Automated detection and monitoring of methane super-emitters using satellite data

Berend J. Schuit[1,2], Joannes D. Maasakkers[1], Pieter Bijl[1], Gourav Mahapatra[1], Anne-Wil van den Berg[1,†], Sudhanshu Pandey[1,‡], Alba Lorente[1], Tobias Borsdorff[1], Sander Houweling[1,3], Daniel J. Varon[2,4], Jason McKeever[2], Dylan Jervis[2], Marianne Girard[2], Itziar Irakulis-Loitxate[5,6], Javier Gorroño[5], Luis Guanter[5,7], Daniel H. Cusworth[8,9], and Ilse Aben[1,3]

[1]SRON Netherlands Institute for Space Research, Leiden, The Netherlands
[2]GHGSat Inc, Montréal, Canada
[3]Department of Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[4]Harvard University, Cambridge, MA, US
[5]Research Institute of Water and Environmental Engineering (IIAMA), Universitat Politècnica de València (UPV), Valencia, Spain
[6]International Methane Emission Observatory, United Nations Environment Program, Paris, France
[7]Environmental Defense Fund, Amsterdam, The Netherlands
[8]Carbon Mapper, Pasadena, CA, US
[9]University of Arizona, Tucson, AZ, US
[†]Now at: Department of Meteorology and Air Quality, Wageningen University, Wageningen, The Netherlands
[‡]Now at: Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, US

Correspondence: Berend J. Schuit (B.J.Schuit@sron.nl)

Abstract.

A reduction in anthropogenic methane emissions is vital to limit near-term global warming. A small number of so-called super-emitters is responsible for a disproportionally large fraction of total methane emissions. Since late 2017, the TROPOspheric Monitoring Instrument (TROPOMI) has been in orbit providing daily global coverage of methane mixing ratios at a resolution of up to 7x5.5 km^2, enabling the detection of these super-emitters. However, TROPOMI produces millions of observations each day, which together with the complexity of the methane data, makes manual inspection infeasible. We have therefore designed a two-step machine learning approach using a Convolutional Neural Network to detect plume-like structures in the methane data and subsequently apply a Support Vector Classifier to distinguish emission plumes from retrieval artefacts. The models are trained on pre-2021 data, and subsequently applied to all 2021 observations. We detect 2974 plumes in 2021 with a mean estimated source rate of 44 t h^{-1} and 5-95th percentile range of 8-122 t h^{-1}. These emissions originate from 94 persistent emission clusters and hundreds of transient sources. Based on bottom-up emission inventories, we find that most detected plumes are related to urban areas / landfills (35%), followed by plumes from gas infrastructure (24%), oil infrastructure (21%) and coal mines (20%). For twelve (clusters of) TROPOMI detections, we "tip-and-cue" targeted observations and analysis of high-resolution satellite instruments to identify the exact sources responsible for these plumes. Using high-resolution observations from GHGSat, PRISMA and Sentinel-2, we detect and analyze both persistent and transient facility-level emissions underlying the TROPOMI detections. We find emissions from landfills and fossil fuel exploitation facilities, for the latter we find up to ten

facilities contributing to one TROPOMI detection. Our automated TROPOMI-based monitoring system in combination with high-resolution satellite data allows for the detection, precise identification and monitoring of these methane super-emitters, which is essential for mitigating their emissions.

1 Introduction

Anthropogenic methane emissions have caused at least 25% of human-induced global warming (Ocko et al., 2018; IPCC, 2021). Methane's atmospheric concentration has increased by a factor of 2.5 since the pre-industrialized era (Szopa et al., 2021) and the rate of increase has accelerated in recent years (NOAA, 2022). Due to its relatively short atmospheric lifetime and large global warming potential (81 times that of CO_2 over a timespan of 20 years (IPCC, 2021)), methane has an important role in the rate of climate warming (Nisbet et al., 2020; Ocko et al., 2021; Szopa et al., 2021). Reducing global methane emissions is therefore vital to achieve the goals set out in the 2015 Paris climate agreement (Nisbet et al., 2020). Since November 2021, over 125 countries signed the Global Methane Pledge (European Commission, 2021; CCAC, 2022), committing them to reduce their methane emissions by 30% in 2030 compared to 2020 levels, this could avoid 0.2°C of global mean warming by 2050 (CCAC, 2022; UNEP and CCAC, 2021). In order to reduce global methane emissions fast and effectively during this decade, it is paramount to identify the largest anthropogenic sources of methane and mitigate those. We therefore propose an automated detection and monitoring system using satellite data with machine learning models to detect methane super-emitters.

The dominant anthropogenic methane emission sources are agriculture (livestock and rice cultivation), oil and gas exploitation, waste management and coal mining; the exact locations and magnitudes of emissions are still uncertain (Saunois et al., 2020). Large fractions of methane emissions in various sectors could be mitigated using existing technology, about a quarter of those at no net cost (Nisbet et al., 2020; Ocko et al., 2021; Lauvaux et al., 2022). Moreover, a small number of emitters is responsible for a disproportionally large fraction of total anthropogenic emissions (Zavala-Araiza et al., 2015; Jacob et al., 2016). These concentrated point sources are often referred to as 'super-emitters' and are difficult to account for in global bottom-up inventories (Zavala-Araiza et al., 2015) as they are often caused by severe malfunctioning or abnormal operating conditions, e.g. dysfunctional natural gas flaring systems (Irakulis-Loitxate et al., 2022a, b; Plant et al., 2022). Super-emitters are not limited to oil and gas production but also occur in the coal mining and waste sectors (Cusworth et al., 2020; Sadavarte et al., 2021; Maasakkers et al., 2022b). Detection, localization, and global monitoring of these methane super-emitters provides a large opportunity to reduce emissions (UNEP and CCAC, 2021; Parry et al., 2022).

One way to obtain more insight in where super-emitters occur is to perform measurements on the ground, with drones, or aircraft campaigns. Several regions with known frequent and large methane emissions have been mapped in detail with aircraft campaigns (e.g. Frankenberg et al., 2016; Duren et al., 2019; Yu et al., 2022; Plant et al., 2022). While ground-based or airborne measurements are limited in spatial and temporal coverage, satellite observations have the potential for global monitoring of methane point-sources with frequent revisits (Jacob et al., 2016; Cusworth et al., 2019; Jacob et al., 2022). The TROPOspheric Monitoring Instrument (TROPOMI) (Veefkind et al., 2012) was launched in 2017 and observes atmospheric methane column-averaged mixing ratios with a pixel size down to 7 km x 5.5 km and daily global coverage (Hu et al., 2018; Lorente et al., 2021), resulting in a point-source detection limit down to ∼5 t h^{-1} under favorable conditions (Jacob et al., 2016). TROPOMI data have been used to quantify global (Qu et al., 2021) and country-level (Chen et al., 2022) distributions of methane emissions as well

as large area sources, such as oil and gas basins like the Permian Basin (Zhang et al., 2020; de Gouw et al., 2020; Schneising et al., 2020; Shen et al., 2022). Lauvaux et al. (2022) performed a study into oil and gas related methane super-emitters using TROPOMI data. Several individual super-emitters have been studied in detail using TROPOMI CH_4 data; e.g. natural gas well blowouts (Pandey et al., 2019; Cusworth et al., 2021; Maasakkers et al., 2022a), and various persistent sources (Varon et al., 2019; Sadavarte et al., 2021; Tu et al., 2022a, b).

Given the intermediate spatial resolution of TROPOMI, it can only be used to pinpoint the sources of emissions from the largest and most isolated point sources. For more challenging sources, high-resolution instruments are more suitable to detect and identify the exact location of super-emitters. So far, the only in-orbit satellite instrument specifically designed to do so are the GHGSat instruments that have a spatial resolution of ~25 m x ~25 m over targeted ~15 km x ~10 km scenes (Varon et al., 2019; Jervis et al., 2021; Ramier et al., 2020; MacLean et al., 2021). More recently it was shown that several Earth surface imagers with spectral sensitivity in the short-wave infrared (SWIR), although not designed for this purpose, can detect signals from methane super-emitters under favorable conditions. As such, the hyperspectral PRISMA instrument (Cogliati et al., 2021) was used to retrieve atmospheric methane plumes related to fossil fuel exploitation using targeted scenes of 30 km x 30 km at a spatial resolution of 30 m x 30 m (Guanter et al., 2021; Cusworth et al., 2021). Varon et al. (2021) demonstrated that the Multi-Spectral Instrument, a band imaging instrument onboard the Sentinel-2 satellite (Drusch et al., 2012), is capable of retrieving large methane plumes with a pixel resolution of 20 m x 20 m for continuous, 290 km wide swaths over favorable terrain. TROPOMI's daily global coverage is particularly well suited to guide observations of these high-resolution instruments, that often have limited viewing domains, to identify large sources of methane at facility level (Irakulis-Loitxate et al., 2022b; Maasakkers et al., 2022b).

TROPOMI has collected over 5 years of methane data, which include numerous methane emission plume signals that cannot feasibly be identified manually. To monitor the growing volume of data for super-emitters, an automated approach is needed. The vast amount of data provides an opportunity for machine learning techniques that require a substantial amount of representative training data. Applications of machine learning in satellite remote sensing have mostly focused on studying the Earth's surface, but also include monitoring anomalous atmospheric conditions to identify plume signatures in large datasets (Valade et al., 2019; Finch et al., 2022). Detecting methane plumes in TROPOMI data is particularly challenging because not every retrieved methane enhancement is a genuine methane emission plume as retrieval artefacts and natural variability can appear like methane plumes. We therefore use a two-step machine learning method to identify methane emission plumes. We first use a Convolutional Neural Network (CNN) to detect plume-like structures in TROPOMI XCH_4 data and then use a Support Vector Classifier (SVC) to evaluate these potential plumes using additional information to distinguish real plumes from artefacts. We train the machine learning models on verified TROPOMI methane plumes from 2018-2020, and then apply the trained models to 2021 data. Based on the 2021 detections, we use observations of three high-resolution satellite instruments (GHGSat-Cx, PRISMA and Sentinel-2) to determine the origin of the emissions down to facility level. The combination of the

automated global monitoring based on TROPOMI with the high resolution of targeted instruments allows the detection and characterization of super-emitters around the world.

2 Data & Methods

We use two machine learning models in sequence to detect plumes in the TROPOMI methane data. First we apply a Convolutional Neural Network (CNN) to detect plume-like structures in TROPOMI methane atmospheric mixing ratio data, then we use additional atmospheric parameters and supporting data to further distinguish between genuine methane plumes and retrieval artefacts using a second machine learning model.

Figure 1 illustrates the full machine-learning pipeline and training process. Section 2.1 describes the pre-processing step to generate scenes used by the CNN and the feature engineering algorithms. Section 2.2 describes the training process of the CNN (Figure 1 *'CNN Training'*). Section 2.3 describes the feature engineering algorithms, which are used to generate feature vectors for each TROPOMI scene. The SVC uses those feature vectors during its training process (Figure 1 *'SVC Training'*), covered in Section 2.4. Then, we apply the full, trained, machine learning pipeline to 2021 TROPOMI observations the models have not been trained on (Figure 1 *'CNN+SVC Application'*). Based on the resulting TROPOMI detections, we perform further analysis (Section 2.5) and use (targeted) high-resolution methane observations (Section 2.6) to pinpoint the responsible sources for 12 of those detections.
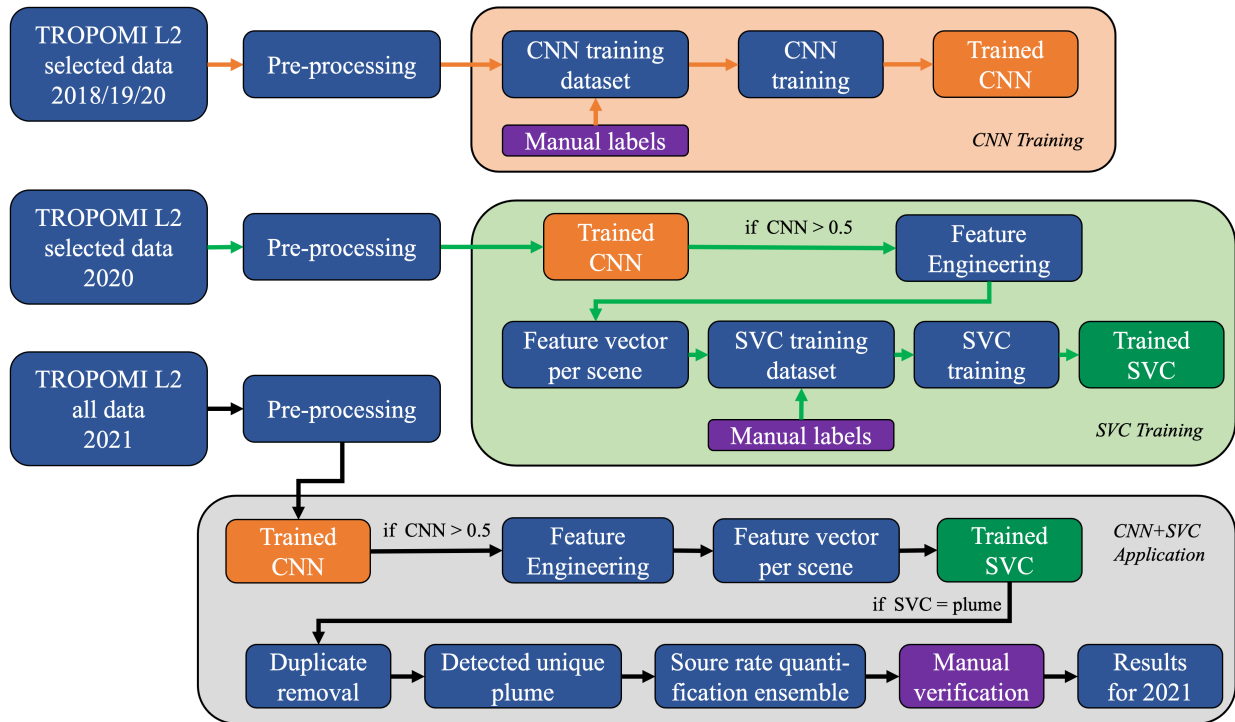
Figure 1. Flowchart showing the employed methodology framework. It consists of three phases, where each next phase uses the output (the trained model) of the previous phase as indicated with matching colors (orange and green). We use the TROPOMI XCH_4 Level 2 scientific data product version 18_17. Pre-processing is equivalent for each of the three phases, and consists of filtering, de-striping of the XCH_4 channel and splitting up the data into 32x32 scenes. The output of the pre-processing is a dataset of [N, M, 32, 32] where N is the number of scenes, M is the number of channels (fields of data used later on, for example the methane concentrations) and 32x32 gives the (pixel) dimensions of the scene. The CNN exclusively uses the XCH_4 channel, both during training and when the trained CNN is used for classification. In the Feature Engineering step of the second and third phase, a feature vector of shape [1, 41] is computed which corresponds to a single scene [1, M, 32, 32]. The SVC exclusively uses the feature vectors, both during training and when the trained SVC is used for classification. Manual verification steps are shown in purple. In the application phase, there is one manual step, which is the verification of detected plumes to make sure the output of the pipeline is correct.

2.1 TROPOMI

We use data over land from the TROPOMI XCH_4 Level 2 scientific data product version 18_17. This product version is con-
sistent with operational version 02.03.01 (Lorente et al., 2021; Hasekamp et al., 2022), but re-processed for the full timespan
of the mission resulting in a homogeneous data product (SRON CH4 L2 team, 2022). We use albedo-bias corrected data with:
QA≥0.4, methane precision < 10 ppb, SWIR aerosol optical depth < 0.13, NIR aerosol optical depth < 0.30, SWIR surface
albedo > 0.02, mixed albedo (2.4 * NIR surface albedo - 1.13 * SWIR surface albedo) < 0.95, and SWIR cloud fraction <
0.02. The loosened filtering compared to the recommended QA=1 filter provides more coverage but also retains more biased
retrievals, especially at the borders of clouds or along coasts. The methane data is de-striped following the approach introduced
by Borsdorff et al. (2018).

To train a machine learning model to recognise methane plumes in TROPOMI data, we created a dataset of scenes consisting
of 32x32 pixels, both with and without methane plumes. 32x32 pixels correspond roughly to an area of ∼176x232 km^2 at nadir,
up to ∼176x448 km^2 for larger viewing angles. For scenes with plumes, we use data over 60 persistently emitting locations
identified using long-term wind-rotated averages (Maasakkers et al., 2022b). By manual inspection, we compile a dataset of
828 positive scenes from 2018-2020 with plumes, of which 195 originate from coal mines, 203 from landfills/urban areas,
and 430 from oil&gas infrastructure (Figure 1, *'CNN Training'*). An example scene is shown in Figure 2a, including the local
windfield at the time of observation.
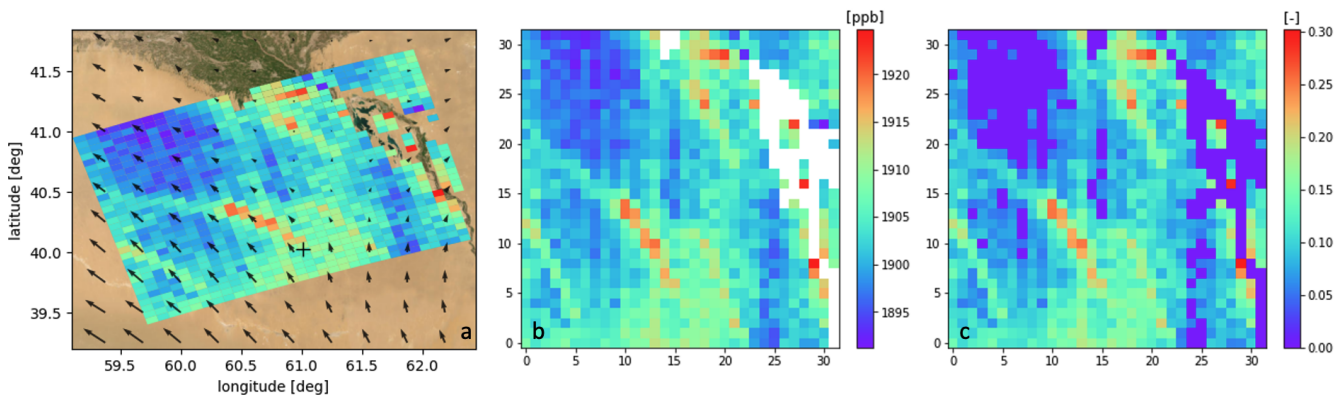
Figure 2. Atmospheric methane mixing ratios of a 32x32 pixel scene containing a methane plume originating from a known persistent source
(indicated by the +) as observed by TROPOMI on 2021-12-05 at 08:47 UTC (not included in the training data). (a) Mercator projection of the
scene over ESRI World Imagery (Esri, Maxar, Earthstar Geographics, and the GIS User Community, 2022), arrows show the local GEOS-FP
10m windfield (Molod et al., 2012). (b) 32x32 pixel scene in along-orbit vs across-orbit direction, indicating filtered pixels. (c) the same
scene after pre-processing as used by the CNN.

A set of (negative) scenes without an emission signal was obtained through manual inspection of six full orbits in different
sections of the orbital repeat cycle, covering a diverse set of surfaces and (meteorological) conditions. We obtain 32x32 pixel

scenes using a moving window algorithm with 50% overlap, resulting in a dataset of 2242 scenes without a plume signal. The moving window algorithm later ensures that if a plume is cut in half in a particular scene, it will be in the center of the adjacent -and partially overlapping- scene. Scenes with <20% valid XCH_4 pixels are discarded. This processing is applied to

125   full orbits, the scenes with plumes originating from known locations were processed to match the same format. Combined, the dataset contains 3070 scenes used for training, the difference in the number of positive (828) and negative (2242) scenes is corrected for later on using class weights. For each scene, we store 46 other channels of supporting information from the same TROPOMI Level 2 methane dataproduct, including co-retrieved atmospheric properties, meteorological parameters and geometric properties. These channels are used in later steps of the machine learning pipeline.

130

In order to correct for differences in local background concentrations (e.g. due to difference in latitude or surface altitude), each scene is normalized from 0 to 1. Values below the mean methane concentration of the scene minus one standard deviation are set to 0. Values above the mean plus 100 ppb minus one standard deviation are set to 1, values in between are linearly distributed. Filtered pixels are set to 0. This pre-processing preserves the information of plume-like enhancements above the

135   local background. Examples of the normalization input and output are shown in respectively Figure 2b and 2c.

2.2 Convolutional Neural Network (CNN)

We use a Convolutional Neural Network (LeCun et al., 2010) to detect methane plumes in the TROPOMI methane data, split into 32x32 pixel scenes. Convolutional Neural Networks (CNN) are a type of machine learning model commonly applied in image recognition and object detection problems (Cheng et al., 2020). A CNN consists of multiple layers, where information moves from an input image, through the layers of the CNN at an increasingly abstract coarse resolution, to the output; the classification of the image. To condense the information of the image to coarser resolution, the CNN uses "convolutional blocks" that consist of two or more convolutional layers, followed by a (max-)pooling layer. A convolutional layer produces "feature maps" that indicate where certain features (e.g. curves, edges or more abstract features) are detected within the image. These feature maps are obtained by convoluting the input image with a convolutional kernel, a small matrix with weights that are optimised during training to best detect features relevant to the particular classification problem. The resulting feature maps (one for each kernel applied) are then the input for the next layer (LeCun et al., 2010; Cheng et al., 2020). A max-pooling layer scans the previous layer with a 2x2 kernel and returns the maximum value, thereby creating a feature map at half the resolution, focused on dominant features (LeCun et al., 2010). After the last convolutional block, the resulting feature maps are flattened and interpreted by one or more fully-connected "dense" layers, consisting of neurons, between which the connections have trainable weights. This part of the network aggregates the information into a single output value. During training, the trainable weights, in the convolutional kernels and dense layers, are optimized to best perform the classification task based on the training dataset (LeCun et al., 2010; Cheng et al., 2020). The trained CNN can then be used to classify new images, in this case as 'plume' versus 'no-plume'.

The main advantages of the CNN compared to regular neural networks or other machine learning models are that: (1) the CNN is capable of better retaining spatial information which is lost in fully-connected networks or machine learning models like decision trees or support vector machines (Selvaraju et al., 2020); (2) the training of a CNN can be done with image level labels (plume or no plume), there is no need to indicate where within the image the feature of interest is located as the CNN learns to localize these features during training; (3) the same convolutional kernels are convoluted with the entire image, which is more computationally efficient compared to fully-connected networks (LeCun et al., 2010; Cheng et al., 2020); (4) the model is rotational and translational invariant when properly trained (LeCun et al., 2010).

This last model property is essential for the automated detection of plumes as a plume can be located anywhere within a scene and the wind can be in any direction. The CNN's output is a prediction between 0 and 1 indicating the confidence of the model about the presence of a plume-like structure. Scenes with prediction scores >0.5 are classified as plumes. Although we use this output for binary classification, the value holds additional information regarding the confidence of the CNN (i.e. 0.6 vs. 0.98), which we use for the second model.

We first select a high-level architecture for the CNN with standard hyperparameters, which we later optimize. Hyperparameters are model settings, separate from the trainable weights, such as the number of convolutional layers and kernel sizes,

170 but also parameters that influence the training process, such as the learning rate. As high-level architecture we selected two convolutional blocks followed by two fully-connected layers and an output neuron (Figure 3). We found that deeper networks (e.g. Resnet (He et al., 2016) or VGG-16 (Simonyan and Zisserman, 2014)) did not yield an improvement in performance for this problem with relatively low resolution, small size (32x32 pixel) scenes.
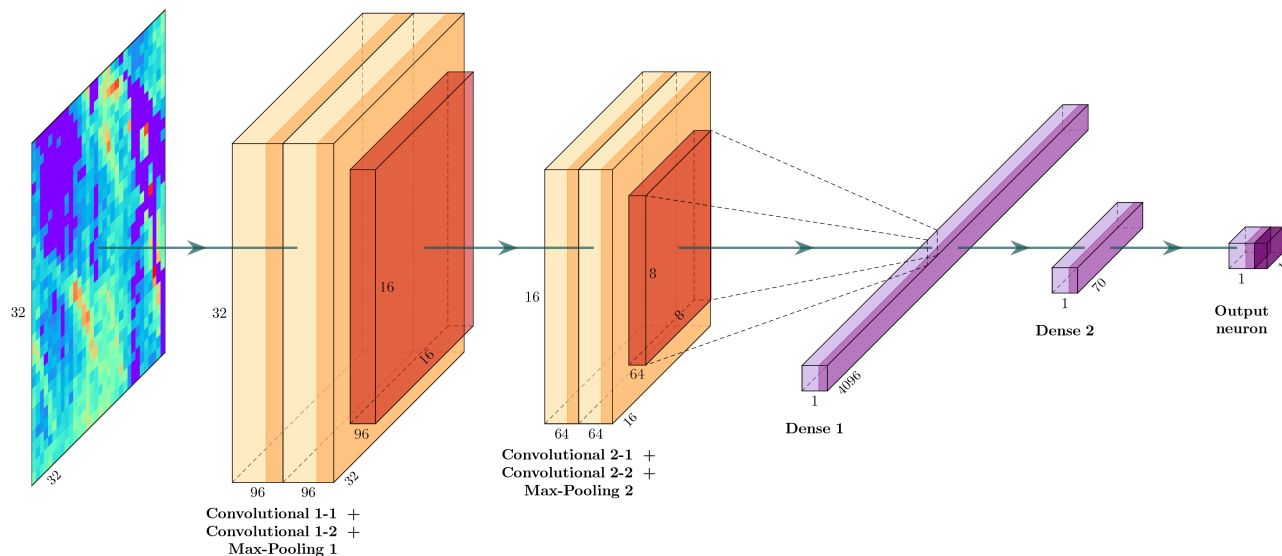
Figure 3. A schematic overview of the Convolutional Neural Network with a pre-processed 32x32 pixel TROPOMI methane scene (left) as input (Figure 2c). The CNN consists of two convolutional blocks (each with two convolutional layers followed by a max-pooling layer) followed by two dense (or fully-connected) layers and an output neuron with sigmoid activation. Numerical values show input dimensions, layer dimensions and the number of feature maps in the convolutional and max-pooling layers. Visualization generated using PlotNeuralNet (Iqbal, 2018).

Our training dataset mostly contains clear positives (828) and clear negatives (2242) to effectively learn distinguishing features. Our dataset has many more negative than positive scenes. When training our CNN however, we want both categories of training samples (classes) to have equal impact to obtain optimal performance (Johnson and Khoshgoftaar, 2019). This balancing can be achieved by applying class weights during training (Johnson and Khoshgoftaar, 2019), giving positive scenes more weight. We set the class weight parameter to the inverse of the ratio between the number of positives and negatives. We randomly split the data into a training (80%) and test (20%) set. 20% of the training scenes are then used as validation subset (Table A1). The validation dataset is used to infer whether there is a generalized performance increase during the training of the CNN to prevent overfitting. We then augment the data in order to obtain larger training and test sets; all scenes are rotated 90° thrice and flipped, enlarging the datasets by a factor 8.

We use the training dataset of 19,648 (augmented) scenes (Table A1) to train the CNN (Figure 1, *'CNN Training'*). The CNN was designed and trained using the machine learning framework Keras (Chollet et al. (2015), Chollet (2021)), at first using the default values for the hyperparameters. The model is trained for a maximum of 100 epochs (iterations of the training process). During training we optimize the validation loss, which measures the error made on the subset of the training data not used in that epoch. We use binary cross-entropy as the loss function and ADAM (an improved version of stochastic gradient descend algorithm, Kingma and Ba (2014)) as the optimizer. We use a 0.4 dropout layer (randomly disabling 40% of the neurons) in the first fully-connected layer during training to prevent overfitting and make the model more robust (Srivastava et al., 2014). The activation function modifies the output before it is passed to the next layer; we apply the ReLU (rectified linear unit, outputs zero when the input is negative and otherwise outputs the input value) activation function in all layers except for the final layer where we apply sigmoid ($\frac{1}{1+e^{-x}}$), which normalizes the output. To force the model to focus more on plume-like signatures during training, the loss weight of plume scenes is set to double that of negatives scenes. Training is halted after the validation loss does not improve for several epochs, the best model weights found up to that point are used. After training, the model performance is inferred by classifying the labeled test dataset.

After training this initial "default" model version, the hyperparameters were further optimized using the KerasTuner optimization framework (O'Malley et al., 2019) and Hyperband (Li et al., 2018). With these methods, we perform a grid-search to find the best hyperparameters for our particular problem. The optimal hyperparameters depend on the size of the training dataset, architecture of the CNN, number of classes and problem type. The search space for the optimization was defined using insights from the initial training, theoretical foundation, and design constraints. We inspected the hyperparameters of the top 10 performing set-ups and selected the optimal hyperparameters by combining the results of this optimization with expert judgement on this particular problem. Figure 3 shows a schematic overview of the CNN with optimized hyperparameters.

We evaluate the performance of the CNN using performance evaluation metrics (Equation 1) calculated from the number of true positives and negatives (TP, TN), and false positives and negatives (FP, FN) (Johnson and Khoshgoftaar, 2019). The Cohen's Kappa score (Cohen, 1960) is a weighted accuracy which takes into account class imbalances and chance agreement.

The Recall indicates which fraction of plumes present in the testset are correctly identified and the precision indicates which fraction of scenes identified as plumes are actually a plume, the F1-score incorporates both into a single metric.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN} \qquad F_1 = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$(1)$$

To test the influence of the split of the training and test datasets (which can be an issue for datasets of limited size), the training of the model with optimized hyper-parameters was repeated 50 times with different splits. We found that model performance is relatively insensitive to different training splits with kappa = 0.943 ± 0.012, recall = 0.956 ± 0.014 and F1 = 0.958 ± 0.009 (standard deviations). We further found that small changes in the hyperparameters have even lower effect compared to different training splits. The consistent performance on the corresponding test datasets shows the model is robust and well generalized. We focus on recall over precision, because the key focus is to have as few potential plumes as possible go undetected. We selected the model which scored best on kappa, second best on F1-score and third best on recall. The performance metrics of the selected CNN are shown in Figure 4a. Manual inspection of the misclassified scenes (30 false negatives and 37 false positives, out of 4912 augmented test scenes) indicates these are borderline cases with difficult to discern morphological structures that are even challenging to a human expert.
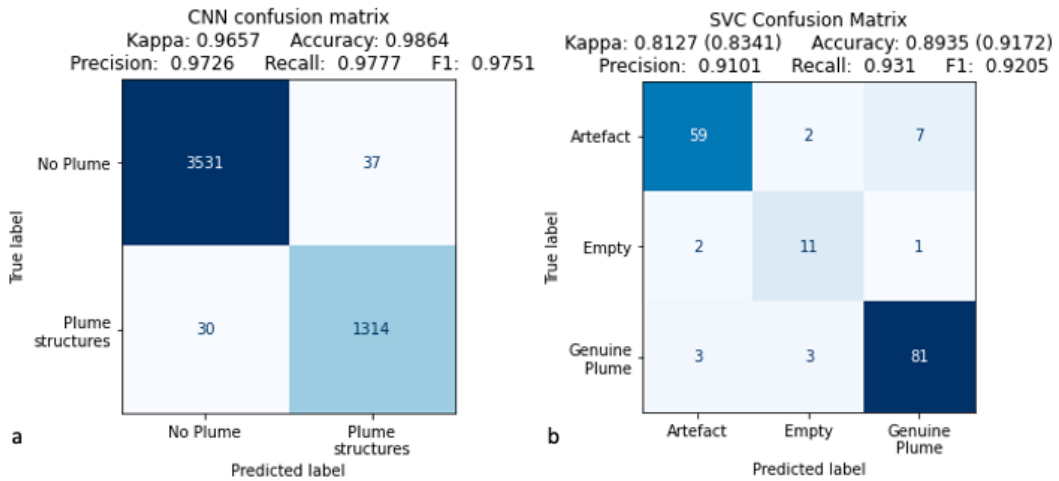
Figure 4. Confusion Matrices showing the performance of the CNN (a) and the SVC (b) on their corresponding test datasets. The performance metrics are defined in Equation 1, the values between brackets for the SVC show the performance when the problem is considered as a binary problem; i.e. when 'Artefact' and 'Empty' are combined as 'No Plume'.

The trained CNN is applied to all 2020 data. Processing of the 5193 orbits of 2020 resulted in 752,890 scenes (only taking into account scenes with >20% valid pixels), of which 25,626 scenes (3.4%) are identified as containing plume-like morpho-

logical structures by the CNN. This number does include artefacts and duplicates due to the moving window algorithm, these are filtered later on. We use a subset of these scenes to train the second step of the machine learning pipeline (section 2.4).

2.3 Feature Engineering

Due to the difficult nature of methane retrievals, not every plume-like morphological structure in the XCH_4 field is an actual methane plume. Different types of surface variability and atmospheric or meteorological conditions are known to affect the retrieval (Lorente et al., 2021); if there is a strong correlation between the methane enhancement and retrieval parameters, e.g. the surface albedo or the aerosol scattering coefficient, the retrieved methane enhancement might be caused by the albedo or aerosol variation and could therefore be a retrieval artefact. Other common examples of artefacts are those on the borders of clouds and coastlines or when the direction of the enhanced structure is not in agreement with the windfield but with surface structures instead.

In order to automate the necessary further inspection we compute numerical values for several "features" of potential plumes through feature engineering. Feature engineering is a commonly applied approach in machine learning problems and is especially helpful when limited amounts of labeled training data are available. These features are a representation of information a human expert would use to inspect the potential plumes in order to determine whether a scene contains a genuine plume or an artefact. We construct feature vectors consisting of features based on the corresponding scene. These vectors are then used to train the second model of the machine learning pipeline, the SVC (Figure 1 'SVC Training'). An overview of all developed features is presented in Table C1.

Fundamental to many of those features is "masking" the plume in order to isolate the plume pixels from the background. For this purpose, we use information about which part of the scene has triggered the CNN detection. For this we use the Class Activation Map (CAM) to visualize the localized activations of a CNN corresponding to a certain class on which it was trained (Zhou et al., 2015). We apply Grad-CAM (Selvaraju et al., 2020), which allows the computation of the CAM for our CNN that includes fully-connected layers. In our binary classification problem, the CAM visualizes which regions of the deepest (coarsest) feature maps (Max-Pooling 2 [8x8] in Figure 3) contribute strongest to an activation of the plume class (output > 0.5) for a given input image. This spatial activation is calculated using the gradients (Selvaraju et al., 2020) between the 64 feature maps (each of 8x8 resolution, resulting in a 64x8x8 array, Figure 3) of the deepest max-pooling layer (Max-Pooling 2) and the first fully-connected, or dense layer (Dense 1 in Figure 3). In order to obtain a CAM of sufficient resolution we limit the depth of the CNN to two convolutional blocks. The CAM is upsampled to match the input resolution (Selvaraju et al., 2020). Figure 5b shows that the CAM correctly identified the plume-like structure in the XCH4 scene in Figure 5a, disregarding noisy high-enhancement pixels elsewhere in the scene.
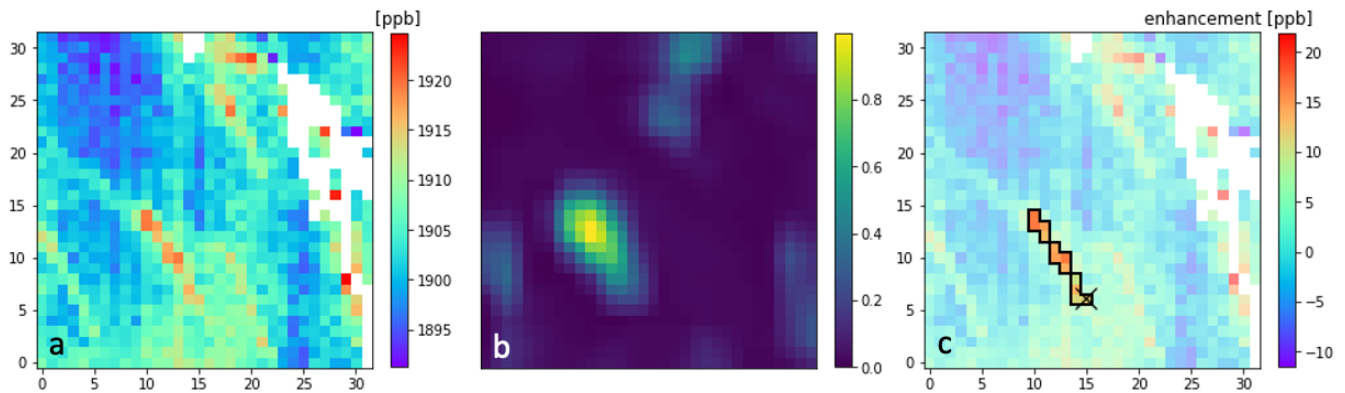
Figure 5. Several Feature Engineering results computed for the 32x32 scene from Figure 2. (a) The atmospheric methane mixing ratio XCH_4, (b) Class Activation Map which highlights the areas identified by the CNN as plume-like structures based on Figure 2c, the pre-processed scene, (c) methane enhancements relative to the local background with the black line indicating the high-confidence plume mask. The pixel with the 'x' is identified as the pixel most likely to contain the source location based on the plume mask and local windfield (shown in Figure 2a).

In addition to using the CAM to analyze CNN performance, we also use it to generate a binary plume mask. We multiply the CAM with the enhancement above the mean XCH_4 value minus the scene's standard deviation. The output is a map highlighting pixels with high methane enhancements that are identified as part of the plume by the CNN; we identify the pixel with the maximum value as starting point for the plume mask. To compute a "high confidence" plume mask, we start from the corresponding pixel in the xch4 image and dilate outwards (including diagonally). We only add adjacent pixels with enhancements of 1.8 standard deviations above the mean (Figure 5c). We repeat this process with a lower threshold of 0.8 standard deviations to also obtain a "low confidence mask"; both thresholds were established empirically. This approach ensures that noise in other parts of the image is excluded from the plume mask. A plume mask can consist of any number of pixels depending on the scene. The minimum is 1 pixel, but this is rare, the average is around 20 pixels. Several statistics of the (potential) plume can be computed using these masks with supporting data.

One of the major indicators of an artefact is a strong correlation with one or more retrieval parameters. If an enhancement in the XCH_4 field is caused by a surface (albedo) feature or by (enhanced) scattering in the atmosphere which is represented by the aerosol optical thickness, we expect their spatial patterns to be similar. Therefore, we calculate the correlation between XCH_4, and the surface albedo (SWIR), aerosol optical thickness, χ^2 (an indicator for retrieval fit quality) and surface pressure across the plume mask. We calculate these correlations for the high and low confidence plume masks as well as 1- and 2-times dilated versions of the low confidence mask, and the entire scene. We account for pixels outside the plume mask as we would expect a strong correlation around the enhancement if it is an artefact. The correlations over the entire scene reflect large scale patterns that do not necessarily imply artefacts.

275   Another major indicator for artefacts is a mismatch between the direction of the plume and the direction of the local 10-m wind field from the ERA5 reanalysis (Hersbach et al., 2020) included in the TROPOMI Level 2 data product (Hasekamp et al., 2022). By applying a Principle Component Analysis (PCA) we compute the two main axes of the pixels in the high confidence plume mask, after re-projecting the pixel centers to meter space and weighting them by their enhancement relative to the background. We use the ratio of the variances described by the axes as a measure of the plume's elongation. For elongated

280   plumes (e.g. Figure 5), the primary axis' variance is much larger than the variance projected to the secondary axis, while for less elongated "blob"-like plumes this ratio is small. Furthermore, we compare the angle of the primary axis of the potential plume to the angle of the wind direction (averaged across the plume mask), the smaller the difference, the more confidence we have in the plume following the wind. We also use the wind field to identify the pixel that most likely contains the plume's source by taking the 'most upwind' pixel within the high confidence plume mask (Figure 5c).

2.4 Support Vector Classifier (SVC)

285

A Support Vector Classifier (SVC) constructs hyperplanes as the optimal decision boundary to separate multiple classes in high-dimensional feature space. SVCs in general perform better with datasets of limited size compared to deep learning algorithms and are in general less prone to erroneous influence from outliers. We use 843 labeled scenes from 2020 classified by the CNN to contain plume-like structures as a training dataset for the Support Vector Classifier (SVC) (Figure 1, *'SVC Training'*).

290 About half of the scenes are randomly selected from within 7 geographical zones with specific types of predominant artefacts and the other half is selected randomly. Scenes are labeled as 'plume' (444 scenes), 'artefact' (341 scenes) or 'empty' (58 scenes, indicating there is not a clear plume or artefact). We have only included scenes with unambiguous labels. The fact that there are relatively few 'empty' scenes in this subset indicates the CNN performs well. We use balanced class weights to correct the imbalance in the number of training samples per class, the weights are inversely proportional to the number of

295 scenes in each class.

The dataformat we use for the SVC is a vector of 41 features, each feature vector corresponds to a 32x32 pixel scene. This feature vector includes: correlations with retrieval parameters for different plume masks, the angle between the wind and elongated direction of the plume, the elongation ratio of the plume, several intermediate outputs of the source rate estimate

300 (Section 2.5.1) and several statistical properties (all engineered features are listed in Table C1). We do not include features such as latitude and longitude or distance to a known source or known infrastructure in order to be unbiased to where a scene is located. We train the SVC to find the optimal classification boundary within this 41 dimensional space based on the 843 labeled feature vectors. Each feature is standardized by subtracting the mean, and scaling the value to the unit variance of that feature in the entire trainingset. We use a Radial Basis Function (RBF) kernel and set the regularization parameter to 1.2, this

305 hyperparameter was optimized using a simple grid search optimization. The gamma value is scaled inversely to the number of features (41 in this case) multiplied with the variance of the training dataset, as is common practice and helps to homogenize the features which have different units and ranges of values (Table C1).

We randomly split the labeled dataset into a training set (80%) and a test set (20%). Contrary to the CNN, when training an

310 SVC no validation set is used. Correctly detecting plumes is of predominant interest, therefore we combine non-plume scenes (artefact & empty classes) when evaluating the performance. We train the model 2000 times for different splits of the dataset. The distribution of these different realizations shows convergence with binary kappa = 0.78 ± 0.04 and recall = 0.88 ± 0.03 (standard deviations), indicating the model setup is not too dependent on the data split. We select a model with relatively high kappa and recall, where performance is similar between the training and test sets. Figure 4b shows the 3-class performance

315 on the test set, indicating distinguishing between plumes and artefacts is the most challenging distinction for the SVC. The binary Cohen's Kappa score of the selected model is 0.83. A Kappa score of above 0.8 is generally seen as a good classification performance. The recall is 0.93, meaning that 93% of the scenes with plumes which were present in the test set are successfully

identified, and only 7% of the plumes are missed.

To evaluate which features are important for the SVC to classify a scene, we performed a permutation importance analysis perturbing each feature 40 times (Breiman, 2001). Based on the resulting mean feature importance, the most important features are: correlation of XCH$_4$ with χ^2, the CNN score, the albedo correlation, the enhancement of the plume, the fraction of valid pixels, the angle with the local wind, and the average quality flag of plume pixels (the top 10 ranking feature importance metrics are presented in Table C1). These correspond to what is important to a human expert labeler.

2.5 Plume characterization

2.5.1 Source rate quantification

To estimate the source rates of the plumes observed by TROPOMI, we apply the Integrated Mass Enhancement (IME) method (Frankenberg et al., 2016; Varon et al., 2018). Some intermediate outputs of the IME method (such as the plume length) are used as features in the feature vector for the SVC (Section 2.4). We perform a full source rate quantification, including uncertainty estimates, for the plumes that pass the machine learning pipeline and are manually verified. The IME method relates the emission rate (Q) to the observed methane enhancement in the plume (IME) and the local windfield (Varon et al., 2018):

$$Q = \frac{1}{\tau}\text{IME} = \frac{U_{\text{eff}}}{L}\text{IME} \quad ; \quad \text{IME} = \sum_{j=1}^{N} \Delta\Omega_j A_j \tag{2}$$

where $\Delta\Omega_j$ denotes the methane column mass enhancement above the local background of pixel j with footprint A_j. The local background is calculated as the median value of the scene's pixels outside the high-confidence plume mask. The IME of all N pixels in the plume is related to the source rate using the average residence time τ of methane particles in the plume, with τ given by the ratio between plume length L and the effective wind speed U_{eff}. The plume length L is approximated as $L = \sqrt{A_M}$, where A_M is the area of the plume mask (Varon et al., 2018). U_{eff} can be expressed as a function of the local (reanalysis) wind speed. Frankenberg et al. (2016) and Varon et al. (2018) developed the IME method for high resolution instruments, for which the 10-meter winds are most representative for U_{eff}. For the larger scale of TROPOMI plumes, both 10-meter (U_{10}) as well as boundary layer average (U_{PBL}) winds have been used (Varon et al., 2019; Schneising et al., 2020; Cusworth et al., 2021; Tu et al., 2022a). As the most representative wind can vary from case to case, we use the mean of quantifications using ERA5 10-meter winds (Hersbach et al., 2020), GEOS-FP 10-meter winds and GEOS-FP PBL winds (Molod et al., 2012).

We calibrate the relation between U_{eff} and these local windspeeds by quantifying 15,336 plumes simulated with the Weather Research and Forecasting model coupled with a Chemistry module, version 4.1.5 (Skamarock et al., 2019; Grell et al., 2005). The model uses 38 vertical levels and three nested domains at a horizontal resolution of up to 4x4 km^2. Physical parameteri-

sations and meteorological initial and boundary conditions are as described in Dekker et al. (2017). We release passive tracers with emission rates mostly between 10-100 t h^{-1} at various locations in West Asia, Mexico, and Argentina for June-September 2019 and 2020. We sample the plume at the TROPOMI overpass time and quantify them as described above.

Using the model wind speeds and known emission rates, we find that U_{eff}'s dependence on the PBL wind is best described by a linear relationship: $U_{\mathrm{eff}} = \alpha_1 \cdot U_{\mathrm{PBL}} + \alpha_2$, with $\alpha_1 = 0.47$ and $\alpha_2 = 0.31$ (r$^2 = 0.78$). For the dependence on U_{10}, we also find a linear relation as optimal and constrain α_2 to be non-negative: $U_{\mathrm{eff}} = \alpha_1 \cdot U_{10} + \alpha_2$, with $\alpha_1 = 0.59$ and $\alpha_2 = 0.00$ (r$^2 = 0.77$). We use the mean of the three resulting U_{eff} values to quantify emissions.

To estimate uncertainty, we create an ensemble of estimates by varying the parameters influencing the quantification. For each of the different input wind speeds: we vary the threshold for masking the plume from 1.3 to 2.3 standard deviations (step 0.1); adjust the background concentration by ± 2 times the mean XCH$_4$ uncertainty in the scene (step 0.4); vary the wind values from -50% to +50% (step 10%); and vary α_1 and α_2 for -5% to 5% (step 1%). We report the standard deviation of the resulting 43,923 member ensembles as uncertainty for each plume.

2.5.2 Removal of duplicate scenes

Due to the moving-window algorithm, each group of 16x16 pixels is seen by the ML pipeline up to four times, as a different corner of a 50% overlapping 32x32 pixel scene. This ensures plumes do not go undetected because they are cut in half and allows multiple nearby plumes to be detected in adjacent scenes, but also leads to duplicate detections. Therefore, plumes for which the generated plume mask overlaps with a plume mask from another scene are grouped into a group. For each group, the scene with the highest IME value is selected.

2.5.3 Anthropogenic source sector estimation

To assess which anthropogenic activity might underlie a detected plume, we use the estimated source location to find the local dominant source type in gridded bottom-up inventories. We exclude sectors that are unlikely to produce point-source emission signals in the TROPOMI, such as rice cultivation and livestock. We include 2019 oil, gas and coal emissions from the Updated Global Fuel Exploitation Inventory (GFEI v2) (Scarpelli et al., 2022) and 2018 landfill emissions from EDGAR V6.0 (Crippa et al., 2021). We identify the dominant source type as the source type with the largest annual flux in a $0.7°$x$0.7°$ square centered around the estimated source location. Based on known emitters, we found that using a window of this size mitigates errors in the estimated source location as well as spatial errors in the emission inventories. We do not use this approach to attribute detections to wetlands. However, we do inspect 2019 fluxes from WetCHARTs v1.3.1 ensemble (Bloom et al., 2021) to identify regions where detections might be influenced by strong wetlands fluxes, such as central Africa (Pandey et al., 2021).

2.5.4 Comparison with a previously studied super-emitter event

In order to test the automated pipeline and feature engineering algorithms, we apply it to data over a September 2019 super-emitter event in Louisiana, US (Maasakkers et al., 2022a). The model detects the emission event on multiple days, including on the first day with large emissions and significant TROPOMI coverage (September 25). The CNN score is >0.999 and the SVC classified the scene as a plume. The estimated source location of the plume is 2.2 km away from the source and our automated quantification estimate is 121 ± 46 t h^{-1}. Our estimate is in good agreement with the quantification by Maasakkers et al. (2022a) of 101 (49-127) t h^{-1} who scale a plume simulated with the WRF atmospheric transport model to match the enhancements seen in TROPOMI using a Bayesian inversion.

2.6 High spatial resolution methane satellite instruments, retrievals and source rate quantification

We use observations from three high spatial resolution instruments, GHGSat, PRISMA and Sentinel-2 to inspect the sources of the detected methane plumes. This section describes the main characteristics of these instruments.

2.6.1 GHGSat

GHGSat-Cx instruments are Fabry-Perot imaging spectrometers launched in 2020-2022 (C1-C5) building on the GHGSat-D instrument (Jervis et al., 2021). The instruments have a spatial resolution of 25 m x 25 m over targeted scenes of ~10 km x ~15 km (Ramier et al., 2020; MacLean et al., 2021). They sample the SWIR part of the spectrum between 1630 and 1675 nm at ~0.3 nm spectral resolution, retrieving methane column density with a precision of 1% of the background concentration and a theoretical detection threshold of down to ~100 kg/h at a windspeed of 3 m s^{-1} (ESA, GHGSat, 2022). During a controlled release experiment comparing methane observing capabilities of different high-resolution instruments by Sherwin et al. (2022), a plume of ~200 kg/h was successfully detected. We use data from GHGSat-C1 and C2 and estimate source rates using the IME method as described in Maasakkers et al. (2022b) for point sources with 10m wind data from GEOS-FP (Molod et al., 2012). The uncertainty in the quantification is estimated as described by Varon et al. (2019), taking into account error contributions from measurement noise, the wind speed and the IME method similar to Maasakkers et al. (2022b).

2.6.2 PRISMA

The Italian Space Agency's hyperspectral instrument PRISMA was launched in March 2019 and generates publicly available targeted hyperspectral 30 km x 30 km images at a spatial resolution of 30 m x 30 m and ~10 nm spectral resolution (Cogliati et al., 2021; Guanter et al., 2021; Cusworth et al., 2021). The revisit time can be as short as 7 days using the instrument's ±20% across-track pointing (Cogliati et al., 2021). The smallest source rate tested during a controlled release experiment (Sherwin et al., 2022) is ~2500 kg/h. The theoretical detection threshold is lower (~300-900 kg/hr for homogeneous scenes) and strongly depends on the surface type/homogeneity (Guanter et al., 2021). The PRISMA instrument is not a continuous mapper, but a data archive of past (targeted) observations is publicly available. PRISMA can also be used to target a location of interest in the future. We perform methane retrievals and IME quantifications as described in Guanter et al. (2021) using plume masking following Varon et al. (2018) and GEOS-FP 10m wind data (Molod et al., 2012).

2.6.3 Sentinel-2

The Sentinel-2 surface-imaging mission (consisting of Sentinel-2A launched in 2015 and Sentinel-2B launched in 2017) was demonstrated by Varon et al. (2021) to be capable of detecting methane super-emitter plumes under favorable conditions. Both satellites carry a MultiSpectral Instrument, with a pixel resolution of 20 m x 20 m for the B11 (~100 nm) and B12 (~200 nm) SWIR bands with sensitivity to methane. The instruments have a 290 km wide swath resulting in a global 2-5 day revisit time (Drusch et al., 2012), Sentinel-2 observes continuously (as opposed to GHGSat and PRISMA) and provides an extensive publicly available archive going back years. A methane absorption signal has to be strong in order to stand out within the

aggregated signal of the entire band, therefore only relatively large quantities of methane can be retrieved and the detection limit worsens considerably over non-homogeneous terrain. The detection limit is estimated at $\sim$1-2 t h^{-1} for homogeneous scenes (Gorroño et al., 2022), in agreement with Sherwin et al. (2022) where a $\sim$1800 kg/h emission was detected and quantified. We apply the methane retrieval and IME quantification approach from Gorroño et al. (2022), that, like Varon et al. (2021) uses a "reference day" without a plume to isolate the difference caused by methane concentration enhancement. We again use GEOS-FP 10m wind data (Molod et al., 2012).

420

3   Results

We apply the trained and optimized CNN and SVC models (Figure 1, *'CNN training'* and *'SVC training'*) in sequence on all 2021 TROPOMI XCH$_4$ data (Figure 1, *'CNN+SVC Application'*). Analysing the full year with the ML pipeline takes approximately three hours on a single core. From the 794,395 (32x32 pixel) scenes, the CNN identifies 26,444 scenes (3.3 %) that contain plume-like XCH$_4$ morphological structures. The SVC classifies 10,430 of these scenes as plumes. After duplicate removal, 4869 scenes are identified as unique. These 4869 scenes are manually inspected to assess the performance of the pipeline. We confirm 2974 scenes as confident plumes. Another 745 scenes are labeled as 'potential plumes', accepting these scenes as plumes results in a precision of 76% for the full pipeline. These potential plumes could not readily be verified as real methane plumes, but are valuable for further inspection. The remaining scenes are either labeled as artefacts or not containing a (concentrated) plume. These misclassifications can be used to further optimize the machine learning pipeline. Here, we will focus on the 2974 'confident plumes', and present the result of our high-resolution satellite instrument analysis to pinpoint the exact sources of twelve (clusters of) detections.

3.1   Overview of the confirmed detections in 2021
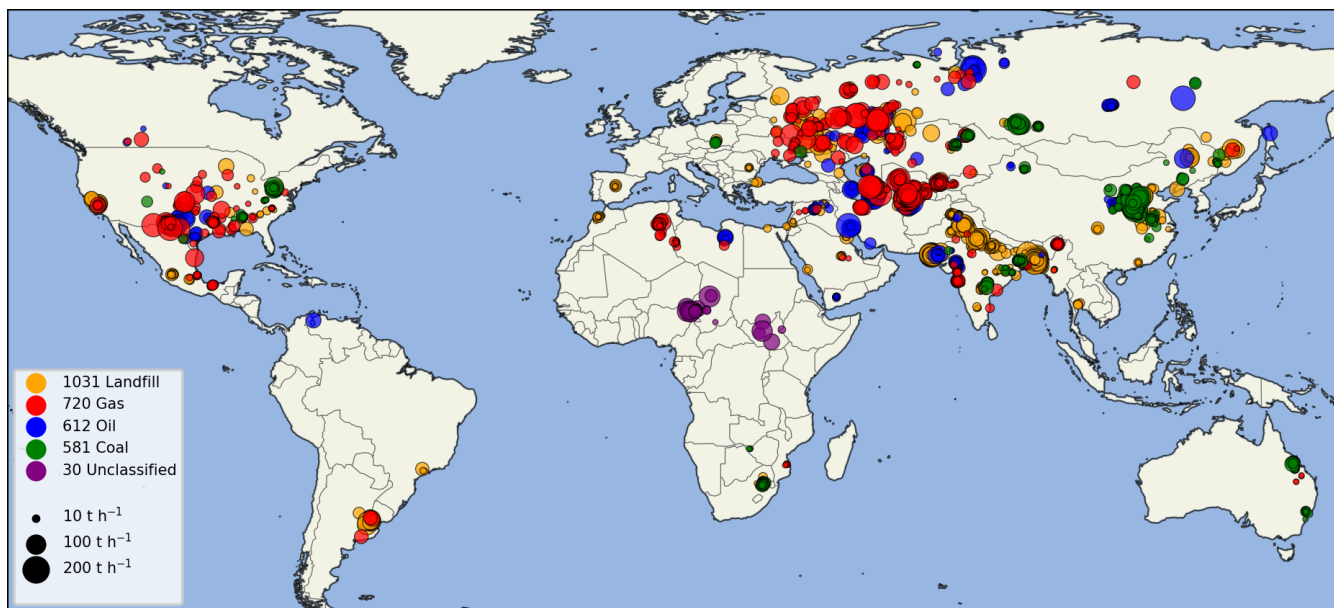
Figure 6. All 2974 confident plume detections for 2021, grouped into one of the four dominant anthropogenic source types and sized by source rate. 30 detections in central Africa are labeled as 'Unclassified'.

Figure 6 shows the spatial distribution of all 2974 detected and confirmed 2021 plumes, classified by source sector based on the three bottom-up inventories (Section 2.5). We find that 1031 plumes predominantly relate to urban areas/landfills, 720 to gas infrastructure, 612 to oil infrastructure and 581 to coal mines. As super-emitters are usually not the result of regular operations, and are therefore not well represented in bottom-up inventories, especially large transient emissions may be misattributed by this approach. Wetlands are not expected to result in point source emissions, but strongly emitting wetlands in central Africa, such as South-Sudan and the Niger delta, can produce large enhancements in TROPOMI data (Pandey et al., 2021; Shaw et al., 2022). In the absence of large anthropogenic emissions, we label plumes from these two regions as 'unclassified'. Wetlands might also contribute to detected signals in areas with large anthropogenic emissions like the Dhaka delta (Bangladesh) and the Mississippi delta (US).

There are many clear hot spot locations with frequent detections. To group the detections into clusters with a common source, we apply the DBSCAN clustering algorithm (Ester et al., 1996; Schubert et al., 2017). We cluster based on the distance between detections in meters and set a threshold of 5 detections within 30 km as the minimum to identify a cluster. We identify 94 clusters, this is a conservative estimate for the number of persistent locations as some known persistent emitters have fewer than 5 detections in 2021. We also observe several areas with extensive plumes from multiple emitters such as the west-coast of Turkmenistan which are grouped into one big cluster. We find the majority of detected plumes (74.8%) to be clustered at a persistent urban or fossil fuel exploitation source and classify the remaining plumes as transients. Zoom-ins of the clusters in several distinct regions as well as source rates for all detections are shown in Figure 7.
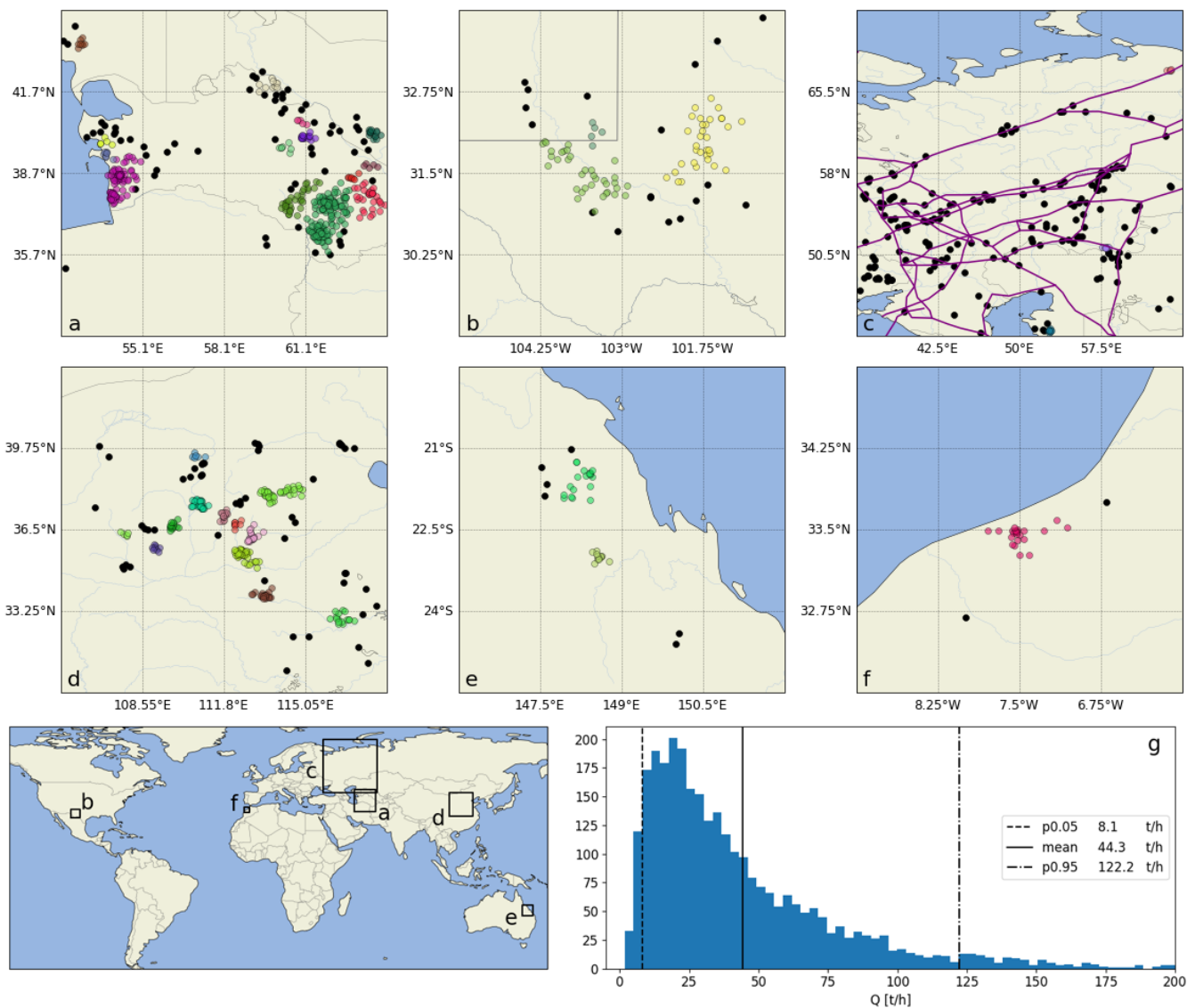
Figure 7. Regional plume detections showing color-coded persistent emission clusters with transient emissions shown in black. (a) Large clusters of detections related to oil/gas exploitation in Turkmenistan. (b) The clearly distinguishable outlines of the Delaware Basin and Midland Sub-Basins within the Permian Basin, US. (c) Detections show the same spatial structure along compressor stations and pipelines in western Russia. (d) Clusters of hotspots in eastern China with the extensive Shanxi coal-mining region in the center. (e) Clusters of coal mining detections in north-east Australia. (f) A clear cluster of detections around the persistent source in Casablanca, Morocco. (g) The distribution of estimated source rates for all 2974 detected plumes in the year 2021, capped at 200 t h^{-1}. The 5th and 95th percentile as well as the mean values of the distribution are shown as vertical lines.

Several of the identified clusters are located over well-known oil/gas production regions such as the west coast of Turkmenistan (Figure 7a) previously studied by Varon et al. (2019) and Irakulis-Loitxate et al. (2022b), Algeria (Varon et al., 2021), Libya, and multiple basins in the US (Shen et al., 2022) including the Eagle Ford, Haynesville, and most prominently the Permian (Figure 7b) where Zhang et al. (2020) quantified emissions based on TROPOMI and we find individual clusters of detections over the Midland and Delaware sub-basins. The fact that many of the detections are clustered around known large sources gives confidence in the performance of the models that did not use prior location information. We also identify oil/gas production clusters which have not been studied in detail, such as in northern Libya, Yemen, and northeastern India.

We also find large transient plumes along the major gas transmission pipelines in western Russia (Figure 7c), similar to what Lauvaux et al. (2022) found for 2019-2020. Clusters of detections are seen over coal mining areas in China (Chen et al., 2022), southern Poland (Tu et al., 2022b), South Africa, Russia, and northeastern Australia (Figure 7e) where Sadavarte et al. (2021) quantified large emissions from these clusters of coal mines. Our approach allows us to detect which specific locations within a larger area of fossil fuel exploitation cause large methane plumes, examples are the super-emitter clusters within the large, spread-out Shanxi coal-mining region in China (Figure 7d).

The majority of our detections are related to urban areas around the world, including four cities with large fluxes (Buenos Aires, Mumbai, Delhi and Lahore), which were also identified by Maasakkers et al. (2022b) based on long-term wind-rotated TROPOMI averages. Urban areas comprise a range of source types but individual landfills can make up a large fraction of total urban emissions (Maasakkers et al., 2022b). When we zoom into the area around Casablanca, Morocco (Figure 7f), we see strong convergence in a cluster. Most plumes within the cluster (19 out of 23) are quantified below 25 t h^{-1}, of which 8 are quantified below 15 t h^{-1}. The estimated source locations of the plumes are on average 12 km away from a landfill later detected and quantified using GHGSat observations (Figure 8, Section 3.2). Other urban clusters include Madrid in Spain; 7 cities in Pakistan; Riyadh in Saudi Arabia; Bucharest in Romania; and Mexico City and Guadalajara in Mexico. The most frequently detected (104 detections) urban cluster is centered around Dhaka, Bangladesh. In India, we see 8 urban clusters and several cities with at least two detections. Detections over India are seasonally limited by meteorology as there are hardly any TROPOMI data during the May-September monsoon season because of persistent cloud cover.

The distribution of the estimated source rates of all 2974 verified plumes is shown in Figure 7g. Our IME-based quantifications show mean emissions of 44 t h^{-1}, with a large 5-95th percentile range of 8-122 t h^{-1}. Many detections are quantified below the detection threshold of previous TROPOMI plume identification & quantification methods of 25 t h^{-1} (Lauvaux et al., 2022; Jacob et al., 2022). We find 1143 plumes quantified under 25 t h^{-1} including 241 plumes under 10 th^{-1}. Many of these originate from persistent emission clusters where emissions have been confirmed using high-resolution instruments. Although the applied mass balance quantification method has significant uncertainty, this shows the plume detection limit of TROPOMI is better than previously reported in literature.

In order to present a rough estimate for the total emissions represented by the detected plumes, we assume each emission event is active for 24 hours (the minimum sampling frequency of TROPOMI). For some transient plumes, such as pulse emissions at compressor stations, the 24 hours estimate can be an overestimate, but we take these as representative for similar transient events occurring outside of the TROPOMI observation window. Using these assumptions, we find detected emissions of 3.1 $\pm$1.3 Tg for 2021. As a conservative uncertainty estimate we use the sum of the standard deviations of the individual ensembles. The number of detected plumes is an underestimate of the true number of plumes as observations are limited by clouds and illumination.

To account for the limited TROPOMI coverage and get an indication of the annual emissions our detections are representative of, we scale our detected emissions by the fraction of days with coverage. We estimate the local number of days with coverage from our 794,395 valid scenes by first mapping their spatial footprints to a 0.1x0.1 degree grid, removing duplicate coverage from overlapping scenes in the same orbit. We then correct for local variations in coverage (such as persistent areas without data) by convoluting this field with the summed footprints of all 2021 TROPOMI data at 0.1x0.1 degrees. We finally aggregate our detected emissions to a 1x1 degree grid and divide those by the fraction of days in 2021 with coverage resulting from the coverage map average to 1x1 degrees. We find a scaled-up annual emission flux of 10.3 Tg, which is approximately 2.7% of total bottom-up 2017 anthropogenic emissions (380 Tg yr^{-1}) (Saunois et al., 2020). Super-emitter plumes from landfills account for 4.1 Tg yr^{-1} (6% of global emissions), those from coal 2.1 Tg yr^{-1} (4.7%), those from oil 2.2 Tg yr^{-1} and those from gas 1.9 Tg yr^{-1} (4.9% of global oil and gas) (Saunois et al., 2020). These estimates are only small fractions of total anthropogenic emissions as our conservative upscaling approach only takes large TROPOMI-detected super-emitter plumes into account. Emissions from smaller point sources and area sources make a large contribution to the total, but are not part of our upscaling. Such emissions are better captured by an atmospheric inversion.

3.2 Synergy of automated TROPOMI detections with high-resolution instruments

We use the detection of persistent methane plumes in TROPOMI data to target high-resolution observations (GHGSat-C1/C2)
515 and data analysis (PRISMA archive, Sentinel-2) following (Maasakkers et al., 2022b). Furthermore, we investigate large transient emissions with data from non-targeted instruments such as Sentinel-2.
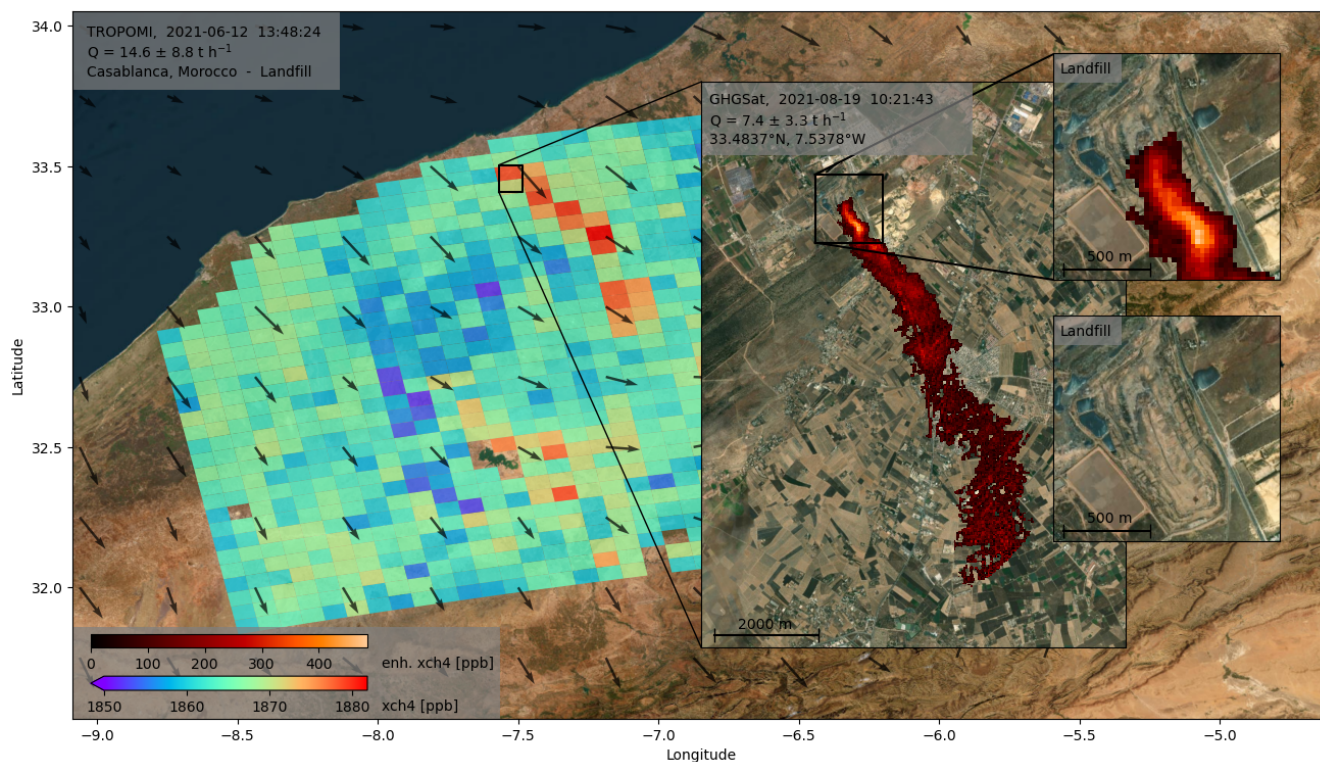
Figure 8. Methane plumes detected from Casablanca, Morocco on two different days, with TROPOMI and GHGSat data overlaid over visual ESRI World Imagery (Esri, Maxar, Earthstar Geographics, and the GIS User Community, 2022). Timestamps are UTC. The plume observed by TROPOMI on 2021-06-12 is quantified at 14.6 ± 8.8 t h^{-1}. The plume observed by GHGSat on 2021-08-19 originates from the landfill between Casablanca and Médiouna (33.483, -7.538) and is quantified at 7.4 ± 3.3 t h^{-1}. The winds are the GEOS-FP 10 meter windfield (Molod et al., 2012).

Figure 8 shows a TROPOMI plume (14.6 ± 8.8 th^{-1}) detected near Casablanca in Morocco on June 12th, 2021. We detected 23 plumes in the area in 2021 with source rates ranging from 8.9 ± 5.1 t h^{-1} to 40.5 ± 18.0 t h^{-1} with a mean of 18.8 t h^{-1}, indicating a persistent source (Figure 7f). Based on wind-rotation analysis (Maasakkers et al., 2022b), we find the landfill lo-
520 cated in between Casablanca and Médiouna to be the optimal target for high-resolution observations. Based on this TROPOMI analysis, we "tip-and-cue" GHGSat to observe this location. The inset image shows a targeted GHGSat-C2 observation on August 19th, 2021, indeed showing a methane plume (quantified at 7.4 ± 3.3 t h^{-1}) originating from the landfill and extending

downwind.

Figure 9 shows two methane plumes detected with TROPOMI in northern Kazakhstan on May 14th, 2021 at 07:53 UTC, quantified at 35.2 ± 13.2 t h^{-1} for the northern and 28.1 ± 11.2 t h^{-1} for the southern plume. The same locations were also detected in an adjacent orbit at 09:33 UTC but the closest days with coverage before and after May 14 do not show emissions, which indicates the plumes are transient. The bottom-up inventories give natural gas systems as the locally dominant anthropogenic source sector because of the presence of a gas transmission pipeline. In Sentinel-2 observations taken 38 minutes before the first detection, we find two emitting locations close to the pipeline in the upwind part of the TROPOMI plume masks. The source rates of the Sentinel-2 plumes are 180 ± 59 t h^{-1} and 75 ± 23 t h^{-1} for the northern and southern plume respectively. The rather large discrepancy between the TROPOMI and Sentinel-2 quantifications can be explained by the uncertain low wind speeds, the not well-developed plume in TROPOMI increasing the uncertainty in the IME, and possibly the partial pixel enhancement effect described by Pandey et al. (2019).
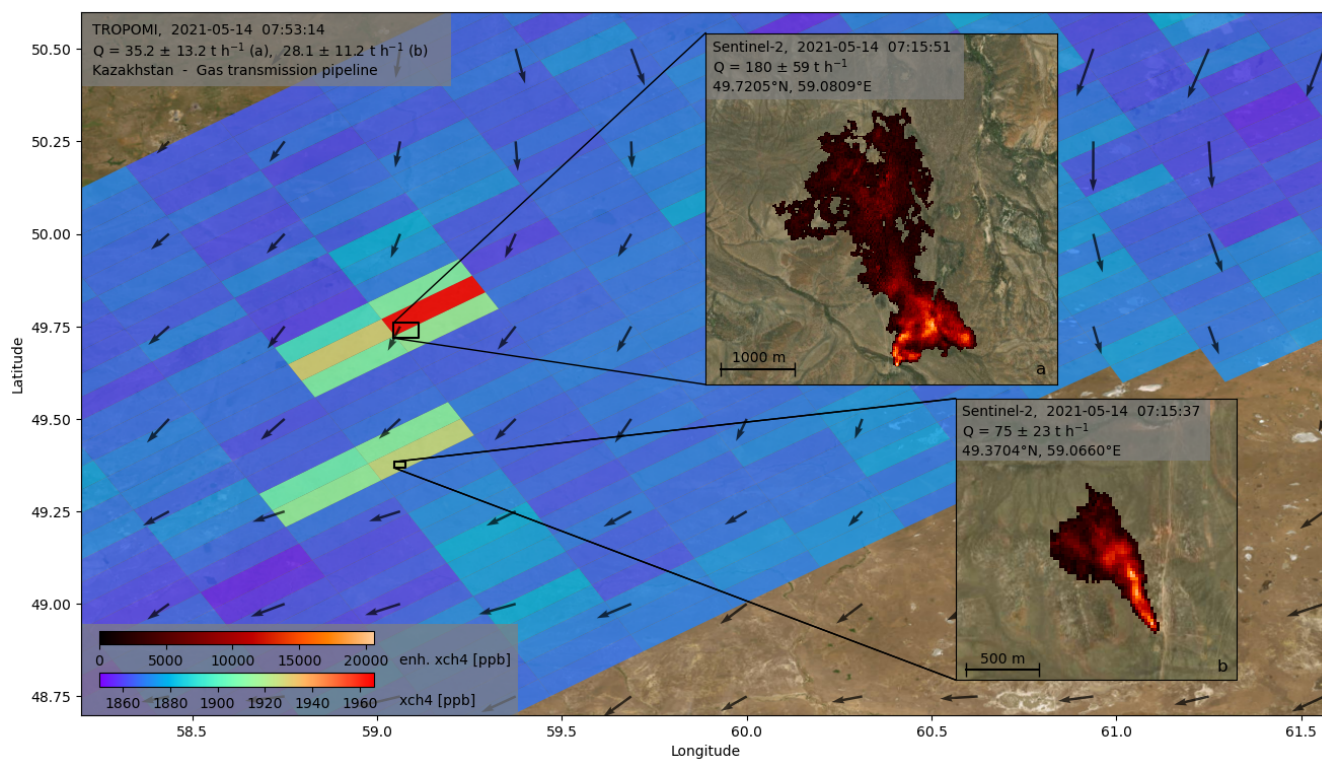
Figure 9. Transient methane plumes detected at two different locations in northern Kazakhstan with TROPOMI (quantified at 35.2 ± 13.2 t h^{-1} for the northern and 28.1 ± 11.2 t h^{-1} for the southern plume) and Sentinel-2 overlaid over visual ESRI World Imagery (Esri, Maxar, Earthstar Geographics, and the GIS User Community, 2022). Timestamps are UTC. The plumes originate from natural gas pipeline infrastructure. The winds are the GEOS-FP 10 meter windfield (Molod et al., 2012).
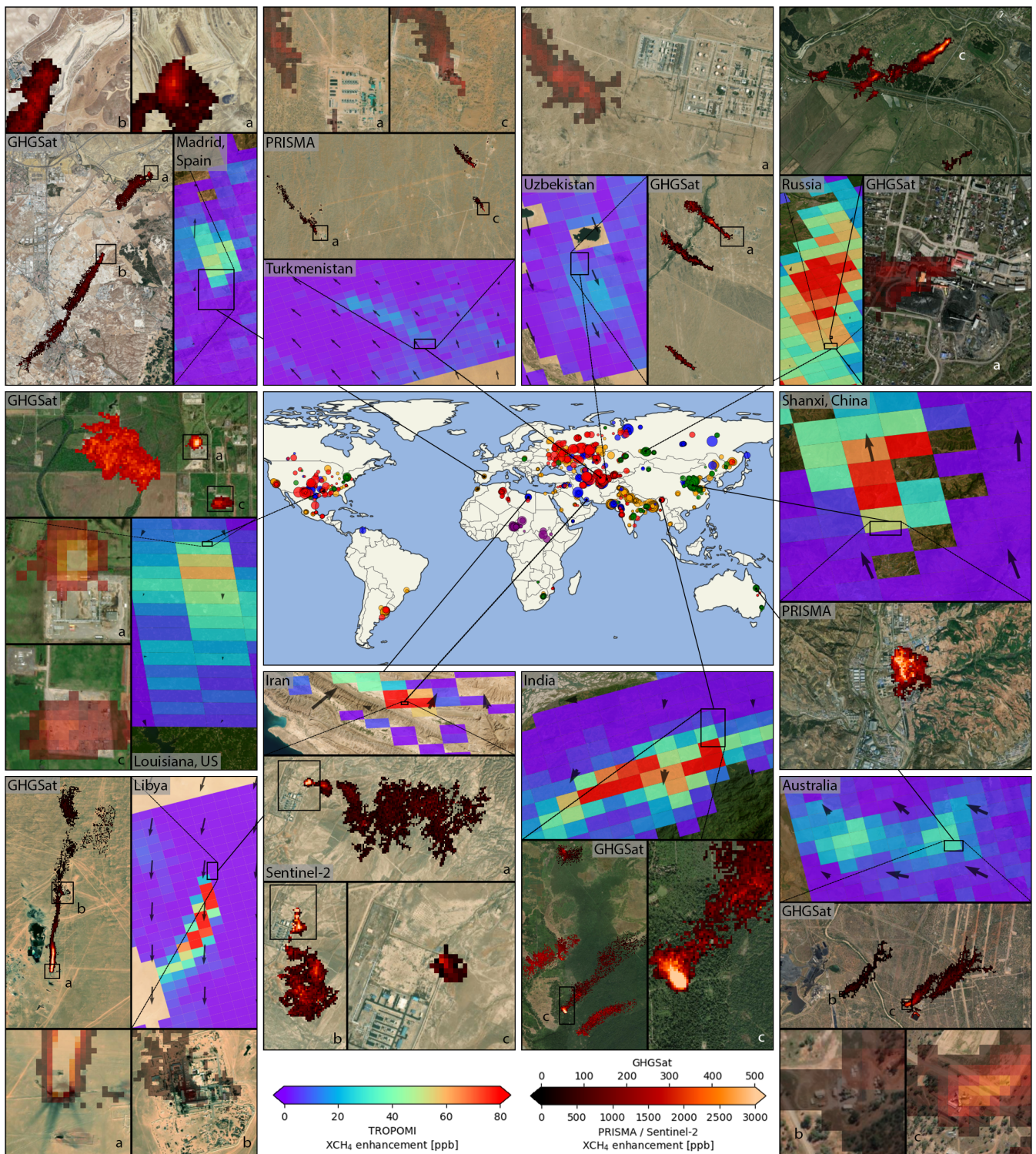
Figure 10. (Caption next page.)

Figure 10. (Previous page.) Plumes detected over 10 locations which were inspected with high-resolution instruments. Observations at the same location with different instruments are most often not on the same day. Details are provided in Tables B1-B3 for the high-resolution instruments and in Table B4 for TROPOMI. TROPOMI data is shown in Mercator projection (EPSG:4326), high-resolution data is shown in the local Universal Transverse Mercator (UTM) projection. The data is overlain over visual ESRI World Imagery (Esri, Maxar, Earthstar Geographics, and the GIS User Community, 2022). The worldmap in the center of the image corresponds to Figure 6, showing all 2974 detected plumes in 2021. TROPOMI data is here displayed as enhancement relative to the median XCH_4 of the 32x32 pixel scene. Several of the zoomed-in views with high-resolution data were set to an opacity of 0.5 in order to reveal the infrastructure at the source of the plume.

535   Figures 8 and 9 show how TROPOMI detections can be combined with high-resolution observations for both persistent and transient emitters. Figure 10 shows ten additional locations analyzed with GHGSat (7 scenes), PRISMA (2), and Sentinel-2 (1) based on TROPOMI detections. These 12 selected locations show the range of typical anthropogenic source types and intermittencies we have observed with both TROPOMI and high-resolution instruments. Supplemental Tables B1-B3 provide details on these high resolution observations and associated (not necessarily on the same day) TROPOMI scenes (Table B4).

540   We find facility-level source rates from 0.3 ± 0.1 t h^{-1} up to 16 ± 5 t h^{-1}. Because of the different spatial footprints, sensitivities, and detection dates, these emission rates cannot be directly compared to the TROPOMI emission estimates (Maasakkers et al., 2022b).

The different specifications of the high-resolution instruments make them suitable for different purposes. The methane designated GHGSat-Cx instruments have the lowest detection limit and are capable of retrieving methane over areas with challenging

545   surface structures such as urban areas. Southeast of Madrid we observe plumes from two separate landfills located 7 km apart. The methane plumes originate from the active areas of the landfills where waste is added. In TROPOMI, the signals from these two landfills appear as a single point source. In the gas production region around Shreveport, Louisiana, US, we observe three emission plumes originating from distinct infrastructure including from two facilities that are only ∼500 m apart. Over a coal mining area in Russia we see in a GHGSat observation that a single TROPOMI-based target has contributions from 10 different

550   point sources with source rates from 0.2 ± 0.1 to 2.4 ± 1.1 t h^{-1}. The emissions originate from coal mining facilities, such as underground mine vents, add up to 8.8 t h^{-1}. In the Assam oil and gas production region in India, we find five plumes adding up to 12.9 t h^{-1}, showing the GHGSat-Cx instruments are also capable of retrieving methane plumes over non-homogeneous areas including forest and agricultural lands. We also use GHGSat to target two less challenging desert scenes with oil and gas production. A GHGSat observation over Libya shows two sources downwind of each other, the most upwind source is

555   an unlit flare stack. At the border of Uzbekistan and Turkmenistan (listed as Uzbekistan in Figure 10), we find emissions at three distinct locations within a single natural gas facility, with source rates ranging from 3.6 ± 1.0 to 6.4 ± 1.8 t h^{-1}. These emissions appear to originate from unlit flaring stacks, similar to what Irakulis-Loitxate et al. (2022b) and Varon et al. (2019) found for other natural gas facilities in the region.

560   For scenes with more homogeneous surfaces or extremely large emissions, PRISMA and Sentinel-2 can also be used. The PRISMA observation in Turkmenistan shows three plumes with an aggregated emission rate of 9.0 ± 2.9 t h^{-1}. The emissions

originate from distinct pieces of gas infrastructure (quantified at 3.5 ± 1.1, 3.0 ± 1.0 and 2.5 ± 0.8 th^{-1}), the sources are located within the footprint of a single TROPOMI pixel (the same source as shown in Figures 2 and 5). We also use the PRISMA archive to detect a plume originating from a coal mine ventilation shaft in Liuzhuang Village in Shanxi, China. Finally, we use Sentinel-2 to investigate a single location in Iran with complex observation conditions (elevation) where we only had a single TROPOMI detection in 2021. The emission therefore appeared to be transient at first, but with Sentinel-2 we find three emission plumes ranging from 1.3 ± 0.4 to 10.0 ± 3.0 t h^{-1} originating from the same oil facility in a timespan of two months. Extensive monitoring of a location of interest over a longer timespan is feasible using Sentinel-2 (Varon et al., 2021; Irakulis-Loitxate et al., 2022b).

570 4   Conclusions

We detected methane emission plumes in 2021 TROPOMI data using an automated, machine-learning-based pipeline. We have trained a Convolutional Neural Network with a relatively small set of manually identified plumes in pre-2021 TROPOMI methane data to detect plume-like morphological structures (kappa score of 0.97 and recall of 0.98 on the test set). We then used a Support Vector Classifier to distinguish real plumes from retrieval artefacts using additional information from the scene

575 and supporting data (kappa score of 0.81 and recall of 0.93 on the test set). This two-step approach can also be applied to other instruments in the future. We tested our detection, source localization, and emission quantification estimate for a specific well-characterized natural gas well blowout and found it was accurately captured by our monitoring system. After application of our pipeline to 2021 data, we targeted high-resolution observations and analyses to find the facilities responsible for twelve (clusters of) plumes seen in TROPOMI.

580

Using our automated machine learning pipeline we scan all 794.395 scenes of 2021 in three hours on a single core. 4869 scenes are automatically classified as plume, of which 2974 are manually verified as confident plumes and 745 as potential plumes; giving the automated pipeline a precision of 61-76%. The most challenging distinction for the SVC is between plume and artefact, a distinction that even for a human expert can be inconclusive for difficult cases. We focus on the manually verified

585 plumes, the remaining 39% of the scenes are mostly difficult to classify and can still be followed up with manual inspection or be used to further train the models. We find that most plumes (74.8%) originate from 94 clusters of detections around both known and new persistent source locations. The other plumes are mainly caused by transient emission events, such as along natural gas transmission pipelines in Russia. We most often detect plumes (based on bottom-up emission inventories) from urban areas / landfills (1031 plumes), followed by 720 plumes from gas infrastructure, 612 from oil infrastructure and 581

590 from coal mining. Many of the identified clusters are located at well-known fossil fuel exploitation regions, or urban areas known to emit methane. We also identify several previously unstudied sources such as in Libya and Assam (India) and identify specific super-emitting locations within spread-out fossil fuel production regions like the Shanxi coal mining area in China. Based on IME quantifications of all plumes, we found mean emissions of 44 t h^{-1} with a 5-95th percentile range of 8-122 t h^{-1}, which is an indication of the TROPOMI detection limit. With 1143 detections under 25 t h^{-1} including 241 plumes under 10

595 th^{-1}, our automated approach has a better detection limit than previously published methods based on TROPOMI data. When we assume all 2944 detected anthropogenic emissions to be active for 24 hours, we find detected 2021 emissions of 3.1 $\pm$1.3 Tg. Accounting for the limited coverage of TROPOMI, these detected emissions are representative of 10.3 Tg yr^{-1}, which is approximately 2.7% of global annual anthropogenic emissions.

600 For twelve locations, we used high-resolution satellite observations (GHGSat-C1/C2, PRISMA, and Sentinel-2) to identify the exact sources responsible for the detected plumes in TROPOMI. We utilized the different strengths of the high-resolution instruments; we made targeted observations with GHGSat over scenes with complex surface reflectance, whereas the archive of Sentinel-2 is used to analyze large transient emission events and track intermittent emissions. We found point sources from

landfills and fossil fuel exploitation with emission rates from 0.3 ± 0.14 t h^{-1} to 180 ± 59 t h^{-1}. Most fossil-fuel-related

605 TROPOMI plumes had contributions from multiple point sources, with one GHGSat observation over Russia revealing emissions from ten different sources.

Over the next few years, the number of global, regional, and point-source mapping instruments capable of retrieving methane plumes will vastly increase; with i.a. Sentinel-5, CO2M, MethaneSAT and Carbon Mapper (Jacob et al., 2022). Our monitor-

610 ing system can incorporate these fast-growing data volumes and can already be used to automatically detect plumes in the operational TROPOMI data, track temporal variability in super-emitter plumes, and "tip-and-cue" high-resolution satellite instruments to find the associated super-emitting facilities. This identification and monitoring of super-emitters with large mitigation potential is paramount to reach the goals of the Global Methane Pledge.

Code and data availability. TROPOMI data are publicly available at https://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/ch4/18_17/.

615 GHGSat-C1/C2 methane plume data used in this study are available at https://doi.org/10.7910/DVN/QQQ9IU. Sentinel-2 data are publicly available at the Copernicus Open Access Hub https://scihub.copernicus.eu/. PRISMA data are publicly available at https://prisma.asi.it/, after registering at https://prismauserregistration.asi.it. GEOS-FP wind data can be downloaded at https://gmao.gsfc.nasa.gov/GMAO_products/. ERA5 wind data are available at https://cds.climate.copernicus.eu. WRF-CHEM code is available at https://github.com/wrf-model/WRF/releases, in this work version 4.1.5 was used. The GFEI (v2) emission inventory

620 is available at https://doi.org/10.7910/DVN/HH4EUM. The WetCHARTs emission inventory is available at https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1502. EDGAR v6 data is available at http://data.europa.eu/89h/97a67d67-c62e-4826-b873-9d972c4f670b. The dataset of detected plumes in 2021 TROPOMI data will be made available upon publication.

Appendix A: CNN trainingdata table

	split	subsplit	Nr of scenes	Nr of scenes augmented
All labeled scenes	1.0		3070	
training set	0.8		2456	19648
- training subset		0.8	- 1965	- 15718
- validation subset		0.2	- 491	- 3930
test set	0.2		614	4912

Table A1. An overview of the split in training and test data for the CNN.

Appendix B: Details on the plumes observed with high resolution instruments

GHGSat		Observation time [UTC]	Latitude [deg]	Longitude [deg]	Source rate [t h^{-1}]	Wind speed U_{10} [m s^{-1}]	Sector
Casablanca, Morocco		2021-08-19 10:21:43	33.4837	-7.5378	7.4 ±3.3	2.6	Landfill
Madrid, Spain	a	2021-12-13 10:04:26	40.3222	-3.5913	4.3 ±2.0	2.7	Landfill
	b	2021-12-13 10:04:26	40.2611	-3.6364	5.6 ±2.5	2.7	Landfill
Libya	a	2021-04-17 08:30:12	28.9089	20.9807	8.8 ±2.0	8.6	Oil exploitation
	b	2021-04-17 08:30:12	28.9400	20.9856	2.7 ±0.6	8.6	Oil exploitation
Australia	a	2021-06-08 23:23:08	-21.8321	148.0099	0.7 ±0.2	6.4	Coal mine
	b	2021-06-08 23:23:08	-21.8847	147.9737	1.1 ±0.3	6.4	Coal mine
	c	2021-06-08 23:23:08	-21.8883	147.9949	2.9 ±0.8	6.4	Coal mine
Uzbekistan	a	2021-12-20 06:22:11	38.7989	64.6410	6.4 ±1.8	6.5	flaring stack
	b	2021-12-20 06:22:11	38.7839	64.6289	4.0 ±1.1	6.5	O&G facility
	c	2021-12-20 06:22:11	38.7347	64.6169	3.6 ±1.0	6.5	O&G facility
Louisiana, US	a	2021-04-19 16:04:48	32.1979	-93.4882	0.5 ±0.4	0.1	O&G facility
	b	2021-04-19 16:04:48	32.1960	-93.4996	2.3 ±2.0	0.1	O&G facility
	c	2021-04-19 16:04:48	32.1900	-93.4856	0.4 ±0.3	0.1	O&G facility
Russia	a	2021-08-25 04:26:02	54.6520	86.1506	1.3 ±0.6	2.8	Coal facility
	b	2021-08-25 04:26:02	54.6313	86.1736	0.3 ±0.1	2.8	Coal
	c	2021-08-25 04:26:02	54.6190	86.1747	2.4 ±1.1	2.8	Coal
	d	2021-08-25 04:26:02	54.6120	86.1498	1.6 ±0.7	2.8	Coal
	e	2021-08-25 04:26:02	54.6122	86.1315	0.9 ±0.4	2.8	Coal
	f	2021-08-25 04:26:02	54.6080	86.1496	0.5 ±0.2	2.8	Coal
	g	2021-08-25 04:26:02	54.5964	86.1828	0.4 ±0.2	2.8	Coal
	h	2021-08-25 04:26:02	54.6168	86.1465	1.0 ±0.4	2.8	Coal
	i	2021-08-25 04:26:02	54.5904	86.1812	0.2 ±0.1	2.8	Coal
	j	2021-08-25 04:26:02	54.5621	86.2065	0.4 ±0.2	2.8	Coal

Table B1. Observation, location and quantification details corresponding to the GHGSat scenes. One of which is Morocco, Casablanca, Landfill in Figure 8. Windspeeds are 10m windspeeds obtained from GEOS-FP (Molod et al., 2012).

Locations Uzbekistan-b and Uzbekistan-c are located just over the border in Turkmenistan, however the main building of the facility (near Uzbekistan-a) is located in Uzbekistan. Because there is another location in Turkmenistan we have chosen this naming.

GHGSat		Observation time [UTC]	Latitude [deg]	Longitude [deg]	Source rate [t h^{-1}]	Wind speed U_{10} [m s^{-1}]	Sector
India	a	2021-12-24 03:21:48	27.4626	95.4788	1.2 $\pm$0.8	0.6	Oil exploitation
	b	2021-12-24 03:21:48	27.3902	94.4695	2.3 $\pm$1.6	0.6	Oil exploitation
	c	2021-12-24 03:21:48	27.3545	95.4733	2.8 $\pm$2.0	0.6	Oil exploitation
	d	2021-12-24 03:21:48	27.3402	95.4830	3.1 $\pm$2.2	0.6	Oil exploitation
	e	2021-12-24 03:21:48	27.3771	95.4472	3.5 $\pm$2.5	0.6	Oil exploitation

Table B1. Observation, location and quantification details corresponding to the GHGSat scenes. One of which is Morocco, Casablanca, Landfill in Figure 8. Windspeeds are 10m windspeeds obtained from GEOS-FP (Molod et al., 2012).

PRISMA		Observation time [UTC]	Latitude [deg]	Longitude [deg]	Source rate [t h^{-1}]	Wind speed U_{10} [m s^{-1}]	Sector
Turkmenistan	a	2021-02-13 06:51:45	40.0106	60.9346	3.0 $\pm$1.0	1.9	O&G exploitation
	b	2021-02-13 06:51:45	40.0496	61.0456	3.5 $\pm$1.1	1.9	O&G exploitation
	c	2021-02-13 06:51:45	40.0240	61.0536	2.5 $\pm$0.8	1.9	O&G exploitation
Shanxi, China		2021-12-22 03:18:33	35.6083	112.5282	11.6 $\pm$3.7	2.7	Coal facility

Table B2. Observation, location and quantification details corresponding to the PRISMA scenes. Windspeeds are 10m windspeeds obtained from GEOS-FP (Molod et al., 2012).

The plume mask of plume Turkmenistan-c (Figure 10, Table B2) was curated in order to exclude an artifact which was caused by a nearby road. Retrieval artefacts in high-resolution methane retrievals from hyper-spectral instruments resulting from surface features such as roads is a known issue (Sánchez-García et al., 2022; Gorroño et al., 2022). PRISMA and Sentinel-2 are more prone to such issues than GHGSat-Cx.

Sentinel-2		Observation time [UTC]	Latitude [deg]	Longitude [deg]	Source rate [t h^{-1}]	Wind speed U_{10} [m s^{-1}]	Sector
Kazakhstan	a	2021-05-14 07:15:51	49.7205	59.0809	179.8 $\pm$59.1	2.6	Natural gas pipeline
	b	2021-05-14 07:15:37	49.3704	59.0660	74.6 $\pm$23.0	2.2	Natural gas pipeline
Iran	a	2021-08-24 07:12:03	27.5345	53.2946	10.0 $\pm$3.0	2.1	O&G facility
	b	2021-09-08 07:11:57	27.5345	53.2946	16.1 $\pm$5.3	2.7	O&G facility
	c	2021-10-23 07:12:06	27.5345	53.2946	1.3 $\pm$0.4	1.8	O&G facility

Table B3. Observation, location and quantification details corresponding to the Sentinel-2 scenes. One of which is Kazakhstan, Natural gas pipeline in Figure 9. Windspeeds are 10m windspeeds obtained from GEOS-FP (Molod et al., 2012).

TROPOMI		Observation time [UTC]	Latitude [deg]	Longitude [deg]	Source rate [t h^{-1}]	Wind speed GEOS-10m / GEOS-PBL / ERA5-10m [m s^{-1}]	Sector estimate (bottom-up)
Casablanca, Morocco		2021-06-12 13:48:37	33.48	-7.54	14.6 ±8.8	6.9 / 2.6 / 2.8	Landfill
Madrid, Spain		2021-01-05 13:13:19	40.30	-3.64	6.8 ±2.6	0.6 / 0.6 / 1.0	Landfill
Libya		2021-07-26 11:40:42	28.88	20.93	87.3 ±28.7	4.4 / 4.6 / 3.4	Oil
Australia		2021-09-25 03:56:01	-21.91	148.06	17.5 ±6.1	3.0 / 2.0 / 2.5	Coal
Uzbekistan		2021-08-10 08:39:36	38.76	64.59	15.2 ±5.6	4.2 / 3.6 / 4.9	Gas
Louisiana, US		2021-09-23 20:02:42	32.13	-93.72	31.1 ±12.1	1.1 / 1.9 / 1.6	Gas
Russia		2021-06-08 06:43:36	54.61	86.10	40.4 ±16.9	1.1 / 1.8 / 1.0	Coal
India		2021-01-04 06:42:43	27.38	95.69	21.1 ±9.0	1.0 / 0.8 / 0.5	Gas
Turkmenistan		2021-12-05 08:47:19	40.10	60.99	11.1 ±3.9	2.4 / 3.2 / 3.3	Oil
Shanxi, China		2021-01-23 05:47:47	35.63	112.52	47.7 ±15.3	3.5 / 4.6 / 3.6	Coal
Kazakhstan	a	2021-05-14 07:53:19	49.85	59.12	35.2 ±13.2	2.2 / 2.5 / 1.4	Gas
Kazakhstan	b	2021-05-14 07:53:14	49.46	59.08	28.1 ±11.2	2.3 / 3.0 / 1.5	Gas
Iran		2021-11-04 10:05:30	27.48	53.41	45.4 ±13.7	2.1 / 1.8 / 2.0	Oil

Table B4. Observation, location and quantification details of the TROPOMI scenes corresponding to the high resolution observations in Figures 8, 9 and 10. Windspeeds presented in this table are the GEOS-10m, GEOS-PBL and ERA5-10m windproducts (Molod et al., 2012; Hersbach et al., 2020) which are used to compute three U_{eff} values, which are then averaged (Section 2.5.1).

625 **Appendix C: Features used as input for the SVC**

Feature name	Category	FIR	Value range	Description
CNN_score	plume-like morphology	1	[0.5, 1]	prediction score assigned by the CNN
valid_pixels	quality of the scene	7	[0.2, 1]	fraction of valid pixels in the scene, N/(32x32)
count_mask_high	plume mask		[1, >	number of pixels in the high confidence plume mask
sum_mask_high	plume mask		[1, >	sum of enhancement of the pixels in the high confidence plume mask
stdev	xch4 statistical distribution		[0, >	standard deviation of the xch4 value of all pixels in the scene
skew	xch4 statistical distribution		-2, >	skewness of the xch4 value of all pixels in the scene
kurtosis	xch4 statistical distribution		[1, >	kurtosis of the xch4 value of all pixels in the scene
IME	magnitude of plume		[0, >	[kg] Integrated Mass Enhancement (section 2.5.1)
L	magnitude of plume		[0, >	[m] plume length (section 2.5.1)
U_10	wind		[0, ~8]	[ms^{-1}] 10 meter windspeed, obtained from ERA5, present in Level 2 dataproduct
cba_sum	clouds		[0, >	sum product of a 3x3 kernel multiplying enhancements of the high conf plume with cloud fraction
cba_count	clouds		[0, >	number of pixels in the high confidence plume mask close to cloudy pixels based on 3x3 kernel
angle_PCA_mean_wind _plume_mask_weighted	wind	8	[0, 90]	[deg] angle between principle axis of the plume mask and mean windvector (section 2.3)
exp_var_ratio _pca_weighted	plume elongation		[0, >	ratio between the variance along the primary and secondary axis
albedo_rvalue_bg	correlation xch4 & supporting data		[-1, 1]	Pearson r-value between xch4 and albedo for the full scene
albedo_rvalue_dil_1	correlation xch4 & supporting data	3	[-1, 1]	Pearson r-value between xch4 and albedo for pixels within 1 dilation around the low confidence mask

Table C1. Overview of the features used by the SVC as input. Each scene is represented as a feature vector with shape [1x41]. This table provides the name of the feature, the category the feature is aimed to provide information about, the ranking of the top-10 features in the feature importance analysis (FIR = Feature Importance Ranking), the possible range of values the feature can attain and a description of the feature.

Feature name	Category	FIR	Value range	Description
aero_rvalue_bg	correlation xch4 & supporting data		[-1, 1]	Pearson r-value between xch4 and AOT for the full scene
aero_rvalue_dil_1	correlation xch4 & supporting data		[-1, 1]	Pearson r-value between xch4 and AOT for pixels within 1 dilation around the low confidence mask
surf_pres_rvalue_bg	correlation xch4 & supporting data		[-1, 1]	Pearson r-value between xch4 and surface pressure for the full scene
surf_pres_rvalue_dil_1	correlation xch4 & supporting data		[-1, 1]	Pearson r-value between xch4 and surface pressure for pixels within 1 dilation around the low confidence mask
chi2_rvalue_bg	correlation xch4 & supporting data	5	[-1, 1]	Pearson r-value between xch4 and chi2 for the full scene
chi2_rvalue_dil_1	correlation xch4 & supporting data	0	[-1, 1]	Pearson r-value between xch4 and chi2 for pixels within 1 dilation around the low confidence mask
cloud_angle_high	clouds		[0, 90]	angle of the principle axis of the high confidence plume mask with the principle axis of a cloud
cloud_angle_low	clouds		[0, 90]	angle of the principle axis of the low confidence plume mask with the principle axis of a cloud
coast_angle	surface conditions		[0, 90]	angle of the principle axis of the high confidence plume mask with a coast
avg_chi2_mask_high	supporting data	4	[0, >	average chi2 value of the pixels within the high confidence plume mask
avg_chi2_mask_low	supporting data		[0, >	average chi2 value of the pixels within the low confidence plume mask
avg_albedo_mask_high	supporting data		[0, 1]	average albedo value of the pixels within the high confidence plume mask
avg_albedo_mask_low	supporting data		[0, 1]	average albedo value of the pixels within the low confidence plume mask
avg_aot_mask_high	supporting data		[0, 1]	average AOT value of the pixels within the high confidence plume mask
avg_aot_mask_low	supporting data		[0, 1]	average AOT value of the pixels within the low confidence plume mask

Table C1. Overview of the features used by the SVC as input. Each scene is represented as a feature vector with shape [1x41]. This table provides the name of the feature, the category the feature is aimed to provide information about, the ranking of the top-10 features in the

Feature name	Category	FIR	Value range	Description
avg_QA_mask_high	supporting data	2	[0.4, 1]	average QA value of the pixels within the high confidence plume mask
avg_QA_mask_low	supporting data	9	[0.4, 1]	average QA value of the pixels within the low confidence plume mask
std_bg_xch4_high	background homogeneity		[0, >	standard deviation of the xch4 outside of the high confidence plume mask, similar to pixel precision Varon2021
std_bg_xch4_low	background homogeneity		[0, >	standard deviation of the xch4 outside of the low confidence plume mask, similar to pixel precision Varon2021
avg_enh_plumemask _high_above_bg	magnitude of plume		[0, >	average enhancement above the background of the pixels within the high confidence plumemask
avg_enh_plumemask _low_above_bg	magnitude of plume		[0, >	average enhancement above the background of the pixels within the low confidence plumemask
max_enh_plumemask _high_above_bg	magnitude of plume	6	[0, >	maximum enhancement above the background of the pixels within the high confidence plumemask
frac_land_pixels_high	surface conditions		[0, 1]	fraction of pixels with surface classification 'land' in the high confidence plume mask
frac_landwater _pixels_high	surface conditions		[0, 1]	fraction of pixels with surface classification 'land+water' in the high confidence plume mask
frac_coast_pixels_high	surface conditions		[0, 1]	fraction of pixels with surface classification 'coast' in the high confidence plume mask

Table C1. Overview of the features used by the SVC as input. Each scene is represented as a feature vector with shape [1x41]. This table provides the name of the feature, the category the feature is aimed to provide information about, the ranking of the top-10 features in the feature importance analysis (FIR = Feature Importance Ranking), the possible range of values the feature can attain and a description of the feature.

Author contributions. BJS, JDM, and IA designed the study. BJS performed the TROPOMI analysis with contributions from PB, GM, SP, and JDM. BJS, JDM, and IA wrote the paper with contributions from all authors. AWB and SH provided the WRF simulations used to calibrate the TROPOMI IME method. AL and TB provided the TROPOMI methane data and associated support. DJV, JMcK, DJ, and MG provided the GHGSat data and supported the GHGSat analysis. II, JG, and LG provided the Sentinel-2 data/analysis. DHC provided the PRISMA data/analysis.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. We thank the team that realized the TROPOMI instrument and its data products, consisting of the partnership between Airbus Defence and Space Netherlands, KNMI, SRON, and TNO, commissioned by NSO and ESA. Sentinel-5 Precursor and Sentinel-2 are part of the EU Copernicus program, and Copernicus (modified) Sentinel-5P data (2018-2021) have been used as well as (modified) Sentinel-2 data (2021). The TROPOMI data processing was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. Financial support is acknowledged from the NSO TROPOMI national programme for AL, the GALES project (15597) by the Dutch Technology Foundation STW-NWO for SP, and ESA through EDAP for GM.

References

Bloom, A., Bowman, K., Lee, M., Turner, A., Schroeder, R., Worden, J., Weidner, R., McDonald, K., and Jacob, D.: CMS: Global 0.5-deg Wetland Methane Emissions and Uncertainty (WetCHARTs v1.3.1), https://doi.org/10.3334/ORNLDAAC/1915, 2021.

Borsdorff, T., Aan De Brugh, J., Hu, H., Hasekamp, O., Sussmann, R., Rettinger, M., Hase, F., Gross, J., Schneider, M., Garcia, O., Stremme, W., Grutter, M., Feist, D. G., Arnold, S. G., De Mazière, M., Kumar Sha, M., Pollard, D. F., Kiel, M., Roehl, C., Wennberg, P. O., Toon, G. C., and Landgraf, J.: Mapping carbon monoxide pollution from space down to city scales with daily global coverage, Atmospheric Measurement Techniques, 11, 5507–5518, https://doi.org/10.5194/amt-11-5507-2018, 2018.

Breiman, L.: Random Forests, Tech. rep., 2001.

CCAC: The Global Methane Pledge: Fast action on methane to keep a 1.5°C future within reach, https://www.globalmethanepledge.org/#about, 2022.

Chen, Z., Jacob, D. J., Nesser, H., Sulprizio, M. P., Lorente, A., Varon, D. J., Lu, X., Shen, L., Qu, Z., Penn, E., and Yu, X.: Methane emissions from China: a high-resolution inversion of TROPOMI satellite observations, Atmospheric Chemistry and Physics, 22, 10 809–10 826, https://doi.org/10.5194/acp-22-10809-2022, 2022.

Cheng, G., Xie, X., Han, J., Guo, L., and Xia, G. S.: Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 3735–3756, https://doi.org/10.1109/JSTARS.2020.3005403, 2020.

Chollet, F.: Deep Learning with Python, Manning, "ISBN-13: 978-1617296864", 2021.

Chollet, F. et al.: Keras, https://keras.io, 2015.

Cogliati, S., Sarti, F., Chiarantini, L., Cosi, M., Lorusso, R., Lopinto, E., Miglietta, F., Genesio, L., Guanter, L., Damm, A., Pérez-López, S., Scheffler, D., Tagliabue, G., Panigada, C., Rascher, U., Dowling, T. P., Giardino, C., and Colombo, R.: The PRISMA imaging spectroscopy mission: overview and first performance analysis, Remote Sensing of Environment, 262, https://doi.org/10.1016/j.rse.2021.112499, 2021.

Cohen, J.: A Coefficient of Agreement for Nominal Scales, Educational and Psychological Measurement, 20, 37–46, https://doi.org/10.1177/001316446002000104, 1960.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Lo Vullo, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J., and Vignati, E.: EDGAR v6.0 Greenhouse Gas Emissions [Dataset], European Commission, Joint Research Centre (JRC), PID: http://data.europa.eu/89h/97a67d67-c62e-4826-b873-9d972c4f670b, 2021.

Cusworth, D. H., Jacob, D. J., Varon, D. J., Miller, C. 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, pp. 5655–5668, https://doi.org/10.5194/amt-12-5655-2019, 2019.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Tseng, E., Thompson, D., Guha, A., Newman, S., Foster, K. T., and Miller, C. E.: Using remote sensing to detect, validate, and quantify methane emissions from California solid waste operations, Environmental Research Letters, 15, https://doi.org/10.1088/1748-9326/ab7b99, 2020.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Pandey, S., Maasakkers, J. D., Aben, I., Jervis, D., Varon, D. J., Jacob, D. J., Randles, C. A., Gautam, R., Omara, M., Schade, G. W., Dennison, P. E., Frankenberg, C., Gordon, D., Lopinto, E., and Miller, C. E.: Multisatellite Imaging of a Gas Well Blowout Enables Quantification of Total Methane Emissions, https://doi.org/10.1029/2020GL090864, 2021.

de Gouw, J. A., Veefkind, J. P., Roosenbrand, E., Dix, B., Lin, J. C., Landgraf, J., and Levelt, P. F.: Daily Satellite Observations of Methane from Oil and Gas Production Regions in the United States, Scientific Reports, 10, 1–10, https://doi.org/10.1038/s41598-020-57678-4, 2020.

Dekker, I. N., Houweling, S., Aben, I., Röckmann, T., Krol, M., Martínez-Alonso, S., Deeter, M. N., and Worden, H. M.: Quantification of CO emissions from the city of Madrid using MOPITT satellite retrievals and WRF simulations, Atmospheric Chemistry and Physics, 17, 14 675–14 694, https://doi.org/10.5194/acp-17-14675-2017, 2017.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, Remote Sensing of Environment, 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

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.

ESA, GHGSat: GHGSat mission overview, ESA, https://earth.esa.int/eogateway/missions/ghgsat, 2022.

Esri, Maxar, Earthstar Geographics, and the GIS User Community: ESRI World Imagery, https://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer, 2022.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Tech. rep., www.aaai.org, 1996.

European Commission: Launch by United States, the European Union, and Partners of the Global Methane Pledge to Keep 1.5C Within Reach, https://ec.europa.eu/commission/presscorner/detail/en/statement_21_5766, 2021.

Finch, D. P., Palmer, P. I., and Zhang, T.: Automated detection of atmospheric NO2 plumes from satellite data: A tool to help infer anthropogenic combustion emissions, Atmospheric Measurement Techniques, 15, 721–733, https://doi.org/10.5194/amt-15-721-2022, 2022.

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 heavytail flux distribution in Four Corners region, Proceedings of the National Academy of Sciences of the United States of America, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, 2016.

Gorroño, J., Varon, D. J., Irakulis-Loitxate, I., and Guanter, L.: Understanding the potential of Sentinel-2 for monitoring methane point emissions, Atmospheric Measurement Techniques Discussions, https://doi.org/10.5194/amt-2022-261, 2022.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957–6975, https://doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

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

Hasekamp, O., Lorente, A., Hu, H., Butz, A., Aan de Brugh, J., and Landgraf, J.: Algorithm Theoretical Baseline Document for Sentinel-5 Precursor methane Retrieval, SRON The Netherlands Institute for Space Research, Leiden, The Netherlands, https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-Methane-retrieval.pdf/f275eb1d-89a8-464f-b5b8-c7156cda874e?t=1658313508597, 2022.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, IEEE Computer Society, Los Alamitos, CA, USA, https://doi.org/10.1109/CVPR.2016.90, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hu, H., Landgraf, J., Detmers, R., Borsdorff, T., Aan de Brugh, J., Aben, I., Butz, A., and Hasekamp, O.: Toward Global Mapping of Methane With TROPOMI: First Results and Intersatellite Comparison to GOSAT, Geophysical Research Letters, 45, 3682–3689, https://doi.org/10.1002/2018GL077259, 2018.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, vol. In Press, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021.

Iqbal, H.: HarisIqbal88/PlotNeuralNet v1.0.0 (v1.0.0). Zenodo., https://doi.org/10.5281/zenodo.2526396, 2018.

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

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

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, Atmos. Chem. Phys, 16, 14 371–14 396, https://doi.org/10.5194/acp-16-14371-2016, 2016.

Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. K., Worden, J. R., and Duren, R. M.: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane, Atmospheric Chemistry and Physics, 22, 9617–9646, https://doi.org/10.5194/acp-22-9617-2022, 2022.

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

Johnson, J. M. and Khoshgoftaar, T. M.: Survey on deep learning with class imbalance, Journal of Big Data, 6, https://doi.org/10.1186/s40537-019-0192-5, 2019.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, http://arxiv.org/abs/1412.6980, 2014.

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

LeCun, Y., Kavukcuoglu, K., and Farabet, C.: Convolutional Networks and Applications in Vision, in: Proceedings of 2010 IEEE International Symposium on Circuits and Systems, https://doi.org/10.1109/ISCAS.2010.5537907, 2010.

Li, L., Jamieson, K., Rostamizadeh, A., and Talwalkar, A.: Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization, Journal of Machine Learning Research, 18, 1–52, http://jmlr.org/papers/v18/16-558.html., 2018.

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., Aan De Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., Pollard, D. F., Shiomi, K., Deutscher, N. M., Velazco, V. A., Roehl, C. M., Wennberg, P. O., Warneke, T., and Landgraf, J.: Methane retrieved from TROPOMI: Improvement of the data product and validation of the first 2 years of measurements, Atmospheric Measurement Techniques, 14, 665–684, https://doi.org/10.5194/amt-14-665-2021, 2021.

Maasakkers, J. D., Omara, M., Gautam, R., Lorente, A., Pandey, S., Tol, P., Borsdorff, T., Houweling, S., and Aben, I.: Reconstructing and quantifying methane emissions from the full duration of a 38-day natural gas well blowout using space-based observations, Remote Sensing of Environment, 270, https://doi.org/10.1016/j.rse.2021.112755, 2022a.

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, Sci. Adv, 8, 9683, https://doi.org/10.1126/sciadv.abn9683, 2022b.

MacLean, J.-P., Jervis, D., McKeever, J., Ramier, A., Strupler, M., Tarrant, E., and Young, D.: Detecting and quantifying methane emissions with the high-resolution GHGSat satellite constellation, in: AGU Fall Meeting Abstracts, vol. 2021, pp. A54F–01, 2021.

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., and Eichmann, A.: The GEOS-5 Atmospheric General Circulation Model : Mean Climate and Development from MERRA to Fortuna, Technical Report Series on Global Modeling and Data Assimilation, Tech. Rep. April, 2012.

Nisbet, E. G., Fisher, R. E., Lowry, D., France, J. L., Allen, G., Bakkaloglu, S., Broderick, T. J., Cain, M., Coleman, M., Fernandez, J., Forster, G., Griffiths, P. T., Iverach, C. P., Kelly, B. F., Manning, M. R., Nisbet-Jones, P. B., Pyle, J. A., Townsend-Small, A., al Shalaan, A., Warwick, N., and Zazzeri, G.: Methane Mitigation: Methods to Reduce Emissions, on the Path to the Paris Agreement, Reviews of Geophysics, 58, https://doi.org/10.1029/2019RG000675, 2020.

NOAA: Trends in Atmospheric Methane, https://gml.noaa.gov/ccgg/trends_ch4/, 2022.

Ocko, I. B., Naik, V., and Paynter, D.: Rapid and reliable assessment of methane impacts on climate, Atmospheric Chemistry and Physics, 18, 15 555–15 568, https://doi.org/10.5194/acp-18-15555-2018, 2018.

Ocko, I. B., Sun, T., Shindell, D., Oppenheimer, M., Hristov, A. N., Pacala, S. W., Mauzerall, D. L., Xu, Y., and Hamburg, S. P.: Acting rapidly to deploy readily available methane mitigation measures by sector can immediately slow global warming, Environmental Research Letters, 16, https://doi.org/10.1088/1748-9326/abf9c8, 2021.

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al.: KerasTuner, https://github.com/keras-team/keras-tuner, 2019.

Pandey, S., Gautam, R., Houweling, S., Denier Van Der Gon, H., Sadavarte, P., Borsdorff, T., Hasekamp, O., Landgraf, J., Tol, P., Van Kempen, T., Hoogeveen, R., Van Hees, R., Hamburg, S. P., Maasakkers, J. D., and Aben, I.: Satellite observations reveal extreme methane leakage from a natural gas well blowout, https://doi.org/10.1073/pnas.1908712116, 2019.

Pandey, S., Houweling, S., Lorente, A., Borsdorff, T., Tsivlidou, M., Anthony Bloom, A., Poulter, B., Zhang, Z., and Aben, I.: Using satellite data to identify the methane emission controls of South Sudan's wetlands, Biogeosciences, 18, 557–572, https://doi.org/10.5194/bg-18-557-2021, 2021.

Parry, I., Black, S., Minnett, D., Mylonas, V., and Vernon, N.: How to Cut Methane Emissions; IMF Staff Climate Note 2022/008. October 2022, Tech. rep., International Monetary Fund, 2022.

Plant, G., Kort, E. A., Brandt, A. R., Chen, Y., Fordice, G., Gorchov Negron, A. M., Schwietzke, S., Smith, M., and Zavala-Araiza, D.: Inefficient and unlit natural gas flares both emit large quantities of methane, Science, 377, 1566–1571, https://doi.org/https://doi.org/10.1126/science.abq0385, 2022.

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

Ramier, A., Deglint, H., Gains, D., Jervis, D., McKeever, J., Shaw, W., Strupler, M., Tarrant, E., and Varon, D. J.: GHGSat-C1-Initial Results, Design, and Characterization, in: AGU Fall Meeting Abstracts, vol. 2020, pp. A247–03, 2020.

Sadavarte, P., Pandey, S., Maasakkers, J. D., Lorente, A., Borsdorff, T., Denier van der Gon, H., Houweling, S., and Aben, I.: Methane Emissions from Superemitting Coal Mines in Australia Quantified Using TROPOMI Satellite Observations, Environmental Science and Technology, 55, 16 573–16 580, https://doi.org/10.1021/acs.est.1c03976, 2021.

Sánchez-García, E., Gorroño, J., Irakulis-Loitxate, I., Varon, D. J., and Guanter, L.: Mapping methane plumes at very high spatial resolution with the WorldView-3 satellite, Atmospheric Measurement Techniques, 15, 1657–1674, https://doi.org/10.5194/amt-15-1657-2022, 2022.

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

Scarpelli, T. R., Jacob, D. J., Grossman, S., Lu, X., Qu, Z., Sulprizio, M. P., Zhang, Y., Reuland, F., Gordon, D., and Worden, J. R.: Updated Global Fuel Exploitation Inventory (GFEI) for methane emissions from the oil, gas, and coal sectors: Evaluation with inversions of atmospheric methane observations, Atmospheric Chemistry and Physics, 22, 3235–3249, https://doi.org/10.5194/acp-22-3235-2022, 2022.

Schneising, O., Buchwitz, M., Reuter, M., Vanselow, S., Bovensmann, H., and P. Burrows, J.: Remote sensing of methane leakage from natural gas and petroleum systems revisited, Atmospheric Chemistry and Physics, 20, 9169–9182, https://doi.org/10.5194/acp-20-9169-2020, 2020.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X.: DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN, ACM Transactions on Database Systems, 42, https://doi.org/10.1145/3068335, 2017.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, International Journal of Computer Vision, 128, 336–359, https://doi.org/10.1007/s11263-019-01228-7, 2020.

820   Shaw, J. T., Allen, G., Barker, P., Pitt, J. R., Pasternak, D., Bauguitte, S. J., Lee, J., Bower, K. N., Daly, M. C., Lunt, M. F., Ganesan, A. L., Vaughan, A. R., Chibesakunda, F., Lambakasa, M., Fisher, R. E., France, J. L., Lowry, D., Palmer, P. I., Metzger, S., Parker, R. J., Gedney, N., Bateson, P., Cain, M., Lorente, A., Borsdorff, T., and Nisbet, E. G.: Large Methane Emission Fluxes Observed From Tropical Wetlands in Zambia, Global Biogeochemical Cycles, 36, https://doi.org/10.1029/2021GB007261, 2022.

Shen, L., Gautam, R., Omara, M., Zavala-Araiza, D., Maasakkers, J. D., Scarpelli, T. R., Lorente, A., Lyon, D., Sheng, J., Varon, D. J.,
825   Nesser, H., Qu, Z., Lu, X., Sulprizio, M. P., Hamburg, S. P., and Jacob, D. J.: Satellite quantification of oil and natural gas methane emissions in the US and Canada including contributions from individual basins, Atmospheric Chemistry and Physics, 22, 11 203–11 215, https://doi.org/10.5194/acp-22-11203-2022, 2022.

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

830   Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, https://doi.org/10.48550/ARXIV.1409.1556, 2014.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., and Barker, D. M.: A description of the advanced research WRF model version 4, Tech. rep., National Center for Atmospheric Research: Boulder, CO, USA, doi:10.5065/1dfh-6p97, 2019.

835   Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research, 15, 1929–1958, 2014.

SRON CH4 L2 team: SRON RemoTeC-S5P scientific XCH4 data product Product User Guide - version 18_17, Tech. rep., SRON The Netherlands Institute for Space Research, https://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/ch4/SRON_CH4_product_Jan2022. pdf, 2022.

840   Szopa, S., Naik, V., Adhikary, B., Artaxo, P., Berntsen, T., Collins, W., Fuzzi, S., Gallardo, L., Kiendler-Scharr, A., Klimont, Z., Liao, H., Unger, N., and Zanis, P.: Short-Lived Climate Forcers, p. 817–922, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.008, 2021.

Tu, Q., Hase, F., Schneider, M., Garciá, O., Blumenstock, T., Borsdorff, T., Frey, M., Khosrawi, F., Lorente, A., Alberti, C., Bustos, J. J., Butz, A., Carreño, V., Cuevas, E., Curcoll, R., Diekmann, C. J., Dubravica, D., Ertl, B., Estruch, C., León-Luis, S. F., Marrero, C., Morgui, J. A.,
845   Ramos, R., Scharun, C., Schneider, C., Sepúlveda, E., Toledano, C., and Torres, C.: Quantification of CH4 emissions from waste disposal sites near the city of Madrid using ground- and space-based observations of COCCON, TROPOMI and IASI, Atmospheric Chemistry and Physics, 22, 295–317, https://doi.org/10.5194/acp-22-295-2022, 2022a.

Tu, Q., Schneider, M., Hase, F., Khosrawi, F., Ertl, B., Necki, J., Dubravica, D., Diekmann, C. J., Blumenstock, T., and Fang, D.: Quantifying CH4 emissions in hard coal mines from TROPOMI and IASI observations using the wind-assigned anomaly method, Atmospheric
850   Chemistry and Physics, 22, 9747–9765, https://doi.org/10.5194/acp-22-9747-2022, 2022b.

UNEP and CCAC: Global Methane Assessment: Benefits and Costs of Mitigating Methane Emissions., Tech. rep., Nairobi: United Nations Environment Programme., iSBN: 978-92-807-3854-4, 2021.

Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., Loibl, D., Hellwich, O., and Walter, T. R.: Towards global volcano monitoring using multisensor sentinel missions and artificial intelligence: The MOUNTS monitoring system, Remote Sensing, 11, 1–31,
855   https://doi.org/10.3390/rs11131528, 2019.

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, Atmos. Meas. Tech, 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.

Varon, D. J., McKeever, J., Jervis, D., Maasakkers, J. D., Pandey, S., Houweling, S., Aben, I., Scarpelli, T., and Jacob, D. J.: Satellite

860   Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production, Geophysical Research Letters, 46, 13 507–13 516, https://doi.org/10.1029/2019GL083798, 2019.

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

865 Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Yu, J., Hmiel, B., Lyon, D. R., Warren, J., Cusworth, D. H., Duren, R. M., Chen, Y., Murphy, E. C., and Brandt, A. R.:

870   Methane Emissions from Natural Gas Gathering Pipelines in the Permian Basin, Environmental Science and Technology Letters, https://doi.org/10.1021/acs.estlett.2c00380, 2022.

Zavala-Araiza, D., Lyon, D., Ramo´n, R., Alvarez, A., Palacios, V., Harriss, R., Lan, X., Talbot, R., and Hamburg, S. P.: Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites, Environ. Sci. Technol, 49, 8167–8174, https://doi.org/10.1021/acs.est.5b00133, 2015.

875 Zhang, Y., Gautam, R., Pandey, S., Omara, M., Maasakkers, J. D., Sadavarte, P., Lyon, D., Nesser, H., Sulprizio, M. P., Varon, D. J., Zhang, R., Houweling, S., Zavala-Araiza, D., Alvarez, R. A., Lorente, A., Hamburg, S. P., Aben, I., and Jacob, D. J.: Quantifying methane emissions from the largest oil-producing basin in the United States from space, Science Advances, 6, 1–10, https://doi.org/10.1126/sciadv.aaz5120, 2020.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A.: Learning Deep Features for Discriminative Localization, http://arxiv.org/

880   abs/1512.04150, 2015.