ be the CH_4 signature. This is modeled as the additive perturbation. The spectrum is represented by $\xi(\mathbf{x}_B)$, when the gas is present. The linear matched filter is modeled as additive perturbation:

$$\xi(\mathbf{r}_i) = \mathbf{r}_i + \epsilon\mathbf{t}, \tag{3}$$

Here $\xi(\mathbf{r}_i)$ is the spectrum when CH_4 is present and ϵ represents the chemical properties of the gas. The CH_4 signature $\mathbf{t}$ represents the change in radiance units of the background caused by adding a unit mixing ratio length of CH_4 absorption [7, 16]. In the ideal scenario where only CH_4 gas is present in signal (i.e. all white background), the matched filter output is $\alpha^T\epsilon\mathbf{t}$. In case there is no gas and just ground terrain and sensor noise, the matched filter output is $\alpha^T\mathbf{r}_i$. The variance (Var) of $\alpha^T\mathbf{r}_i$ for the latter is represented as :

$$Var(\alpha^T\mathbf{r}_i) = \langle(\alpha^T\mathbf{r}_i - \alpha^T\boldsymbol{\mu})^2\rangle = \alpha^T\mathbf{Cov}\alpha, \tag{4}$$

Here **Cov** and μ are covariance and mean respectively. Inspired from [16, 7] we define the Methane-to-Ground terrain Ratio (MGR) as

$$\text{MGR} = \frac{|\alpha^T\epsilon\mathbf{t}|^2}{\alpha^T\mathbf{Cov}\alpha}, \tag{5}$$

We can see that the magnitude of α does not affect MGR. According to [30, 7, 16], the MGR can be maximized subject to constraints(zero mean and $\alpha^T\mathbf{K}\alpha$ constraint to 1). The matched filter α is then represented by:

$$\alpha = \frac{\mathbf{Cov}^{-1}\epsilon\mathbf{t}}{\sqrt{\epsilon\mathbf{t}^T\mathbf{Cov}^{-1}\epsilon\mathbf{t}}}, \tag{6}$$

where $\mathbf{t}$ is the CH_4 signature compute from HITRAN database [15], ϵ represents the chemical properties of the gas, and **Cov** is covariance of the ground terrain. In ideal instances when there is no background (i.e. all white background) and just CH_4 gas present. The matched filter is directly proportional to $\mathbf{t}$. In simple terms, it is just the target signature ($\mathbf{t}$) itself scaled so that the filtered output has variance of one. The methane enhancement per column can be computed as follows:

$$\hat{\alpha}(\mathbf{r}_i) = \frac{(\mathbf{r}_i - \mu)^T\mathbf{Cov}^{-1}\epsilon\mathbf{t}}{\sqrt{\epsilon\mathbf{t}^T\mathbf{Cov}^{-1}\epsilon\mathbf{t}}}, \tag{7}$$

where $\mathbf{t}$ is the CH_4 signature compute from HITRAN database [15], ϵ represents the chemical properties of the gas, and **Cov** is covariance of the ground terrain. $\hat{\alpha}(\mathbf{r}_i)$ is the column enhancement of methane per pixel. The covariance matrix (**Cov**) used is not known as $prior$ and is estimated from data. It is computed as outer product of the mean subtracted radiance over all the pixels. In other words, The standard matched filter from equation 7 computes the covariance (**Cov**) of ground terrain with an underlying assumption that in all elements have similar absorption pattern. The standard matched filter from equation 7 computes the covariance (**Cov**) of ground terrain with an underlying assumption that in all elements have similar absorption pattern. But in realistic scenarios, the type of terrain changes frequently, there is water bodies, bare soil, vegetation, dense vegetation, building structures in cities, roads etc in a single image. For example, water have a strong absorption of solar radiations, therefore the methane on such backgrounds have a very weak visibility. On the other hand, bare soil, rocks, etc have lower absorption, the methane present on such background have strong visibility. A simple and single approximation of the covariance (**Cov**) of ground distribution can not provide the right and effective estimate of methane enhancement. To tackle this limitation, we do land cover segmentation and classification. **Cov per class:** In practice we have 20 classes, each with a segmentation mask. We merge two or more adjacent classes into one if the number of pixels in that class is less 20000. The Number of pixels in each class is kept higher to ensure that while computing the covariance (**Cov**) matrix, the methane signal does not have any or have negligible effect. It is okay to merge adjacent classes into one because they have almost similar radiance/reflectance, for example, light vegetation and normal vegetation have similar reflectance, etc. The covariance $\mathbf{Cov}_k$ of k^{th} class is computed as:

$$\mathbf{Cov}_k = \frac{1}{N}\sum_{i=1}^{i=j}(\mathbf{r}_i - \mu_k)(\mathbf{r}_i - \mu_k)^T \;\forall\; j \in k, \tag{8}$$

where N is the number of pixels (> 10000) in k^{th} class and μ_k is the mean of k^{th} class. For each class we compute the mean μ_k, covariance matrix $\mathbf{Cov}_k$ and matched filter α_k. While iterating through each pixel of hyperspectral image, we check to which class k the pixel $\mathbf{r}_i$ belongs to and use those pre-computed values. The final Sentinel Enhanced Matched Filter is shown in algorithm 1

$$\hat{\alpha_k}(\mathbf{r}_i) = \frac{(\mathbf{r}_i - \mu_k)^T \mathbf{Cov}_k^{-1} \epsilon \mathbf{t}}{\sqrt{\epsilon \mathbf{t}^T \mathbf{Cov}_k^{-1} \epsilon \mathbf{t}}} \; \forall \; i \in k, \tag{9}$$

where $\mathbf{Cov}^{-1}$ is the inverse of covariance matrix. Our approach is simple and effective, it can be implemented with basic python pseudo code as shown in algorithm 1.

Data: *Sentinel-2 12 channel dataset*
Result: *CH_4 concentration map*
initialization;
for $datum$ *in* $DATASET$ **do**
    1. create memory map $datum$;
    2. seg_mask = compute segmentation mask;
    for *mask in seg_mask* **do**
        data.**append**($datum$[mask])
        if (len(data) $<$ 100000): **continue**
        Cov, μ = compute_stats(data);
    end
    3. **for** r_i *in data* **do**
        k = seg_mask[i];
        $\hat{\alpha_k}(\mathbf{r}_i) = \frac{(\mathbf{r}_i - \mu_k)^T \mathbf{Cov}_k^{-1} \epsilon \mathbf{t}}{\sqrt{\epsilon \mathbf{t}^T \mathbf{Cov}_k^{-1} \epsilon \mathbf{t}}}$
    end
    $\hat{\alpha_k}(\mathbf{r}_i) \; \forall \; classes \; and \; i \in datum$
end

Algorithm 1: $SEMF$