U-Plume: Automated algorithm for plume detection and source quantification by satellite point-source imagers

Jack H. Bruno[1,2], Dylan Jervis[2], Daniel J. Varon[1], Daniel J. Jacob[1]

[1]Earth And Planetary Sciences, Harvard University, Cambridge, MA, 02138, USA
[2]GHGSat Inc, Montreal, Quebec, H2W1Y5, Canada

Correspondence to: Jack H. Bruno (jackbruno@g.harvard.edu)

Abstract. Current methods for detecting atmospheric plumes and inferring point source rates from high-resolution satellite imagery are labor intensive and not scalable to the growing satellite dataset available for methane point sources. Here we present a two-step algorithm called U-Plume for automated detection and quantification of point sources from satellite imagery. The first step delivers plume detection and delineation (masking) with a machine learning U-Net architecture for image segmentation. The second step quantifies point source rate from the masked plume using wind speed information and either a convolution neural network (CNN) or a physics-based Integrated Mass Enhancement (IME) method. The algorithm can process 62 128×128 images per second on a single 2.6 GHz Intel Core i7 CPU. We train the algorithm with large-eddy simulations of methane plumes superimposed on noisy and variable methane background scenes from the GHGSat-C1 satellite instrument. We introduce the concept of point source observability $O_{ps} = Q/(UW\Delta B)$ as a single dimensionless number to predict plume detectability and source rate quantification error from an instrument as a function of source rate Q, wind speed U, instrument pixel size W, and instrument-dependent background noise ΔB. We show that O_{ps} can powerfully diagnose the ability of an imaging instrument to observe point sources of a certain magnitude under given conditions. U-Plume successfully detects and masks plumes from sources as small as 100 kg h^{-1} in GHGSat-C1 images over surfaces with low background noise and succeeds for larger point sources over surfaces with substantial background noise. We find that the IME method for source quantification is unbiased over the full range of source rates while the CNN method is biased toward the mean of its training range. The total error in source rate quantification is dominated by wind speed at low wind speeds and by the masking algorithm at high wind speeds. A wind speed of 2-4 m s^{-1} is optimal for detection and quantification of point sources from satellite data.

Keywords: Methane, Point Sources, Satellites, Machine Learning, Image Segmentation

1 Introduction

A number of satellite instruments can now detect and image methane column plumes with spatial resolution finer than 60 m by solar backscatter in the shortwave infrared (SWIR), enabling quantification of large point sources from individual facilities (Jacob et al., 2022). As the satellite observing system expands for both methane and other gases, there is a growing need for efficient methods of detecting and quantifying these point sources by automated processing of vast amounts of data. Here we

Deleted: core.

Deleted:

Deleted: 50

present the U-Plume algorithm, a generalized machine-learning method to address this need, and we apply it to methane observations from GHGSat (Jervis et al., 2021).

Several methods have been proposed for inferring point source rates from satellite imagery of atmospheric pollution plumes, including for methane (Varon et al. 2018), CO_2 (Nassar et al. 2017), NO_2 (Valin et al. 2013; De Foy et al. 2015; Beirle et al. 2021), CO (Pommier et al. 2013), SO_2 (Fioletov et al. 2015; McLinden et al. 2016), and NH_3 (Clarisse et al. 2019; Dammers et al. 2019; Noppen et al. 2023). Gaussian plume inversion (Bovensmann et al., 2010; Krings et al., 2011, 2013) and mass balance methods (Jacob et al. 2016; Buchwitz et al., 2017) have both shown success in estimating emissions for very large plumes (>10 km) but they fail when applied to sub-km plumes due to stochastic turbulence that leads to non-Gaussian behavior and dominance of eddy flow (Varon et al., 2018). Temporal averaging of plumes over multiple satellite passes with wind rotation has been used extensively for large plumes from well-identified point sources to decrease noise and enhance Gaussian behavior (Pommier et al. 2013; Fioletov 2015; McLinden et al. 2016; Varon et al., 2020; Massakers et al. 2022), but is difficult to implement for methane plumes that are typically smaller and intermittent (Frankenberg et al., 2016; Cusworth et al., 2021, Irakulis-Loitxate et al., 2021, Ehret et al., 2022, Thorpe et al., 2023).

Formatted: Font: Not Italic, Font color: Auto, Complex Script Font: Not Italic

Two methods that have shown success in estimating emissions from high-resolution instantaneous plume imagery are the Integrated Mass Enhancement (IME) and Cross-Sectional Flux (CSF) methods (Krings et al. 2011, 2013; Frankenberg et al. 2016; Varon et al. 2018). The IME method relates the total observed plume mass to a source rate, using information from the plume size and the local wind speed. The CSF method integrates concentrations over plume cross-sections perpendicular to the wind direction and multiplies them by the local wind speed to infer a source rate. Both methods require the plume to be identified and masked (i.e., delineated) within the image.

Detection and masking of plumes in satellite scenes has generally been done by human analysts (Guanter et al., 2021) but this is not practical operationally. Simple statistical thresholding approaches combined with adjacency criteria have been developed to detect methane enhancements above background for plume masking (Varon et al., 2019; Duren et al. 2019) but these are vulnerable to retrieval artifacts, particularly from surface features mistakenly identified as methane plumes (Cusworth et al., 2019).

Machine learning is a promising avenue for automating plume detection and inference of point source emissions. Jongaramrungruang et al. (2022) proposed a Convolutional Neural Network (CNN) to estimate source rates directly from aerial AVIRIS-NG methane imagery. Their method (called MethaNet) provides source rate estimates based solely on methane column retrieval fields, forgoing the use of external information on wind speed by exploiting wind information contained in the morphology of the plumes (Jongaramrungruang et al., 2019). Training and testing of MethaNet has been done on aircraft imagery with relatively clean backgrounds and high pixel resolution (3 m), but to our knowledge it has not been applied to

satellite imagery which is coarser and more vulnerable to surface artifacts. MethaNet does not produce a plume mask (outline of plume boundaries), but such a mask offers important visual information and can help identify artifacts. Joyce et al. (2023) used a U-Net architecture to mask methane plumes in PRISMA satellite imagery directly from radiances and again without wind speed information. Separate networks were then used for inferring methane column concentration and estimating emission rates. In both MethaNet and the Joyce et al. (2023) approach there is a bias towards the training mean in the final source rate estimates, resulting in overestimate of low emitting plumes and underestimate of high emitting plumes. This is a common issue in machine learning applications.

Here we perform binary segmentation of methane column imagery from GHGSat using a U-Net neural network architecture to produce plume masks, and we then use these masks together with wind speed information to infer point source rates either with a CNN or with the physics-based IME method. We use wind speed information because some is always available, either from local measurements or from regional/global reanalysis datasets. The U-Net architecture (Ronneberger et al. 2015) has previously demonstrated to be effective in feature recognition by binary image segmentation from satellite imagery (Joyce et al. 2023), avoiding the false positives present in more traditional methods reliant on thresholding (Rezvanbehbahani et al. 2020). Combining the U-Net architecture to detect plumes in column enhancement with subsequent steps to infer point source rates forms an end-to-end algorithm that we call U-Plume.

2 Methods

We present here the U-Plume algorithm for identifying point source plumes in satellite imagery of methane concentrations and inferring point source rates. The algorithm has two components: (1) a U-Net to detect plumes by binary segmentation of the pixels in the satellite scene (0 = background, 1 = plume), producing a plume mask; (2) two alternative CNN and IME approaches for inferring source rates from the plume mask. We demonstrate U-Plume on the GHGSat-C1 instrument for methane, but the architecture and training process presented here are potentially applicable to any satellite point source imagers for any species.

GHGSat-C1 is the first operational instrument launched in 2020 by GHGSat Inc., following the demonstration instrument GHGSat-D launched in 2016 (Jervis et al., 2021). The instrument is a shortwave IR interferometer with an observation domain of ~12×12 km^2 at 25×25 m^2 pixel resolution. Retrieval of the backscattered solar spectrum at the 1.65 µm methane absorption band yields an estimated methane column mass enhancement over background reported as mol m^{-2} with 1-2% estimated precision. The instrument is in polar sun-synchronous orbit with observation at ~9:30 am local time. Subsequently launched GHGSat-C2-C11 instruments have less background noise than -C1. Our results can be extended to C2-C11 and any other instruments using the background noise metrics presented below.

Deleted: 8

Deleted: 8

Our first step is to create a training data set of plume-containing satellite images in which synthetic instantaneous plumes of known source rates are superimposed on actual plume-free images. Following Varon et al. (2018) and Jongaramrungruang et al. (2019), we use the Weather Research and Forecasting Model Large Eddy Simulation (WRF-LES) to create a diverse plume dataset of atmospheric methane concentration enhancements where the relationship between source rate and plume concentrations is known. Point source methane emission and atmospheric transport are simulated on a 25×25 m^2 horizontal and 15-m vertical grid over a 9×9×2.4 km^3 domain, with the point-source emitting at the surface 5/6 from the downwind edge of the domain. We conduct 4 simulations for different meteorological conditions, each run for 3 hours with the first hour used as spin-up, resulting in 2 hours of usable plume images per simulation. Time steps for WRF-LES integration are 1/4 second and instantaneous plumes are sampled every 30 seconds. Mean wind speeds for the simulations are 3-9 m s^{-1} with sensible heat fluxes of 100-300 W m^{-2} and mixed layer depths of 500-2000 m. This is the same ensemble as used by Varon et al (2021) where the specific details of the simulations are provided.

To create an effective neural network, the training imagery must be as close as possible to the real observations that the model will be applied to. Previous work by Varon et al. (2018) using WRF-LES to simulate synthetic plumes superimposed these plumes on a white-noise background to generate a training data set, but variable surface albedo and terrain can lead to heterogenous noise fields with complex structure. To create realistic noise for our training images we start from a set of actual GHGSat-C1 observations of plume-free scenes and add the methane column enhancements from the WRF-LES simulation. We use the standard deviation of the pixel enhancements in the plume-free scene (ΔB, kg m^{-2}) as measure of background noise:

$$\Delta B = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}} \quad (1)$$

Here n is the number of pixels in the image, x_i is the column concentration for individual pixel i, and x is the mean column concentration in the scene. When discussing ΔB we express it as a percentage of the global mean background concentrations (taken as x_b = 0.011 kg m^{-2}; Jervis et al., 2021). We use 28 12×12 km^2 GHGSat-C1 plume-free observations corresponding to a variety of surface types. Scenes with $\Delta B < 5\%$ correspond to homogenous bright surfaces including arid and grassland terrain. Scenes with $5\% < \Delta B < 10\%$ correspond to moderately heterogeneous surfaces. Scenes with $\Delta B > 10\%$ correspond to very heterogeneous or dark surfaces including forests, wetlands, and urban areas. Values of ΔB are instrument dependent. The more recent GHGSat-C2 and beyond instruments have ΔB < 2% across most observing conditions (Ramier, 2022). On the other hand, hyper/multi-spectral land surface mappers can have background noise exceeding 10% (Cusworth et al., 2019; Varon et al., 2021).

Deleted: ¶

Deleted:

Deleted: and

Deleted: all scenes and

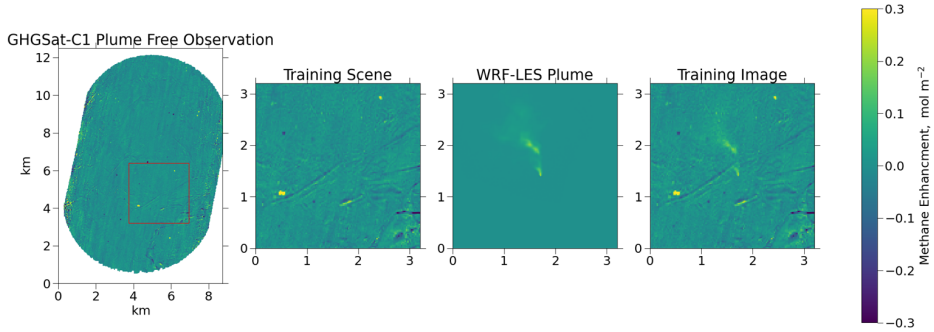


Figure 1: Sample generation of U-Plume training imagery. The first panel shows a sample plume-free observation of methane column concentration enhancements observed by GHGSat-C1. Panel 2 is a 128×128-pixel scene randomly selected from the full observation domain. Panel 3 shows a sample WRF-LES instantaneous plume for a source rate of 2000 kg h^{-1} (we choose a large plume here for easy visualization). Panel 4 shows the result of adding panels 2 and 3 to form a training image. 0.1 mol m^{-2} corresponds to a dry column mixing ratio enhancement of 284 ppb.

Fig 1 illustrates the steps for creating the synthetic plume observations used in training and testing our method. For each training image, a 128×128-pixel scene is randomly selected from a plume-free GHGSat-C1 observation. A random plume from the WRF-LES dataset is then selected, with a random rotation and translation applied to situate the plume anywhere in the scene but with a safeguard to ensure that the plume is fully contained in the scene. Since the methane enhancement in the plume is proportional to source rate, we can scale the plume randomly to correspond to a given source rate. This plume image is then added to the selected scene to create a training image. A training range of plumes emitting in the 500-2000 kg h^{-1} is chosen to ensure that the plumes are at least partially visible in most training images. Though we want the network to be able to detect smaller source rates, it is important that the structural features being searched for are visible above noise within the training dataset. Because the U-Net recognizes structure, not just enhancement, the network trained on the 500-2000 kg h^{-1} range is still usable outside of this range as we will see. Because pixel values for methane enhancement generally fall in the -0.3-0.3 mol m^{-2} range we found that it is not necessary to normalize the data for training.

Our set of training/test images is created using 28 plume-free observations. We step the 128×128-pixel scene in 16-pixel steps to create a set of scenes. The ΔB values of these scenes range from 1% to more than 50%. We only use scenes with $\Delta B < 20\%$ to avoid training on scenes with extremely high background variability where detection of any but the largest plumes would be intractable. This filtering leaves us with a total of 6870 scenes with a median ΔB of 8% and interquartiles of 5% and 12%. We then apply the random plume placement and source magnitude as described above to create 6870 simulated plume observations. This is a relatively small set for training by machine learning standards, but it is sufficient to provide a successful

Formatted: Indent: First line: 0"

Deleted: 5% being the most common value of .

Deleted: xx

Deleted: xx

Deleted: xx

Deleted: ,

Deleted: (>1000 images), a small number of scenes (<400 images) at $\Delta B < 4.5\%$, and the remaining images being an approximately linear descent in frequency down to <250 images with $\Delta B > 19\%$.

network which could be easily bolstered with a future collection of additional LES simulations and background images. The true plume mask is defined as pixels where the enhancement from the LES simulation is higher than the ΔB of the scene, which is the criterion for the pixel contributing more information than noise (Varon et al., 2018). 90% of the images are used for training and 10% are set aside for testing. We use a relatively standard configuration of the U-Net model with a modest training period and therefore do not separate a validation set for specific learning rate stopping criteria. The metric we use for masking success is the Jaccard score, which is the intersection of the predicted mask (A) and true mask (B) over the unity of the two:

$$J(A,B) = \frac{A \cap B}{A \cup B} \tag{2}$$

We train the model for 20 epochs with a batch size of 32 images minimizing a loss function (L) combining binary cross-entropy (BCE, Jardon 2020) and the Jaccard score defined as:

$$L = -\ln\big(J(A,B)\big) + \mathrm{BCE} \tag{3}$$

Formatted: Font: Italic, Complex Script Font: Italic

Fig. 2 shows the two-step process used to obtain source rate estimates from methane enhancement images. First, a U-Net masking network is used to identify pixels in which a plume is present. The binary mask created by this network is then fed along with the original enhancement image and external wind speed data to estimate the source rate from the identified plume. The enhancement image with mask is preserved for visualization and as a quality control diagnostic. Processing the test dataset of 687 images, preloaded from their relevant files, in 32 image batches on a single 2.6 GHz Intel Core i7 CPU takes 11 seconds, corresponding to 62 images per second.

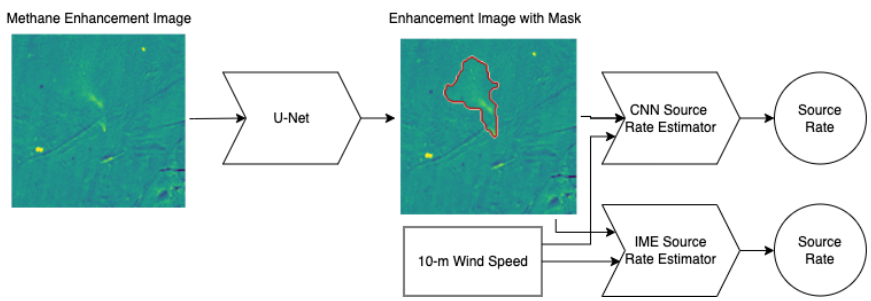


Figure 2: U-Plume architecture. Starting from a methane enhancement image (the sample image of fig.1), we use the U-Net machine learning algorithm to produce a plume mask (in red) of pixels containing more information than noise. All masked pixels in a single 128×128-pixel (3.2×3.2 km^2) scene are assumed to originate from the same plume. We then use the original enhancement image, the mask, and local 10-m wind speed to quantify emissions using either a machine learning method or an Integrated Methane Enhancement (IME) method.

The U-Net neural network for plume masking (Ronneberger et al. 2015) uses an encoder-decoder framework to recognize structural features within imagery. The network receives a methane enhancement image as in Fig. 2 and classifies each pixel with a confidence level for the presence of a plume. We apply a threshold of 50% confidence to produce a binary plume mask. The main advantage of this method over traditional threshold masking is that high enhancement features that do not have plumelike shape characteristics will not be falsely captured as plumes.

The encoder portion of the network applies a series of convolutional filters and pooling layers to identify features of increasing complexity within the image. At the end of the encoding the image has been compressed spatially and has many channels corresponding to the various feature maps created by the convolutional filtering. By compressing the image spatially in this way, the network can recognize connections between distant features.

The decoder portion of the network serves to interpret these identified features as criteria for classification. An inverse convolutional operator is used to expand the spatially compressed image back to the original size of the input image while reducing the number of filters applied, which compresses the number of channels in the image. The final step of the network uses the information gathered from the encoder-decoder process to estimate a confidence level (0-1) that each pixel is or is not part of a plume. Output from this model for a single image is a confidence mask for plume location across the enhancement image as shown in Figure 2. In practice, the confidence mask is generally either very close to 1 or very close to 0. Intermediate values fall off very quickly at the edge of the mask. We differ from the Ronneberger et al. structure only in that we begin with 16 convolutional layers rather than 64 and thus reach a maximum channel depth of 256 rather than 1024 at the end of the encoder path. A loss function combining binary cross entropy and Jaccard score is minimized for the training set. The code used for the model creation and loss function can be found alongside our dataset in the repository at the end of this text (https://doi.org/10.7910/DVN/YFRQU4)

Once the U-Net mask has been generated as the first step of our U-Plume algorithm, we infer the source rate as a second step using as options either a CNN method or an IME method. The CNN method takes in a 3-channel input image containing the original image, the binary mask, and 10-m wind speed. Initial testing indicated that including wind direction did not improve performance. The wind speed channel is populated with a uniform 15-minute scene average value from the LES simulation. The CNN uses a series of convolutional and pooling layers which connect to a set of fully connected layers, rather than a decoder, to yield a scalar estimate of emissions. The CNN structure is identical to that of the U-Net truncated at the end of the encoder path and connected to 32 and then 16 node dense layers with a single node output layer. All 3 of the added layers use a relu activation function and the model is trained using a mean squared error loss. We train the model for 10 epochs with a batch size of 32 images and use U-Net generated masks for the mask input channel (limited to successfully masked examples).

Deleted: .

Moved (insertion) [1]

Moved up [1]: Initial testing indicated that including wind direction did not improve performance. The wind speed channel is populated with a uniform 15-minute scene average value from the LES simulation. The CNN uses a series of convolutional and pooling layers which connect to a set of fully connected layers, rather than a decoder, to yield a scalar estimate of emissions.

The IME method follows Varon et al. (2018), in which the plume mass enhancement (IME) is combined with a parameterized effective wind speed (U_{eff}) and plume length scale (L) to infer the source rate (Q, kg h^{-1}):

$$Q = \mathrm{IME}\frac{U_{eff}}{L} \tag{4}$$

The plume length scale L is defined as the square root of the plume mask area. The effective wind speed is fitted to the 15-minute averaged wind speed U_{10} at 10-m altitude using the training data set where the relationship between Q, IME, and L is known for a given U_{10}. This yields

$$U_{eff} = 0.7 + 0.23\, U_{10} \tag{5}$$

with U_{eff} and U_{10} in units of m s^{-1}. The intercept of 0.7 m s^{-1} arising from the fit can be physically interpreted as a minimum turbulent diffusion for the plume at low wind speeds. This calibration is specific to the GHGSat-C1 instrument and should be recalculated when applying U-Plume to other platforms.

3 Dimensionless Point Source Observability Number

We introduce the concept of point source observability (O_{ps}) as a dimensionless number to determine the ability of a satellite instrument to detect and quantify point sources. Observability is a function of source rate (Q, kg s^{-1}), wind speed (U, m s^{-1}), instrument pixel resolution (W, m) which is 25 m for our case, and scene-dependent background noise (ΔB, kg m^{-2}, as defined by equation (1)).

The idea behind the O_{ps} concept is that the observability of point sources is determined by the signal-to-noise ratio in the plume (Varon et al., 2018). The column concentration enhancement in the plume (kg m^{-2}) scales as Q/UW (Jacob et al., 2016). The noise for a given scene (kg m^{-2}) is given by ΔB which is a scene-dependent property of the instrument that can be characterized as a function of surface type. ΔB can be viewed as the precision of the instrument accounting for the contribution from surface artifacts. The dimensionless point source observability number O_{ps} is then given by

$$O_{ps} = \frac{Q}{UW\Delta B} \tag{6}$$

and is a measure of the signal-to-noise ratio. We will use it here to interpret our results, but it can be applied more generally to any remote sensing observation of point sources.

4 Results

4.1 Plume Detection and Masking

Fig. 3 shows the relationship between source rate and U-Net masking success. In 2% of cases no mask is produced for the image, and in 30% of cases there is no overlap between the true and predicted masks. These failures occur for high background noise ($\Delta B > 5\%$) and lower emission rates. On the other hand, Jaccard scores generally exceed 0.5 for low background noise ($\Delta B < 5\%$).

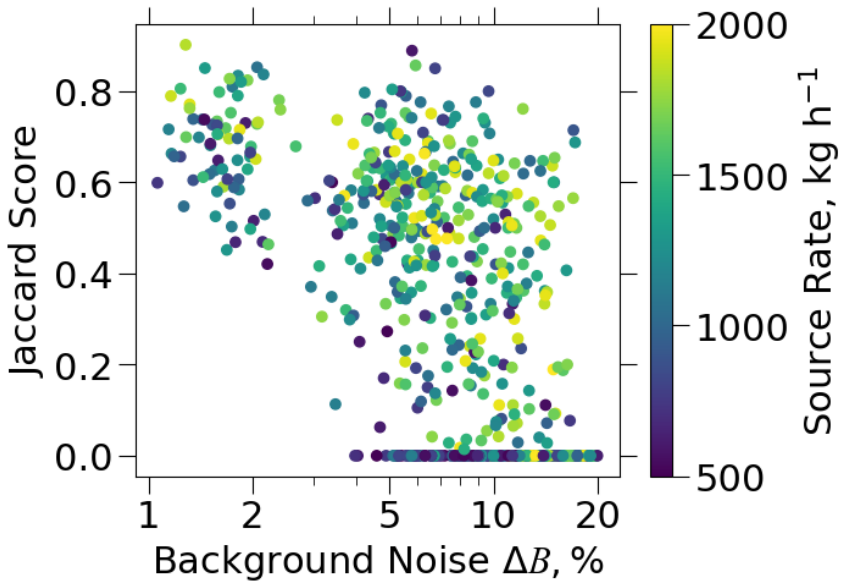


Figure 3: Ability of the U-Net machine learning algorithm to delineate (mask) plumes emitted from point sources. Masking success is measured by the Jaccard score (equation (2)) and is plotted as a function of background noise ΔB (equation (1)) for a range of point source rates from 500 to 2000 kg h^{-1}. Results are for the test set of WRF-LES plume images superimposed on noisy plume free observations from the GHG-Sat C1 satellite instrument.

Deleted: The metric we use for masking success is the Jaccard score, which is the intersection of the predicted mask (A) and true mask (B) over the unity of the two:

Deleted: <

Deleted:

Deleted: 5

We examine false positives by applying the network over the test dataset of backgrounds without plumes added. This yields masks in 14% of cases (false positive detection rate). The masks produced in these false positives are generally very small, with a median size of 5 pixels. Applying a minimum 5-pixel mask size brings the false positive rate down to 6% while losing only 1.6% of the true positive detections in the test data set and removing 92% of the zero overlap masks. False positives occur most comonly in scenes with high ΔB values and particularly in scenes with complex topography. The mask sizes for these false positives are generally small, as is apparent in the mask size filtering effectiveness, and their corresponding estimated source rates are therefore also generally small.

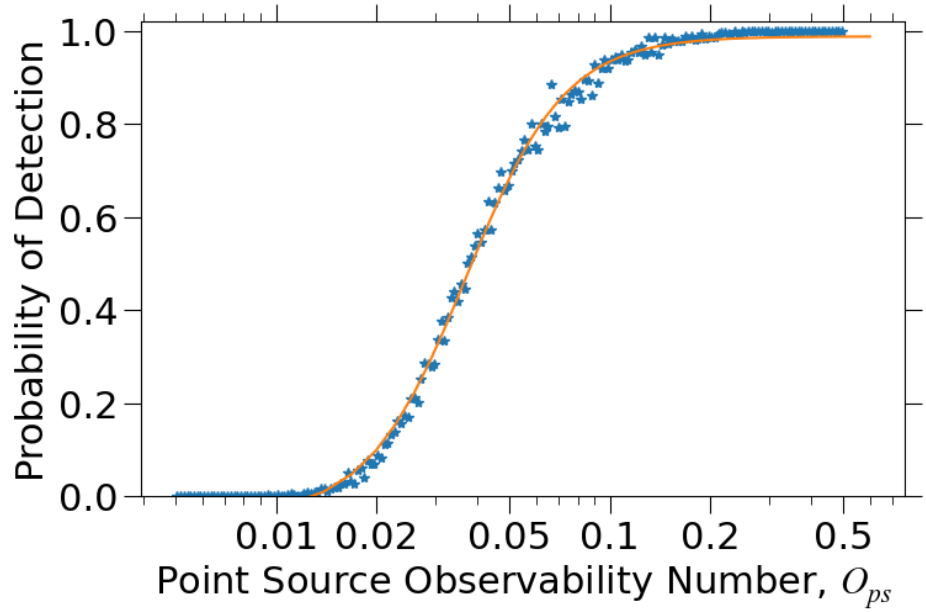
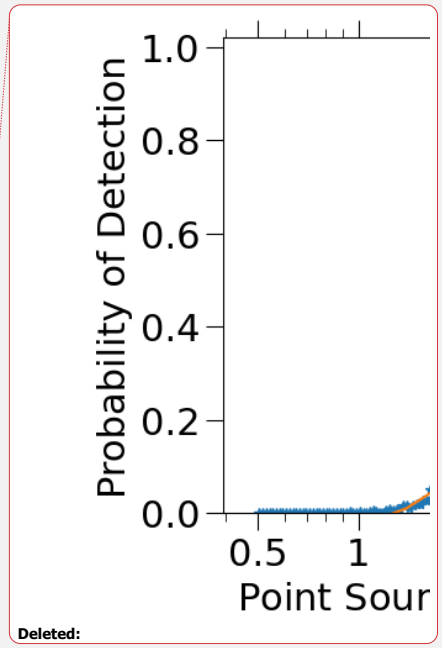


Figure 4: Probability of detection ($J > 0.1$) for point sources in the GHGSat-C1 augmented test data set as a function of the point source observability dimensionless number O_{ps} (equation (6)). Calculation of O_{ps} uses the WRF-LES 10-m wind speed for U and the GHGSat-C1 pixel size $W = 25$ m.

The point source observability (O_{ps}) in equation (6) provides a scaling dimensionless number to better understand how point source detection for a given instrument relates to the combination of source rate, background noise, wind speed, and pixel size.

Deleted: re



Deleted:

Deleted: 4

Deleted: 4

Here we use the 10-m wind speed (U_{10}) in the calculation of O_{ps}. We create an augmented test dataset with 10 subset images from each of the 28 plume-free observations (Figure 1) and 20 emission levels for each image from 100 to 2000 kg h^{-1} in 100 kg h^{-1} increments. We bin images in the augmented test dataset by their O_{ps} ranging from 0.005 to 0.5 in 201 log-spaced bins, and for each bin with a minimum of 50 images we calculate the probability of detection as the fraction of plumes in the bin with $J > 0.1$. Figure 4 shows the relationship between the probability P of detection and O_{ps}. The relationship for $O_{ps} > 0.014$ can be tightly fit to a sigmoid function of the form:

$$P = \frac{1.03}{1 + e^{-2.9(log(O_{ps})+3.3)}} - 0.05 \qquad (O_{ps} > 0.014) \tag{7}$$

Detection probability is 10% for $O_{ps} = 0.02$, 50% for $O_{ps} = 0.04$, and 90% for $O_{ps} = 0.08$. The transition from low detectability (<10%) to high detectability (>90%) is very sharp in O_{ps} space, demonstrating the power of the O_{ps} metric for evaluating the ability of plume imaging instrument to observe point sources. The value $O_{ps} = 0.04$ for the 50% detection threshold can be interrpreted as a characteristic column-averaged concentration enhancement $0.04Q/UW$ in the plume. The maximum column-averaged concentration enhancement in the source pixel would be Q/UW (Jacob et al., 2016).

Deleted: 50

Deleted: 1.4

Deleted: –

Deleted: *1.3*

Deleted: *1.4*

Deleted: *6*

Deleted: 4

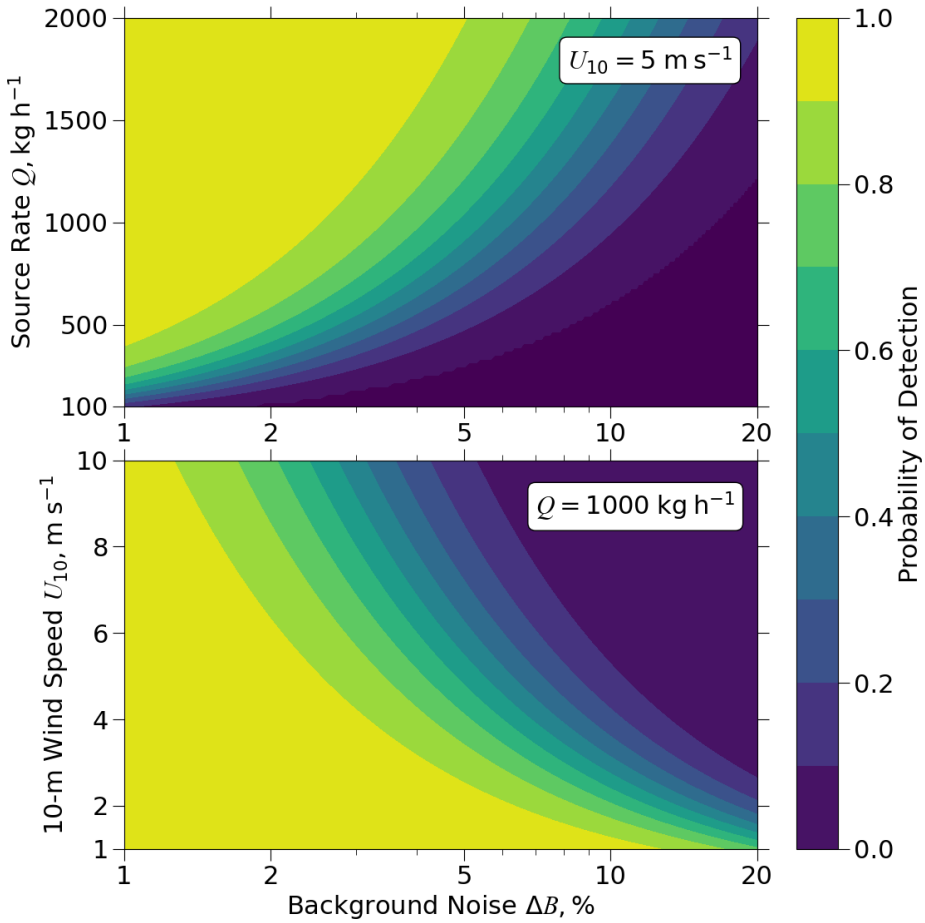


Figure 5: Probability of detection ($J > 0.1$) for point sources calculated using equation (2) as a function of background noise (equation (1)), source rate (top panel), and 10-m wind speed (bottom panel).

Deleted: 6

Fig. 5 shows the probability of detection calculated using equation (7) as a function of source rate, 10-m wind speed, and background noise. For low values of ΔB, the probability of detection is very sensitive to changes in Q and less sensitive to changes in U_{10}. With ΔB = 1%, U = 5 m/s, and W = 25 m, 90% probability of detection is achieved for Q > 400 kg h^{-1}.

Deleted: 6

Commented [DJ1]: Need to explicitly mention these along with Delta B since result depends on them

Deleted: With ΔB = 1%,

Deleted: Conversely, for high values of ΔB the probability of detection is very sensitive to changes in U_{10} and less sensitive to changes in Q.

4.2 Source Rate Estimation

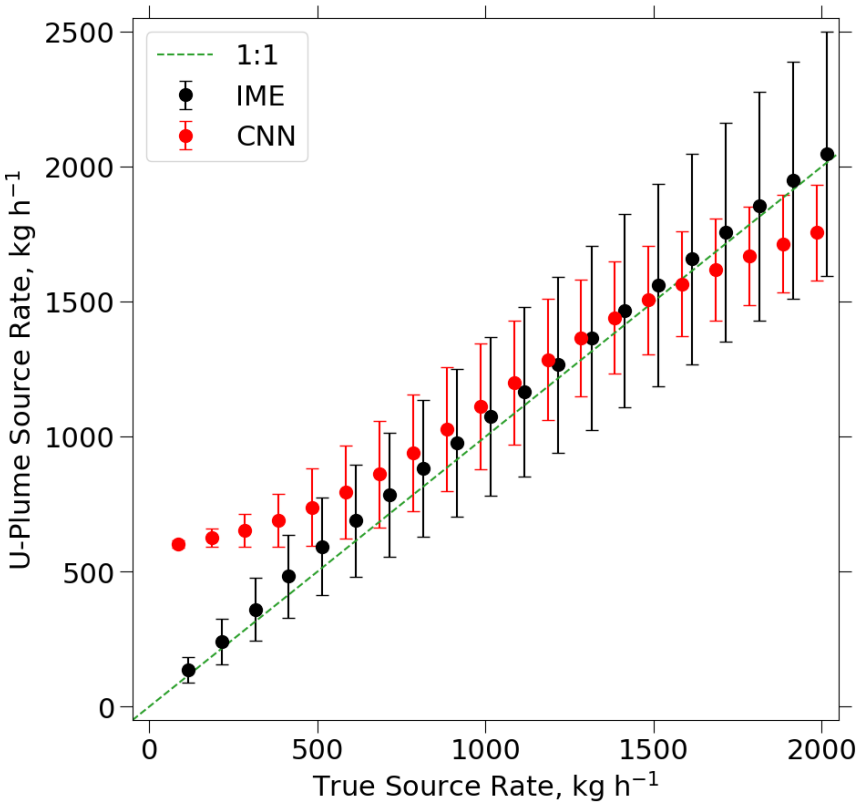


Figure 6: Source rate estimates from the alternative CNN and IME methods (Figure 2) for point sources emitting in the 100-2000 kg h^{-1} range (discrete 100 kg h^{-1} bins) in the augmented test dataset. The methods were trained with source

rates in the 500-2000 kg h^{-1} range. Only images with Jaccard scores > 0.1 are processed. Error bars show 1 standard deviation of estimates.

Fig. 6 shows the source rate estimates for the complete U-Plume workflow, starting from the plume imagery and applying either the CNN or the IME method for source rate quantification. We restrict our analysis to plumes with Jaccard score >0.1 to enable generalization to instruments with less background noise than GHGSat-C1 (such as GHGSat-C2) or to the use of improved filtering such as requiring a minimum number of plume pixels. Results are for the augmented dataset covering the 100-2000 kg h^{-1} range but trained only over the 500-2000 kg h^{-1} range. We assume for now no error on wind speed.

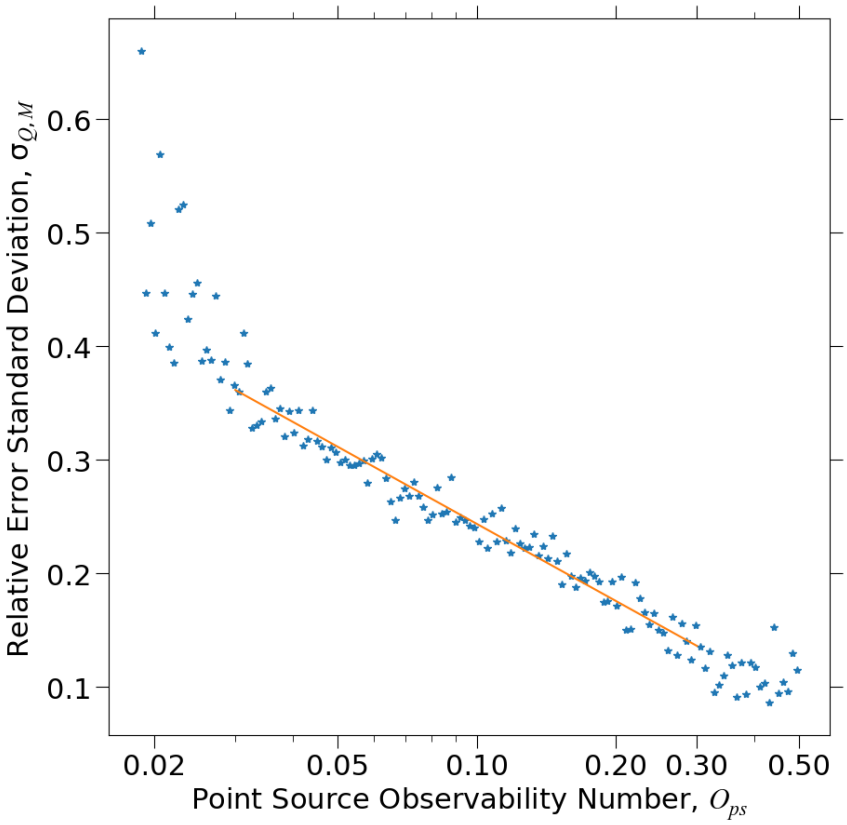
The CNN method has a smoothing bias where low source rates are overestimated, and high source rates are underestimated. This smoothing towards the mean is a common issue in machine learning methods and appears in the work of Jongaramrungruang et al. (2022) and Joyce et al. (2023). In addition, the CNN method is unable to extrapolate outside its training range. By contrast, the physically based IME method can successfully fit source rates all the way down to 100 kg h^{-1} as long as the plumes are detectable. Discussion from this point forward will focus on the IME method as performing better than the CNN method over the full range of source rates. We retain the CNN method as an option in the U-Plume algorithm because it offers an alternative source rate estimate for verification purposes and because it could be improved in the future with a more extensive training dataset. A potential avenue for future work would be the testing of other CNN architectures and the expansion of training data to include the range of all possible source rates.

Deleted:

4.3 Error Analysis

The uncertainty on source rate estimates using U-Plume with the IME method can be expressed as the relative error standard deviation (σ_Q) of the U-Plume estimate versus the true value. It has two independent components: (1) error in the U-Plume algorithm, including the masking process and the IME parameterization ($\sigma_{Q,M}$); and (2) error in the 10-m wind speed ($\sigma_{Q,U}$). Here we derive expressions for $\sigma_{Q,M}$ and $\sigma_{Q,U}$ and add them in quadrature to quantify the overall error and determine the driving factors.

The relative error standard deviation $\sigma_{Q,M}$ from the U-Plume algorithm can be parameterized using the augmented test data set binned by point source observability O_{ps} following the approach in section 4.2 using only bins with >50 positive detections (J > 0.1). Figure 7 shows the relative error standard deviation plotted against O_{ps}.



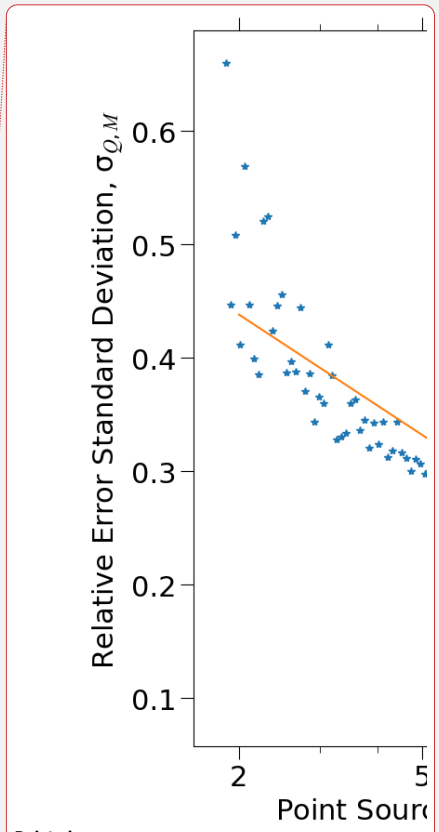


Figure 7: Relative error on source rate estimation by the U-Plume algorithm. The figure shows the relative error standard deviation $\sigma_{Q,M}$ versus the point source observability dimensionless number O_{ps} (equation (6)). Errors are calculated for point sources emitting in the 100-2000 kg h^{-1} range (discrete 100 kg h^{-1} bins) in the augmented test dataset, not including error in wind speed (for which the error is calculated separately, see text). Errors from individual images are binned by values of their point source *ob*servability (equation (6)) for O_{ps} = 0.005, 0.5, in 201 log-spaced intervals. The orange line is a reduced-major-axis (RMA) regression fit (R^2 = 0.89) for $\sigma_{Q,M}$ versus O_{ps} (equation (8)).

Fitting the error to log (O_{ps}) yields:

Deleted:

Formatted: Font: 9 pt, Complex Script Font: 9 pt

Deleted:

Deleted: 4

Deleted: 4

Deleted: 5

Deleted: 50

Deleted: 7

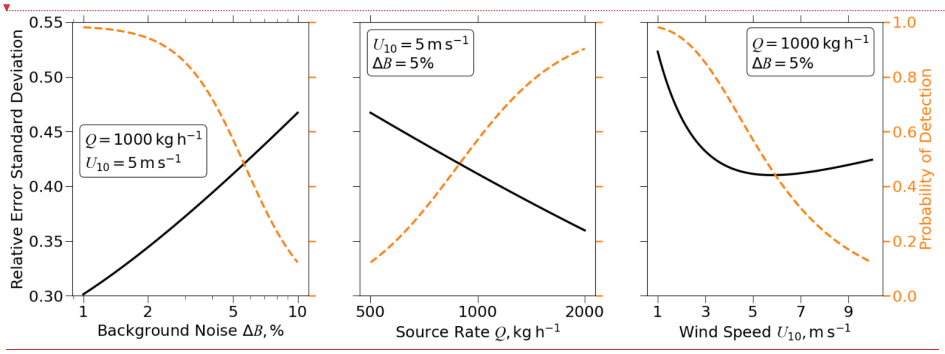
$$\sigma_{Q,M} = \max(0.1, (0.018 - 0.098 \log(O_{ps})) \quad (0.03 < O_{ps} < 0.3) \tag{8}$$

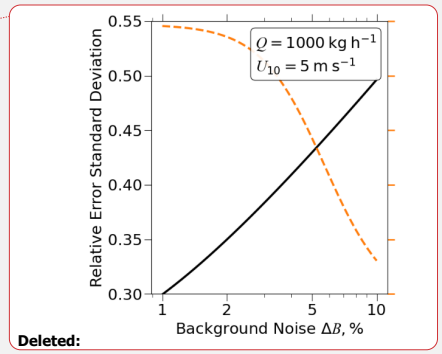
For values of $O_{ps} < 0.02$ the probability of detection is less than 10% (section 4.1). The relative error standard deviation bottoms out at 10% for highly observable plumes ($O_{ps} > 0.3$), reflecting irreducible error inherent to the IME method.

Uncertainty in wind speed adds another source of error. Let σ_U (m s^{-1}) denote the error standard deviation in the 10-m wind speed U_{10}. Replacing in equations (4) and (5) yields a corresponding relative error standard deviation $\sigma_{Q,U}$ on the inferred IME source rate Q:

$$\sigma_{Q,U} = \frac{0.23\,\sigma_U}{0.7 + 0.23\,U_{10}} \tag{9}$$

Wind speed data may be available locally or from meteorological analysis datasets. A global default is the NASA Goddard Earth Observation System – Fast Processing (GEOS-FP) data publicly available on a 0.25° × 0.3125° grid at 1-h temporal resolution. Varon et al. (2018) estimated an error standard deviation $\sigma_U = 2$ m s^{-1} for the GEOS-FP data, independent of wind speed magnitude, by comparison to 5-minute observations at US airports. Replacing into equation (8) indicates a maximum relative error standard deviation of 66% at low wind speed, decreasing with increasing wind speed. At high wind speeds the error becomes small but the IME is then small so that the masking error is large as inferred from the inverse dependence of the point source observability on wind speed (equation (6) and Figure 7).





Deleted: 0.52
Deleted: 0.12
Deleted: >
Deleted: 2
Deleted: 7
Deleted: 30
Deleted: 2
Deleted: 3
Deleted: 4
Deleted:

Figure 8: Total relative error standard deviation σ_Q on point source rate estimates from the U-Plume algorithm with the IME method (equation (9)) as a function of three variables: background noise ΔB (%), source rate Q (kg h^{-1}), and 10-m wind speed U_{10} (m s^{-1}). The dashed orange line represents the probability of point source detection calculated with equation (6).

Adding the error variances from equations (8) and (9) in quadrature gives the total relative error variance ${\sigma_Q}^2$ on the estimated sources rates from the U-Plume algorithm including error on wind speed:

$$ {\sigma_Q}^2 = \left(\max\left(0.1, \left(0.018 - 0.098 \log\left(O_{ps}\right)\right)\right)\right)^2 + \left(\frac{0.23\ \sigma_U}{0.7 + 0.23\ U_{10}}\right)^2 \qquad (0.03 < O_{ps} < 0.3) \qquad (10) $$

Fig. 8 shows the total relative error standard deviation σ_Q as a function of background noise ΔB, source rate Q, and 10-m wind speed, U_{10}. computed from equation (10) for common observing conditions. Equation 10 can be used to calculate the values at any desired combination within the limits of the composing equations. The total relative error standard deviation increases linearly with log (ΔB), decreases linearly with log (Q), and is minimum for U_{10} = 4 m s^{-1}.

Deleted: 7

Deleted: 8

Deleted: max*(0.1,0.52−0.12*log*Ops)*

Deleted: *(Ops>2)*

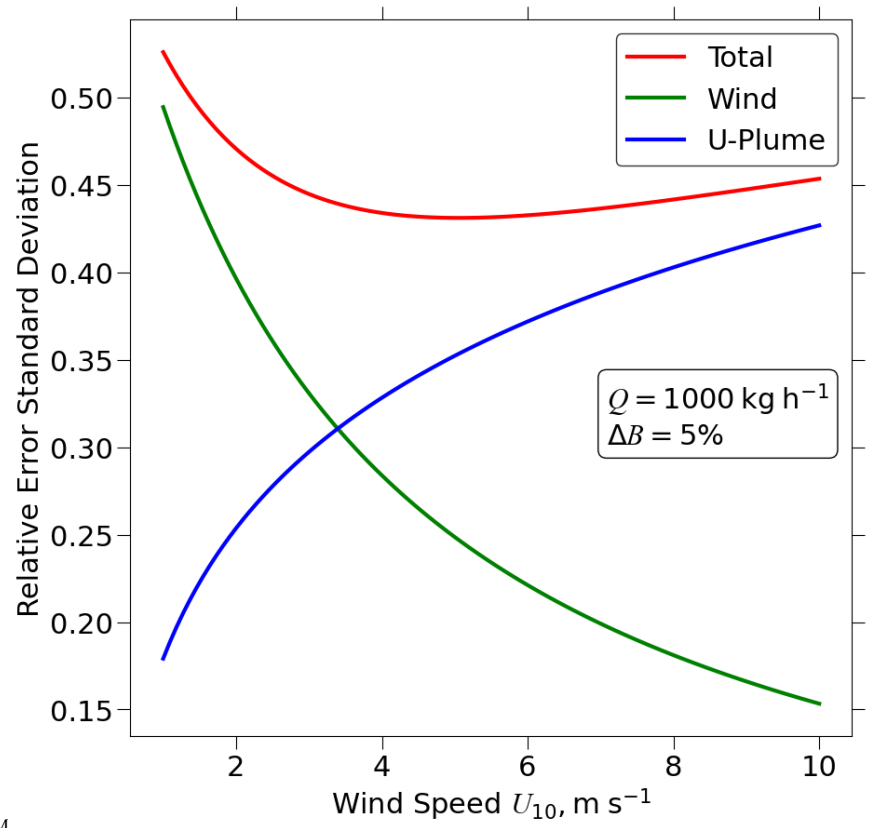
Deleted: *9*

Deleted: We attempt to represent the range of

Deleted: , but

Deleted: e

Deleted: The



M

Figure 9: Dependence of errors in point source estimates on wind speed. The Figure shows the relative error standard deviation on the source rate estimates inferred from U-Plume with the IME method (equation (10)) as a function of 10-m wind speed (U_{10}). The total relative error standard deviation σ_Q is shown as the red line (same as Figure 8) and is decomposed into the contributions from errors in the U-Plume algorithm ($\sigma_{Q,M}$ equation (8)) and errors in wind speed (σ_U equation (9)).

Deleted: 9

Deleted: 7

Deleted: 8

The dependence of the total relative error standard deviation on wind speed is partitioned further in Figure 9 into the contributions from the U-Plume algorithm and wind speed errors. The error contribution from the U-Plume algorithm increases

with rising wind speed, because methane enhancements in the plume become smaller, but this levels off at high wind speeds as the plumes are increasingly likely to be undetected instead. The error contribution from wind speed decreases with rising wind speed. The total error is dominated by wind speed for $U_{10} < 4$ m s^{-1} and by the U-Plume algorithm for $U_{10} > 4$ m s^{-1}. For the range of values of ΔB and Q in our dataset we find that the cross-over between the two regimes is at 2-4 m s^{-1}, which can be regarded as an optimum wind speed range for plume quantification.

5 Conclusions

We developed the U-Plume algorithm for automated detection and quantification of point source rates from high-resolution satellite imagery of atmospheric column concentrations. Our work was motivated by the need for fast identification of plumes in the operational processing of the vast amounts of methane data from the rapidly expanding constellation of GHGSat and other satellite instruments targeting methane point sources. With instrument-specific training, the method can also be applied to any chemically inert plume such as for CO_2, due to the inherent similarity in the physics governing plume formation and transport.

U-Plume involves two steps. The first step is a machine learning U-Net architecture for plume detection and delineation (masking) in noisy satellite images. The resulting plume mask is then used together with wind speed information in a second step of source rate quantification using either a convolutional neural network (CNN) or a physically based Integrated Mass Enhancement (IME) method. Having plume mask as well as source rate information available from U-Plume output is important for visualization, quality control, and pinpointing the source location. The end-to-end algorithm can process 62 images per second on a single core.

We trained U-Plume with an ensemble of plume free observation images from the GHG-Sat C1 satellite instrument covering a range of surfaces from homogeneous to highly heterogeneous and superimposing instantaneous methane plumes from large-eddy simulations with known point source rates. Evaluation with data unseen by the model during the training process shows that successful plume detection and masking is strongly dependent on four state variables: instrument background noise ΔB, source rate Q, 10-m wind speed U_{10}, and pixel resolution W. We combine these variables into a new dimensionless number metric, the point source observability $O_{ps} = Q/(WU_{10}\Delta B)$, and show that this metric can successfully predict the ability of the GHGSat-C1 imager to detect the plumes and quantify source rates for given observing conditions. As the O_{ps} metric is dependent on the fundamental state variables for the detection of an enhancement produced by a methane point source, it should have universal value for predicting the detectability and quantification accuracy of other point source imagers. For example, it can assist in determining when automation is possible for a given target detection level. For the GHGSat-C2 instruments with background noise $\Delta B < 1\%$, we thus expect that U-Plume should reliably detect and quantify point sources with $Q > 400$ kg h^{-1} at $U = 5\ m/s$ wind speeds. For reference, GHGSat's manual plume detection method has achieved a 50% probability of detection in 3 m/s at 117 kg/hr in a self-administered controlled release campaign

Deleted: T

Deleted: with appropriate instrument specific training

Deleted: an independent dataset

Deleted:), and

Deleted: a

Deleted: given plume imaging instrument

Deleted: it

Deleted: of the

Deleted: we expect that the point source observability will be a valuable metric for evaluating

Deleted: after appropriate tuning

Deleted: I

Deleted: find

Deleted: couldan

(https://go.ghgsat.com/validation-and-metrics-for-emissions-detection-by-satellite) and detected the only <200 kg/hr emission in a 3rd-party organized single-blind campaign (https://www.nature.com/articles/s41598-023-30761-2). U-Plume's value comes in the ability to process large sets of plume data rather than pushing the limits of detectability. Human operators with knowledge of point source location would be expected to have greater ability to detect small sources (Sherwin et al., 2022).

We find that the physically based IME method is superior to the CNN method for inferring point source rates from the plume masks. The CNN has high bias for low source rates, and low bias for high source rates, as is typical for machine learning methods. The superiority of the IME method would be expected considering that there is a simple physically-based linear relationship between plume mass and source rate modulated by wind speed. Although a CNN method may not require information on wind speed, such information is always available either from local measurements or from regional/global databases.

We developed an end-to-end error model for the point source rates inferred from U-Plume as a function of O_{ps}. We find that at low wind speeds the error is dominated by wind speed whereas at higher wind speeds the error is dominated by the U-Plume algorithm. The U_{10} value where this shift occurs is typically 2-4 m s^{-1}. This also represents an optimal wind speed range for plume detection and quantification.

U-Plume thus offers a capable tool for fast automated processing of vast amounts of satellite imagery to detect plumes from point sources and quantify point source rates. As the ability of satellite observations to decrease background noise continues to improve, the U-Plume capability to detect and quantify points sources will correspondingly increase.

Acknowledgments. DJJ acknowledges funding from the NASA Carbon Monitoring System. We thank an anonymous reviewer for identifying a coding error in the original version.

Data Availability

The model weights and both the original training data and the augmented test dataset are available through Harvard Dataverse (https://doi.org/10.7910/DVN/YFRQU4).

References

Beirle, S., Borger, C., Dörner, S., Eskes, H. J., Kumar, V., De Laat, A., and Wagner, T.: Catalog of NOx emissions from point sources as derived from the divergence of the NO2 flux for TROPOMI, Earth Syst. Sci. Data, 13, 2995–3012, https://doi.org/10.5194/essd-13-2995-2021, 2021.

Bovensmann, H., Buchwitz, M., Burrows, J. P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., and Erzinger, J.: A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications, Atmospheric Measurement Techniques, 3, 781–811, https://doi.org/10.5194/amt-3-781-2010, 2010.

Buchwitz, M., Schneising, O., Reuter, M., Heymann, J., Krautwurst, S., Bovensmann, H., Burrows, J. P., Boesch, H., Parker, R. J., Somkuti, P., Detmers, R. G., Hasekamp, O. P., Aben, I., Butz, A., Frankenberg, C., and Turner, A. J.: Satellite-derived methane hotspot emission estimates using a fast data-driven method, Atmospheric Chemistry and Physics, 17, 5751–5774, https://doi.org/10.5194/acp-17-5751-2017, 2017.

Clarisse, L., Van Damme, M., Clerbaux, C., and Coheur, P.-F.: Tracking down global NH_3 point sources with wind-adjusted superresolution, Atmospheric Measurement Techniques, 12, 5457–5473, https://doi.org/10.5194/amt-12-5457-2019, 2019.

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 Discussions, 1–29, https://doi.org/10.5194/amt-2019-202, 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, Environ. Sci. Technol. Lett., 8, 567–573, https://doi.org/10.1021/acs.estlett.1c00173, 2021.

Dammers, E., McLinden, C. A., Griffin, D., Shephard, M. W., Van Der Graaf, S., Lutsch, E., Schaap, M., Gainairu-Matz, Y., Fioletov, V., Van Damme, M., Whitburn, S., Clarisse, L., Cady-Pereira, K., Clerbaux, C., Coheur, P. F., and Erisman, J. W.: NH3 emissions from large point sources derived from CrIS and IASI satellite observations, Atmospheric Chemistry and Physics, 19, 12261–12293, https://doi.org/10.5194/acp-19-12261-2019, 2019.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O., and Miller, C. E.: California's methane super-emitters, Nature, 575, 180–184, https://doi.org/10.1038/s41586-019-1720-3, 2019.
Fioletov, V. E., McLinden, C. A., Krotkov, N., and Li, C.: Lifetimes and emissions of SO2 from point sources estimated from OMI, Geophysical Research Letters, 42, 1969–1976, https://doi.org/10.1002/2015GL063148, 2015.

de Foy, B., Lu, Z., Streets, D. G., Lamsal, L. N., and Duncan, B. N.: Estimates of power plant NOx emissions and lifetimes from OMI NO2 satellite retrievals, Atmospheric Environment, 116, 1–11, https://doi.org/10.1016/j.atmosenv.2015.05.056, 2015.

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, Environ. Sci. Technol., 56, 10517–10529, https://doi.org/10.1021/acs.est.1c08575, 2022.

Irakulis-Loitxate, I., Guanter, L., Liu, Y.-N., Varon, D. J., Maasakkers, J. D., Zhang, Y., Chulakadabba, A., Wofsy, S. C.,

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, PNAS, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, 2016.

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, 112671, https://doi.org/10.1016/j.rse.2021.112671, 2021.

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, 14371–14396, https://doi.org/10.5194/acp-16-14371-2016, 2016.

Jadon, S.: A survey of loss functions for semantic segmentation, in: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), arXiv:2006.14822 [cs, eess], 1–7, https://doi.org/10.1109/CIBCB48159.2020.9277638, 2020.

Jervis, D., McKeever, J., Durak, B. O. A., Sloan, J. J., Gains, D., Varon, D. J., Ramier, A., Strupler, M., and Tarrant, E.: The GHGSat-D imaging spectrometer, Atmos. Meas. Tech., 14, 2127–2140, https://doi.org/10.5194/amt-14-2127-2021, 2021.

Jongaramrungruang, S., Frankenberg, C., Matheou, G., Thorpe, A. K., Thompson, D. R., Kuai, L., and Duren, R. M.: Towards accurate methane point-source quantification from high-resolution 2-D plume imagery, Atmospheric Measurement Techniques, 12, 6667–6681, https://doi.org/10.5194/amt-12-6667-2019, 2019.

Formatted: Line spacing: 1.5 lines

Deleted: ¶

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, 112809, https://doi.org/10.1016/j.rse.2021.112809, 2022.

Joyce, P., Ruiz Villena, C., Huang, Y., Webb, A., Gloor, M., Wagner, F. H., Chipperfield, M. P., Barrio Guilló, R., Wilson, C., and Boesch, H.: Using a deep neural network to detect methane point sources and quantify emissions from PRISMA hyperspectral satellite images, Atmospheric Measurement Techniques, 2023.

Krings, T., Gerilowski, K., Buchwitz, M., Reuter, M., Tretner, A., Erzinger, J., Heinze, D., Pflüger, U., Burrows, J. P., and Bovensmann, H.: MAMAP – a new spectrometer system for column-averaged methane and carbon dioxide observations from aircraft: retrieval algorithm and first inversions for point source emission rates, Atmospheric Measurement Techniques, 4, 1735–1758, https://doi.org/10.5194/amt-4-1735-2011, 2011.

Krings, T., Gerilowski, K., Buchwitz, M., Hartmann, J., Sachs, T., Erzinger, J., Burrows, J. P., and Bovensmann, H.: Quantification of methane emission rates from coal mine ventilation shafts using airborne remote sensing data, Atmospheric Measurement Techniques, 6, 151–166, https://doi.org/10.5194/amt-6-151-2013, 2013.

Maasakkers, J. D., Varon, D. J., Elfarsdóttir, A., McKeever, J., Jervis, D., Mahapatra, G., Pandey, S., Lorente, A., Borsdorff, T., Foorthuis, L. R., Schuit, B. J., Tol, P., van Kempen, T. A., van Hees, R., and Aben, I.: Using satellites to uncover large methane emissions from landfills, Science Advances, 8, eabn9683, https://doi.org/10.1126/sciadv.abn9683, 2022.

McLinden, C. A., Fioletov, V., Krotkov, N. A., Li, C., Boersma, K. F., and Adams, C.: A Decade of Change in NO2 and SO2 over the Canadian Oil Sands as Seen from Space, Environ. Sci. Technol., 50, 331–337, https://doi.org/10.1021/acs.est.5b04985, 2016.

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., and Crisp, D.: Quantifying CO2 Emissions From Individual Power Plants From Space, Geophysical Research Letters, 44, 10,045-10,053, https://doi.org/10.1002/2017GL074702, 2017.

Noppen, L., Clarisse, L., Tack, F., Ruhtz, T., Merlaud, A., Van Damme, M., Van Roozendael, M., Schuettemeyer, D., and Coheur, P.: Constraining industrial ammonia emissions using hyperspectral infrared imaging, Remote Sensing of Environment, 291, 113559, https://doi.org/10.1016/j.rse.2023.113559, 2023.

Deleted: ¶

Pommier, M., McLinden, C. A., and Deeter, M.: Relative changes in CO emissions over megacities based on observations from space, Geophysical Research Letters, 40, 3766–3771, https://doi.org/10.1002/grl.50704, 2013.

Ramier, A., Girard, M., Jervis, M., MacLean, J.P, Marshall, D., McKeever J., Strupler, M., Tarrant, E., and Young, D.: High-Resolution Methane Detection with the GHGSat Constellation, in: GLOC 2023 conference proceedings, 75002, 2023.

Rezvanbehbahani, S., Stearns, L. A., Keramati, R., Shankar, S., and van der Veen, C. J.: Significant contribution of small icebergs to the freshwater budget in Greenland fjords, Communications Earth & Environment, 1, 1–7, https://doi.org/10.1038/s43247-020-00032-3, 2020.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597 [cs], 2015.

Valin, L. C., Russell, A. R., and Cohen, R. C.: Variations of OH radical in an urban plume inferred from NO2 column measurements, Geophysical Research Letters, 40, 1856–1860, https://doi.org/10.1002/grl.50267, 2013.

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.

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., Jacob, D. J., Jervis, D., and McKeever, J.: Quantifying Time-Averaged Methane Emissions from Individual Coal Mine Vents with GHGSat-D Satellite Observations, Environ. Sci. Technol., 54, 10246–10253, https://doi.org/10.1021/acs.est.0c01213, 2020.

Formatted: Line spacing: 1.5 lines