https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

U-Net Semantic Segmentation of Sentinel-2 Imagery for Landfill Detection and European Inventory Validation

Emma Julie Schoenmakers[1,2], Abdullah Al Nayeem[2,a], Anchal Chawla[2,b], Hugo Denier van der Gon[2], Sander Houweling[1,3]

[1]Earth Sciences Department, Vrije Universiteit, Amsterdam, 1081HV, Netherlands
[2]TNO, Netherlands Organisation for Applied Scientific Research, Utrecht, 3584CB, Netherlands
[3]SRON, Space Research Organisation Netherlands, Leiden, 2333AL, Netherlands

[a]Now at: Spatial Health Lab, College of Health, Oregon State University, Corvallis, OR 97331, United States
[b]Now at: D2Win, Digital Twin | Spatial Specialisten, Nieuwegein, 3436 ZZ, Netherlands

Correspondence to: Emma Schoenmakers (e.j.schoenmakers@vu.nl)

Abstract. The waste sector accounted for 19 % of the global anthropogenic methane emissions in the previous decade, according to emission inventories. To evaluate these emission inventories, correctly located facilities are necessary. The quality of these locations affects downstream applications, such as atmospheric transport modelling and satellite tasking for emissions monitoring. Despite this importance, comprehensive knowledge of locations of solid waste disposal (landfills) sites across the European continent remains limited. This study introduces a method to investigate the locations and spatial extent of landfills. To do so, we use a U-Net model, trained on Sentinel-2 satellite bands using a limited manually labelled dataset of European landfill and non-landfill sites in order to perform pixel-level segmentation and a corresponding image-level classification. When tested onto 873 methane-reporting landfill sites from the TNO-GHGco inventory, largely based on the European Industrial Emissions Portal Regulation (IEPR) reporting, our model had a 79 % accuracy in detecting correct landfill locations and a 92 % recall, indicating that few true landfill locations were missed. Notably, this study detected a mismatch between reported coordinates and imagery-detected extent for at least 51 landfills, here representing ~6 % of the data, suggesting that these erroneous landfill locations could be corrected in the official IEPR register. By correctly segmenting and confirming more than 638 landfills, 462 more than the traditional CORINE land cover map, this method offers an avenue for routine validation of reported landfill locations from space. However, false positive rates associated with misreporting were high in France, likely driven by red-edge similarities between landfills and croplands, though this effect may extend beyond this region. The model exhibits a moderate Matthews Correlation Coefficient (MCC) of 0.25, due, in parts, to the number of false positives likely driven by spectral overlap and indicating the U-Net performs best at identifying correctly reported landfills. These findings underline the importance and challenge of refining landfill location data for European methane reporting and of routinely detecting both correctly reported locations and misreporting to minimise (future) satellite mistargeting. The U-

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Net algorithm we present fills a critical monitoring gap, paving the way for methane source attribution, activity monitoring and infrastructure location validation through satellites to support subsequent methane measurement campaigns.

1 Introduction

Global methane (CH_4) emissions have increased by 15-20 % over the past 20 years with the waste sector contributing 15 Tg CH_4 yr^{-1} to that rise (Jackson et al., 2024). In the European Union, the waste sector, including wastewater treatment processes, solid waste disposal and incineration, was the second largest contributor (24 %) to total reported EU methane emissions in 2024 after agriculture (58 %) and preceding the energy sector (8 %) (EEA, 2026a). To reach the EU's commitment to reduce methane by 30 % by 2030, the implementation of the EU regulation on reducing methane in the energy sector (EU/2024/1787) has been introduced. However, reaching this 30 % target will also inevitably require significant reductions in methane emissions from the waste sector. New regulations based on the EU Methane Strategy (COM/2020/663) support this direction through technological innovations, such as satellite detection, biogas development or an emphasis on the diversion of organic waste and enforcement with penalties based on the associated EU Methane Regulation (2024/1787). European methane emissions from solid waste landfills, accounting for the majority of the reported waste sector's methane emissions (66 %), are estimated to have been declining by 36 % between 1990 and 2024 thanks in parts to the Landfill Directive (1999/31/EC, revised Directive (EU) 2018/850) and national efforts, particularly in the Netherlands, Germany and Sweden (EEA, 2026b). However, this trend masks a slowdown in annual methane emission reduction from landfills which have stabilised near 2400 kt CH_4 since 2016, according to the latest EU greenhouse gas inventory (EEA, 2026b). This poses a new challenge for targeting landfill methane reduction further.

As per the Landfill Directive, a landfill refers to a managed waste 'disposal site for the deposit of waste onto or into land' such as a 'permanent site, of more than a year', or an 'internal waste disposal sites', for example from a producer disposing of their own waste (1999/31/EC, revised Directive (EU) 2018/850). It excludes sites that are used for unloading waste for preparation, disposal, recovery or treatment (1999/31/EC, revised Directive (EU) 2018/850). As of 2024, the percentage of total municipal waste being landfilled in the EU-27 (24 %) shows notable differences across countries but is prominent in countries such as Romania (80 %) and Greece (81 %), where limited incineration, recycling and composting capacity often causes municipal waste to end up in landfills (Eurostat, 2026). This makes detecting landfills an important and complex issue to tackle for countries with highly varying degrees of waste composition and infrastructure for waste processing across Europe.

Satellites offer a new avenue for assessment of methane emissions from space (Varon et al., 2018). This ranges from detecting CH_4 from coal mines (Sadavarte et al., 2021), oil and gas sites (Naus et al., 2023) or landfills (Maasakkers et al., 2022). Methane emissions from landfills are mostly due to disposal of organic waste, including food waste and paper onto landfills

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

over years. There is a strong dependence of methane emissions on, but not limited to, the microorganisms present in the landfill (Grégoire et al., 2023), the quantity and type of organic waste (Eleazer et al., 1997), the size and structure of landfills and their individual cells (Scharff et al., 2024). An additional parameter is the methane oxidation capacity which is related to the topsoil or landfill cover and parameters such as soil temperature, composition, moisture, pH and nutrient content (Rachor et al., 2013). Due to the highly localised spatial patterns of the landfill, coal and oil and gas sectors, future satellites operating at high spatial resolution are currently in development, namely the CO2M mission (~2 x 2 km) or the point-source mapper ESA TANGO (~300 x 300 m) mission. However, the trade off from many high-resolution Earth Observation (EO) satellites is a lower revisit time than their daily mapper counterparts like Sentinel-5p (TROPOMI). As such, the efficient targeting and tasking of known facilities is a vital part of both current and upcoming point source detector satellites. This means that prior knowledge of facility location at high precision local scale but also with a comprehensive coverage is crucial to any future greenhouse gas (GHG) satellite mission becoming a success.

Facility location detection and segmentation is a relatively understudied field within open-science with many datasets utilising 'bottom-up' national emission inventories at the point source scale as a key resource to support upcoming GHG satellite tasking, particularly for methane. However, some efforts have been developed to open-source facility scale information, namely the citizen science based OpenStreetMap (https://www.openstreetmap.org/) or inventory-based Waste Map (https://wastemap.earth/) to name a few. Beyond not-for-profit initiatives, research has been performed on the 'top-down' classification of Sentinel-2 imagery, particularly encouraged by its sub-100m spatial resolution. Applications have ranged from coal mine detection using pre-trained Convolutional Neural Network (CNN) architectures (Madhuanand et al., 2021) to general land cover classification tasks on a global scale (Zanaga et al., 2022). However, little has been investigated for the landfill sector, particularly in Europe. A recent study by Dogniaux et al. (2026) used Sentinel-2 imagery to identify active landfill clusters using changes in their structures as an indicator of landfilling activity. Regarding waste segmentation, most of the focus tends to be towards plastic waste instead, such as through the use of hyperspectral data (Bouchelaghem et al., 2026).

Landfills represent a major source of methane with a large mitigation potential (Scharff et al. 2024). A better understanding of landfill facility-scale reporting in Europe is thus needed as inaccurate location data directly limits inventory assessments, the ability to target mitigation measures and the monitoring of their effectiveness. This work aims to (i) provide a methodology for the validation of solid waste landfill inventory locations across Europe at image-scale and (ii) understand some of the particularities of landfills at pixel-scale and their reporting throughout the continent and (iii) identify the potential for improvements of the methodology to futureproof it for the increasing use of incoming satellite methane data.

We therefore developed a U-Net model to perform landfill segmentation at pixel-level and classification at image-level to investigate whether this methodology allows us to validate and/or flag lacking landfill facility reporting across Europe. This is done by leveraging a training set (Sect. 2.1) based on both model and satellite measurements from landfills all over Europe and testing it on the TNO-GHGco inventory, an emission inventory based on IEPR (Sect. 2.6). The methodological

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

contribution, in addition to the use of a U-Net (Sect. 2.4), focuses on the sampling strategy, novel train set construction and ablation studies (Sect. 4.2) suited to the landfill detection and segmentation task.

2 Methods

For this study, we adopt a standard U-Net architecture, chosen for its proven performance under limited labelled data conditions. The following sections will introduce the strategy to build the landfill segmentation model, starting with an explanation of the manually labelled training set based on open-source landfill location data (Sect. 2.1). The next section will explain the validation set (Sect. 2.2) used for parameter tuning. Thereafter the most appropriate loss function will be investigated (Sect 2.3). In Section 2.4, the U-Net model is introduced in detail whilst the application to the test set based on the emission inventory locations, the inference, is explained in Sections 2.5 and 2.6. Finally, the metrics used to evaluate the performance of the model are introduced in Section 2.7, including a comparison with the more classical land cover dataset CORINE (CORINE, 2020).

2.1 Training Set

Solid waste landfill facilities were used to train the U-Net. Their locations were based on Climate TRACE solid waste facilities, which draws on Global Waste Map, IEPR and OpenStreetMap for Europe (Climate TRACE, 2023). Carbon Mapper detections of methane emissions over landfills were also used, mainly from NASA EMIT at the time of the study (Duren et al., 2025). These locations (Fig. 1) were selected to obtain a training set that supports applications on facilities that emit satellite-detectable amounts of methane. The 99 locations were then exported through Google Earth Engine as Sentinel-2 Level-2A .tif files from the years 2022-2024 (matching the respective years of Climate TRACE or Carbon Mapper detections) after drawing a variable bounding box of approximately 300 m East and West of the landfill coordinates, depending on the landfill size. This was done to add background information to contrast with the landfill itself. In addition, we introduced random locations of 'background' imagery (non-landfill) comprising of 29 locations all over Europe of urban, agricultural, coastal, forested and mixed classes to increase the robustness of our model in preventing false positives. The focus of the background class was on representing the European landscape diversity rather than the intra-class diversity as it was assumed that different landscapes such as bare soil or mountainous areas would be more likely to trigger false positives rather than adding more urban sites in one country for example. As such, our training set consisted of both Sentinel-2 imagery of a landfill or background and its corresponding label. To introduce diversity, images were kept at their original different resolutions and image sizes to prevent overfitting to a specific image resolution in the training set. The bands span a range of spatial resolutions from 10m to 60m, thus capturing fine-grained surface detail and broader spectral context. Our manual sample set therefore comprises of 99 sites across 23 countries, each site represented by all 12 Level-2A Sentinel-2 spectral bands [442-2202 nm]. The use of 12 bands

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

introduces important inter-band correlations which are crucial for landfill detection such as albedo or vegetation relationships (Papale et al., 2023).

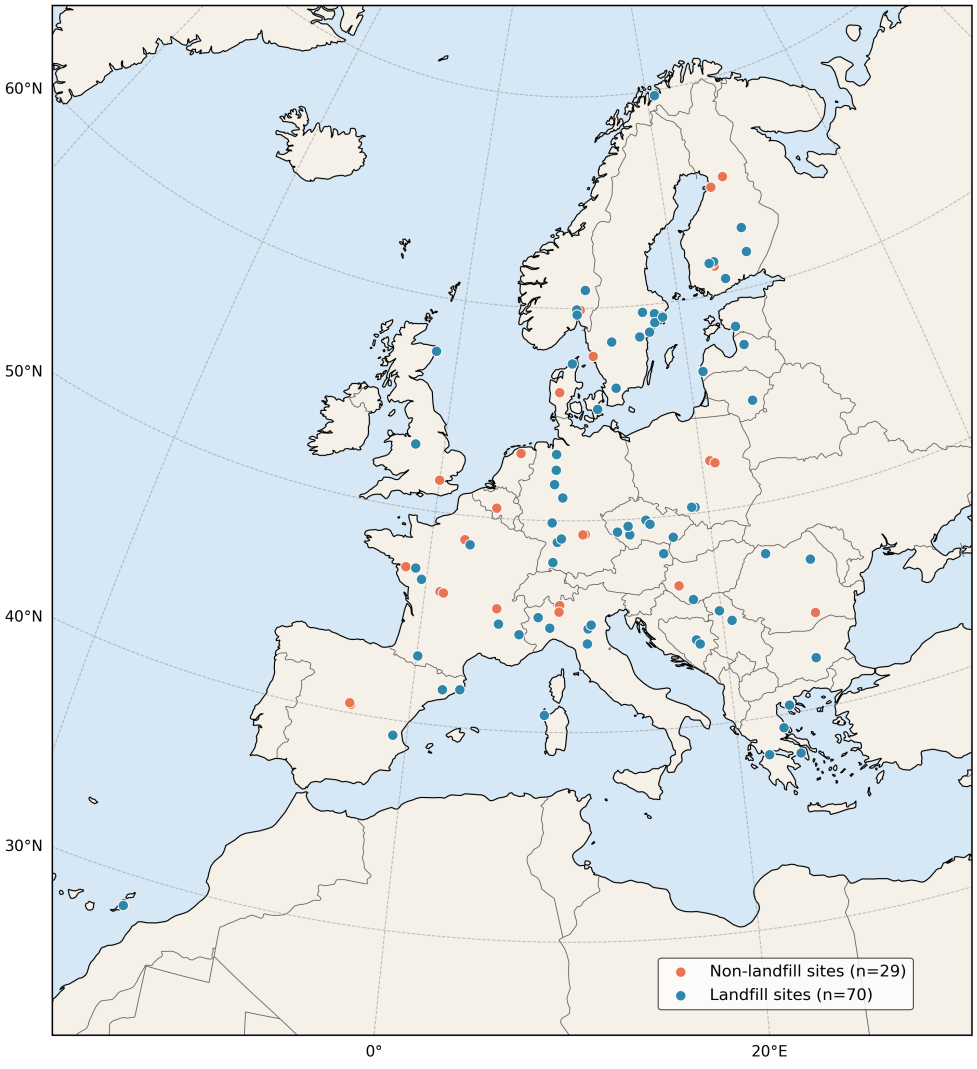


Figure 1. Sentinel-2 landfill and 'background' non-landfill locations manually sampled over Europe, made with Natural Earth.

Once our imagery was downloaded, we then applied a robust min-max normalisation using the 2nd and 98th percentile per band to preserve the individual statistical distribution of each band (Eq.1).

$$X_{\text{scaled,i,j,b}} = \frac{X_{\text{i,j,b}} - P_{2,\text{b}}}{P_{98,\text{b}} - P_{2,\text{b}}} \quad (1)$$

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

where $X_{i,j,b}$ is the pixel value at row i, column j in a specific band b, and $P_{2,b}$ and $P_{98,b}$ are the 2nd and 98th percentile values of each band b.Values falling outside the 2nd percentile or 98th percentile are then clipped to 0 or 1 respectively.

This method caps extreme values often present in earth observation data, such as water body or urban rooftop reflectance, but preserves the original distribution whilst removing the sensitivity to outliers. This normalisation thus provides inputs between [0,1] and will allow the convolutional filters in the U-Net to train more stably and prevent exploding gradients (Glorot et al., 2010). In QGIS, we also manually created and exported the corresponding pixel mask, also called label, for both landfill (1 label) and background classes (0 label). The per-pixel label focused on the active section of the landfill, defined as cells with active waste dumping activity and including leachate ponds, typically seen in the Nordics and Germany in our training set. Hence the vegetation covered areas, which typically look green, or areas with no active waste dumped, were not encapsulated in the labels. For both the training and validation set, we start with the afore-mentioned normalised image and corresponding label, one-hot encode the labels and start the sampling process. Regarding the training set sampling strategy for the U-Net, we used random image and patch selection from the dataset at each epoch so that the model sees a different combination of an image 'piece' (the patch) at each step. Each single patch represented information across all 12 Sentinel-2 bands. This aims to improve generalisation by introducing spatial diversity but at the detriment of complete coverage of patches per epoch. For the U-Net training sampling strategy, this is called random whilst the validation set uses a different approach allowing for reproducibility, referred to as deterministic patch sampling.

2.2 Validation Set

For the validation set, we held out 40 % of images from our training data with their corresponding labels. On average the ratio of landfill to non-landfill images was 2:1 within the validation and training sets. The non-landfill and landfill difference in composition between the training and validation set was 5 % on average. This is computed by taking the proportion of landfill labels (and background) with imagery for each set and compared with the absolute difference between sets. Composition difference was always checked to be under 15 % to prevent class imbalance. Based on this validation set we then changed the model parameters to derive the best outcomes of the following metrics: Intersection over Union (IoU) mean, class-specific IoU, accuracy, class-specific F1, loss, and qualitative expert visualisation for both training and validation. Thanks to our manually created labels, we were able to determine the effectiveness of our sampling strategy, model parameters or model splits using this validation set. A summary of the workflow for the model procedure can be found in the Appendix (Fig. A1). As previously mentioned, we use a deterministic patch sampling approach to the validation set. This means that every patch is seen exactly once per epoch always at the same time. We use a sliding window by tracking the top left corner of the image's patch currently being processed and scan to the right and then down through the next row in the image. This approach ensures result comparability between different model runs.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.3 Loss Function

Loss functions were tested to address class imbalance within images of the training and validation set as labelled data had an average of 4 % landfill pixels per image. As such, we did denial experiments including Dice loss (Sudre et al., 2017), Focal loss (Lin et al., 2018), Cross-Entropy (CE) loss (Zhang and Sabuncu, 2018) and a Combined Cross-Entropy loss incorporating tuned weights from Dice Loss and CE loss (Wang and Li, 2021). We found that Cross-Entropy loss yielded the best segmentation results both visually, with a higher number of True Positives in the validation set predictions compared to other loss functions, and a higher overall train and validation accuracy ranging from 7-25 % improved performance in train accuracy and 2-4 % improvement in validation accuracy by the end of the training. The CE loss also yielded higher Intersection over Union (Eq. 2) results for landfills ranging from a 5 % improvement from Focal loss to 12 % improvement from the Combined CE by the end of the training for example.

2.4 U-Net Architecture

For the purpose of this study, we adopt a standard U-Net architecture (Fig. 2), initially developed for biomedical image segmentation and chosen to handle the lack of large, labelled training datasets (Ronneberger et al., 2015). This characteristic U-shaped architecture can be described as a series of convolutional neural network layers with the aim to localise fine features in an image.

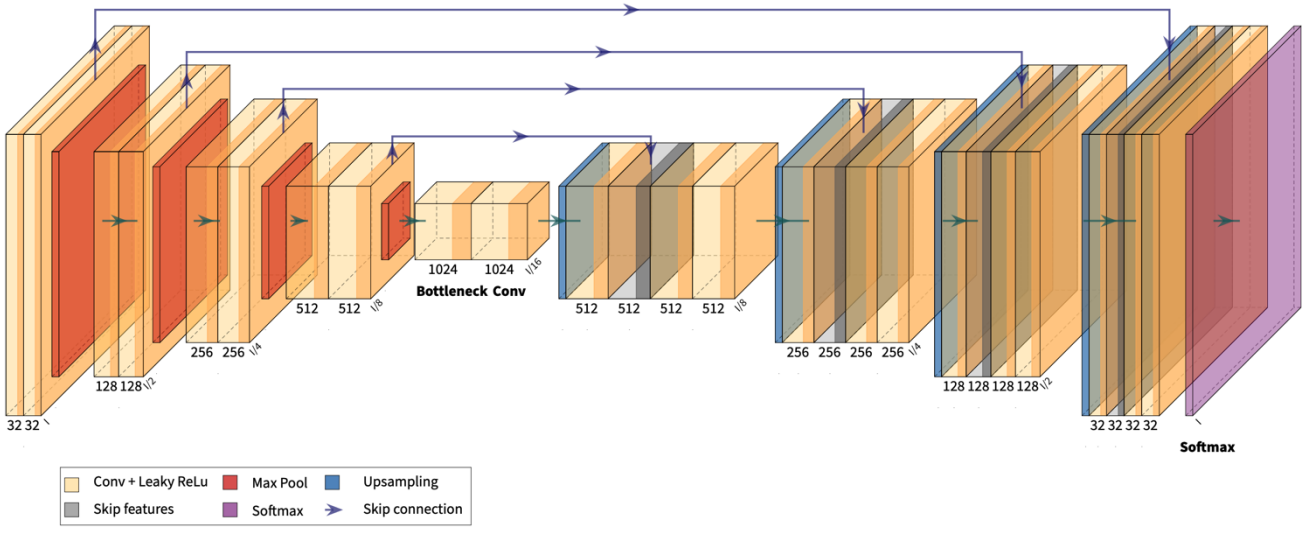


Figure 2. U-Net architecture used for our landfill segmentation model. The skip connection represented by blue arrows represent the flow of the propagated context information from one respective contracting path (left) layer to its equivalent in the upsampling layer in the expansive path (right), made with PlotNeuralNet by Haris Iqbal.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

For a U-Net specifically, this is done by decreasing the resolution at each layer, taking into account the propagated context information (the contracting path, left side), and reconstructing the input using the upsampling operators (the expansive path, right side). As both the contracting and expansive path are symmetric, they yield this characteristic U-shaped network. Each step in our network follows a classic U-Net architecture with 4 layers on both the contracting and the expansive path. The contracting path uses a succession of convolutional layers along with pooling operations leading to a progressively smaller spatial resolution but a larger number of channels. As the number of channels increases, this allows the network to capture a richer set of features and vice versa. At every block a set of convolutional filters slides across the input grid producing feature maps that capture important features pertaining to a landfill (or background), the granularity of which is defined by the number of the afore-mentioned channels. Each convolutional layer uses a Leaky Rectified Linear Unit (Leaky ReLu) activation function. This is introduced to offer better generalisation than the Rectified Linear Unit (ReLu) and mitigate the issues of negative gradients leading to unused 'dead neurons' often found in ReLu activation functions (Maas et al., 2013). The expansive path uses the upsampling of the feature map and is followed by a 2x2 convolution, thus halving the number of feature channels per block. The expansive path also contains a concatenation with the corresponding cropped feature map taken from the contracting path's respective layers via skip connections. The last layer of our U-Net uses a two-channel softmax activation function, converting the outputs from the previous layer to explicit per-pixel probabilities between 0 and 1 representing the probability of a pixel belonging either to the background or the landfill class. This formulation was chosen over a single channel sigmoid output to directly produce probability estimates at inference for each class and offers the possibility of extension to multi-class problems. Parameters for our U-Net including learning rates, batch sizes, patch sizes and loss functions weightings were trialled using both hyperparameter tuning using the Python Optuna package and by doing ablation studies (Sect. 2.3.) also supported by our validation set, as described in Sect. 2.2. The parameters used for our model are summarised in the table below (Table 1.)

Table 1. U-Net model parameters optimised using the validation set.

Parameter	**Value**
Patch size	64
Batch size	16
Learning rate	10^{-3}
Loss Function	Cross-Entropy (no class weighting)
Train/Val split	60/40

As seen in Fig.2, we use 32 channels in the first block and its corresponding upsampling block for a faster computation time, compared to the traditional 64 in Ronneberger et al. (2015). With 59 images in the aforementioned training set and 40 in the

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

validation set (approximately 60/40 train/val split), with each set itself divided into patches, one epoch consisted of 1,130 steps (batches) across all 12 bands. Since each training step draws on a fresh, random patch per sample, approximately 18,000 patch samples are drawn per epoch (1130 x 16), meaning that some patches may be revisited with slight positional variation in the same epoch.

2.5 Inference

After our custom U-Net model has been trained, the model is applied to the full Sentinel-2 imagery per facility, to classify all the pixels as either background (0) or landfill (1). This is called the inference step. Each image is divided into the same patch size as in our training but using a sliding window with 50 % overlap. This technique aims to allow for a smoother transition between each patch and prevent boundary effects in the segmentation. This sliding window goes from left-to-right and top-to-bottom. Taking the probabilities computed by the U-Net's final softmax layer, as mentioned in Sect. 2.4, the inference step accumulates softmax probabilities from all overlapping patches at the per-pixel level. The class with the highest accumulated score will be converted into the final label. Since patches will have overlap, the probabilities will receive contributions from multiple patches, thus smoothing out boundary artifacts. As such, inference is the last step in the segmentation pipeline and forms the basis for testing the model on 'unseen data'.

2.6 Landfill Facility Test Set over Europe

The European landfill locations were taken from the TNO greenhouse gas and co-emitted species (GHGco) inventory (TNO-GHGco v5). This is an updated version of the TNO-GHGco_v4 developed for the EU-horizon project CoCO2 (Denier van der Gon et al., 2022). The TNO-GHGco inventory focuses on Europe and follows the methodology as outlined in Kuenen et al. (2022) but point sources are given at exact location. The point source locations are largely based on reporting to the Industrial Emissions Portal Regulation (IEPR, formerly E-PRTR) (EEA, 2026c) and supplemented by other facilities in a minority of cases (non-EU countries). 94 % of test locations are geographically distinct from training locations with the remaining 6 % (n=20) falling within 500m of a training point, due to the use of IEPR as a supporting dataset in Climate TRACE and Carbon Mapper data.

E-PRTR is a public database for individual facilities that release pollutants to the air, water and land as well as off-site transfers. This includes sectors like the chemical industry, aquaculture, energy production and waste and wastewater management sector. For the wastewater and solid waste sector these facilities must quantify and report these releases every year, provided these are above the threshold for reporting. For the landfill waste sector specifically, facilities must report if they receive more than 10 tonnes of waste per day and/or have a total capacity of more than 25,000 tonnes (European Commission, 2006). For solid waste landfills, this falls under category 5(d) in E-PRTR and usually comes with reporting of CO_2, CH_4 and air pollutant

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

emissions such as Non-Methane Volatile Organic Compounds (NMVOCs), NO_x or heavy metals (any of the 91 listed pollutants). For methane specifically, once the tonnage threshold is met, a landfill must report if it is estimated that the facility emits more than 100,000 kg year^{-1} of CH_4 (European Commission, 2006). Landfills are exempt from E-PRTR reporting if they closed before 16 July 2001 or accept only inert waste such as concrete or sand. Landfills are also exempt if they are in the after-care phase, as defined in Article 13 of the Landfill Directive, meaning they are closed but still under monitoring by the 'competent authorities' in partnership with the operator (European Commission, 2006). This is particularly relevant for the detection of landfills, as in practice, some landfills report under the methane threshold, or sometimes closed landfills are still reported under E-PRTR for decades. E-PRTR does not only collect facility-reported emissions but also associated metadata such as company name and facility location. Typically, companies send their facility-scale emissions reporting to the 'competent authority', which is usually a national environmental authority (European Commission, 2006). This information is compiled and checked at the national scale, sent over to the European Commission and with the support of the European Environment Agency (EEA), submitted to E-PRTR.

The ongoing transition from the current E-PRTR system to the IEPR system in 2028 offers opportunities for both improved reporting and enforcement. Indeed, the IEPR system now shifts from facility scale reporting to installation-level and makes use of the revised Industrial Emissions Directive (IED 2.0) as the main legal instrument for implementation of more than 50,000 reporting industrial facilities, including landfill facilities (European Commission, 2024). Some examples of new compliance enforcements include an environmental inspection on site every three years by 'competent authorities', enhanced financial fines for infringements and allowing people to seek compensation in the case of damages to their health due to illegal pollution. For the sake of consistency, we refer to the acronym IEPR to designate both past E-PRTR and upcoming IEPR data.

Using the coordinates from this TNO-GHGco dataset, we used the Google Earth Engine python API to download Sentinel-2 imagery in a 1 km circular buffer around the reporting coordinates and exported the imagery over its bounding box (approximately 2x2 km). The Sentinel-2 Level-2A imagery was downloaded as a median composite in 2018 to match the inventory year present in our test set and allow for enough cloud-free (<20 % cloud cover) images to be taken at all sites. After data cleaning, including of duplicate data, the test set represents 873 locations for facilities that reported methane emissions over 100 tons per year and with either a daily intake of 10 tonnes of waste per day or with a total capacity of 250 000 tonnes. In practice, our unseen test set consists of solid waste landfill locations in 29 countries with the majority of facility entries from the United Kingdom (n=195), France (n=137), Spain (n=94) and the least in Montenegro (n=2), Estonia (n=2) and Malta (n=2). As such, our unseen test set represents a greater number of countries than our training set to provide a better assessment of our model's generalisation capability to a variety of countries' respective landfill infrastructure, topography and landscape diversity. The model is therefore run on the 873 locations and gives an output with the location id, coordinates, country, prediction classification and landfill percentage to determine the area size of the landfill itself. In addition, it plots the original

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Sentinel-2 RGB imagery along with the segmentation mask of where it determines the landfill to be located and its size, or an empty mask if it doesn't detect any.

2.7 Model Evaluation

To identify the quality of the inventory, this model outputs segmentation imagery, statistics and classification outputs in a csv. Since we do not have ground truth data on this unseen test set, the confusion matrix can only be established using a manual check on the model predictions visually. This means that landfills that are classified as landfills in an image of the inventory location get assigned a True label if they visually look like landfills and vice versa. This includes visual checks on landfill structure, presence of gas pipes, leachate ponds, or, in tough cases, using Google Street View and even local Google Reviews at site locations. As this process is costly, it is only done to check the model's capability on the test set once. All other iterations on the test set are then compared to this 'ground truth'. On the other hand, landfills that are classified by the U-Net as background in an inventory location get assigned a True label if they indeed do not look like a landfill, using the same method. The case with a correctly detected landfill will thus be considered a True Positive (TP) whilst the correctly detected background case would be a True Negative (TN). For False Positives (FP), this would occur if it were predicted that there is a landfill when a human check in fact does not agree and sees background. In our methodology, the human decision and independent thorough checking supersede the U-Net model. For False Negatives (FN), this would occur if it were predicted that there is a background when a human check does not agree and sees a landfill. These FNs and FPs are therefore a grey area where either the model has failed at detecting a landfill or background (FNs) or confuses it with another class (FPs). Lastly, we estimated the contribution of each Sentinel-2 band to the landfill segmentation by computing the sensitivity (gradient) of the model's landfill predictions to each input band, restricted to pixels labelled as landfill in the ground-truth masks of the validation sequence (Simonyan et al., 2014).

To further evaluate our model and demonstrate the added value of the U-Net system compared to more classical data evaluation techniques, this study also uses the CORINE Land Cover European raster for the year 2018 (v2020_20_u1) at 100m resolution (CORINE, 2020). The dataset itself is based on the classification of satellite images including Sentinel-2, Sentinel-1 and gapfilling with Landsat 8 by all EEA members and cooperating countries (CORINE, 2020). Taking the artificial surface class 132, 'dumpsite', as there is no landfill class in the dataset, the polygons are extracted and overlayed with the inventory test set. If the coordinates from the inventory fall within 500 m of those polygons, they are thus considered validated.

As mentioned in Section 2.2, we used the validation set to evaluate the optimal parameters for the U-Net model. These parameters were assessed per batch to inform parameters using the following metrics: Intersection over Union (IoU) mean, class-specific IoU, accuracy, class-specific F1. The IoU (Eq.2) is the defined as the area of overlap between the predicted landfill or background pixels and the landfill or background ground truth mask over the total area of both prediction and ground

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



truth combined. This is computed per image and only for images which contain landfills. For background IoU, the same equation is used (Eq.2) for all images. In the case of the mean IoU, the mean of both landfill and background IoU are macro-averaged, so that they are not pixel frequency weighted.

$$IoU = \frac{A \cap B}{A \cup B} \tag{2}$$

For the test set we use the following metrics: accuracy (Eq.3), precision (Eq. 4), recall (Eq. 5), F1 score (Eq. 6) and Matthews Correlation Coefficient (MCC)(Eq.7). The F1 score is the same equation for both the validation set and test set, measuring the balance of the precision and the recall. Note that on the validation set, in contrast to the test set, we also use TPs, FPs and FNs to calculate the F1 score, however this is computed using the ground truth masks for reference and not the aforementioned expert judgement.

$$Accuracy = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ number\ of\ predictions} \tag{3}$$

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \tag{4}$$

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \tag{5}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{7}$$

The precision equation (Eq. 4) for the purpose of the test set, calculates out all the instances where the model predicted positive detections and how many of those were correct. The recall (Eq. 5) assesses the model's ability to detect all the actual positive class, without missing detections. The F1 score (Eq. 6) is a mean of both precision and recall making sure precision doesn't hide the performance of the recall, specifically for our imbalanced case whereby landfills are a minority. The MCC, in contrast to F1, also accounts for an imbalanced dataset and assesses all four parts of the test set result: namely TPs, FNs, FPs and also TNs (Chicco & Jurman, 2023). This is particularly useful when predicting negatives are an important goal of the study. For the case of inventory validation where we focus on getting as many TPs as possible right, this is a less relevant but insightful metric to compute to understand the limitations of the model.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

3 Results

3.1 Testing on an Emission Inventory

To assess the emission inventory locations, we present both a confusion matrix at the image level and a geographical performance assessment at national level. The more TNs and TPs, the better our model is performing at either detecting background, where there might be misreporting or inventory errors, or landfills, where the reported location is detected within the 2x2 km S-2 box and therefore validated. In addition, we present a country specific case study whereby the performance of the segmentation at pixel-scale is also investigated (Sect 3.2.). An overview of the results of the test set per landfill location according to our U-Net model can be seen in Fig.3, below.

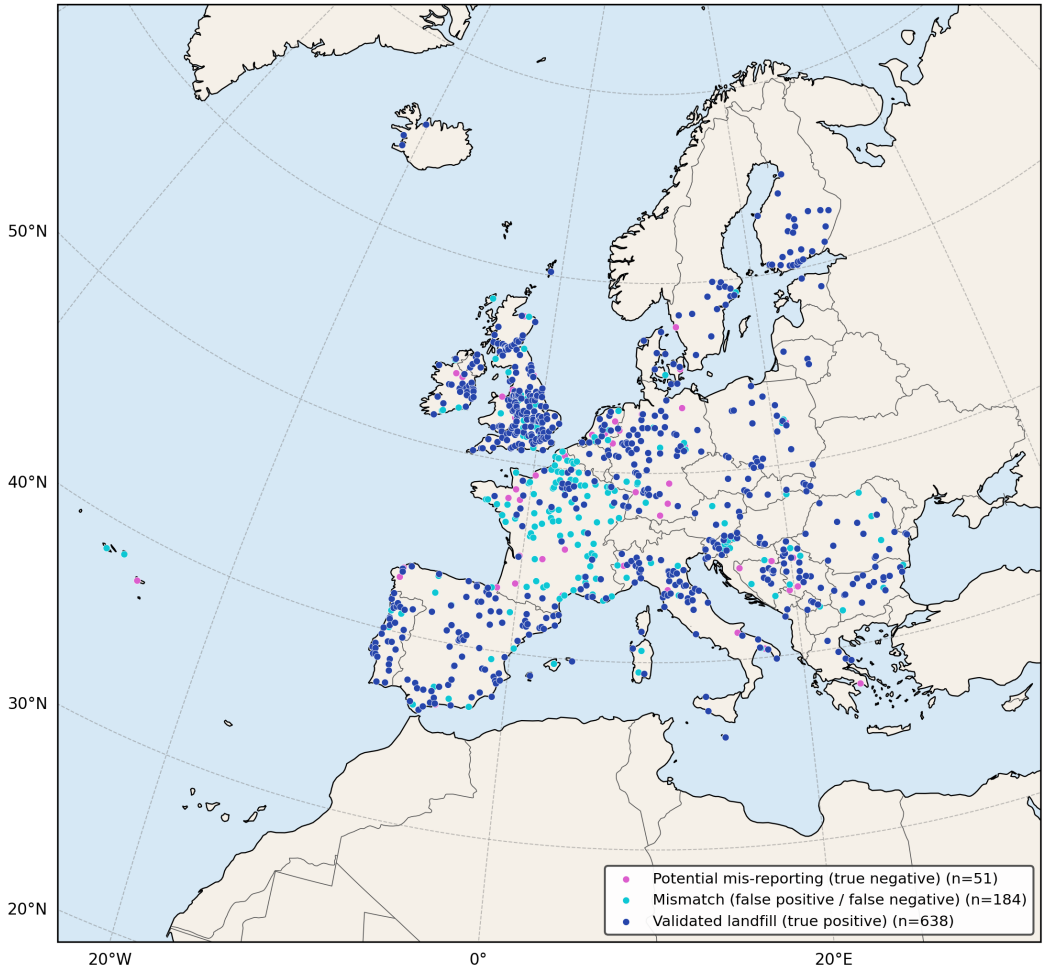


Figure 3. U-Net image-level statistics on the inventory test set for the 2020 reporting year, made with Natural Earth.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

On the test set, the model achieves a precision of 83 %, recall of 92 %, F1 of 87 %, an accuracy of 79 % but MCC of 0.25 (Fig. 4). It successfully detects 638 landfills in the reporting with 54 missed (FNs). Moreover, our U-Net successfully detects 51 cases where no landfill was found.

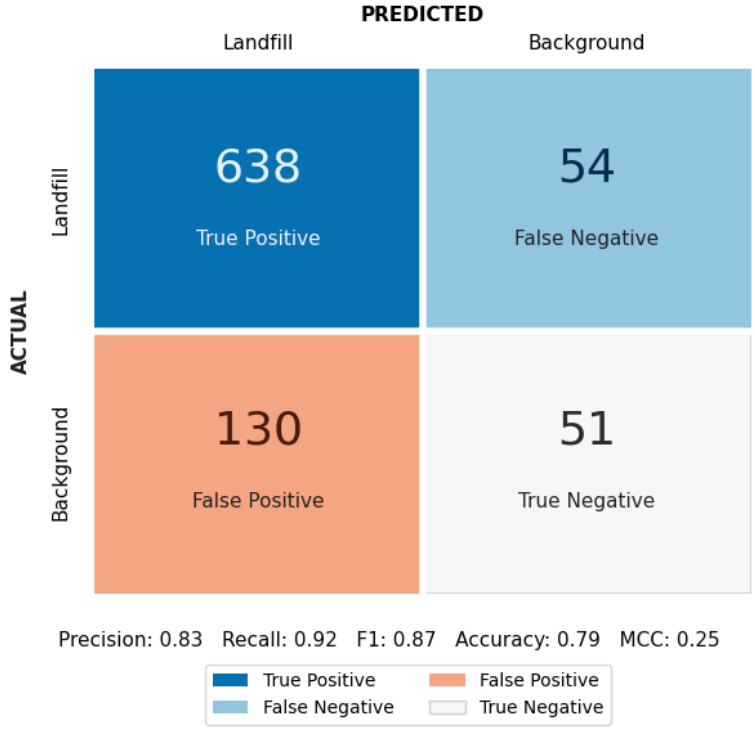


Figure 4. Confusion matrix on the test set. From top-right to bottom-left, the plot respectively represents: True Positives (TPs), False Negatives (FNs), False Positives (FPs) and True Negatives (TNs).

As previously discussed in Section 2.7, the results of the segmentation and classification at image level were checked. The result of the assessment displays a high recall (92 %) and F-1 score (87 %). The recall signifies that the model does a good job at detecting actual cases of landfill in the test set while minimising its instances of missing the landfills altogether. This is particularly important as the goal of the study is to validate as many landfills as possible, as is often the case in reporting validation. There is a lower result on the precision (83 %) as well as accuracy (79 %) though it remains reasonable. Regarding precision, this metric shows that false positives remain relatively high with 130 instances whereby the U-Net predicted landfills where there were none. This is further explained in the following section (Sect. 3.2). The F1 score on the other hand is a key metric in assessing our model, as the nature of our problem is imbalanced, namely there tends to be less pixels with landfills than background per image in the test set segmentation ($\bar{x}$ =8.3 %). As such, the F1 score of 87 % shows that the model is

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

effective at detecting TPs whilst reasonably avoiding both FPs and FNs in an imbalanced problem. The MCC score of 0.25 highlights that against a zero value, whereby 0 would be that the predictions are made purely by chance, it performs better than randomly guessing outcomes. However, this moderate performance, due to its different accounting of FPs (Eq.7) compared to F1, highlights that this model performs better at detecting positive cases of landfills correctly (recall) rather than identifying the absence of a landfill, which was not the goal of the study.

Looking at the country scale, when applied to the inventory test set, the model seems to generalise across all 29 countries (Fig. A4; Fig. A5) except France (Fig. 5). This is evidenced by the good performance on detecting 92 % of actual landfill cases, thus meaning that the model only missed 8 % of potentially actual landfills.

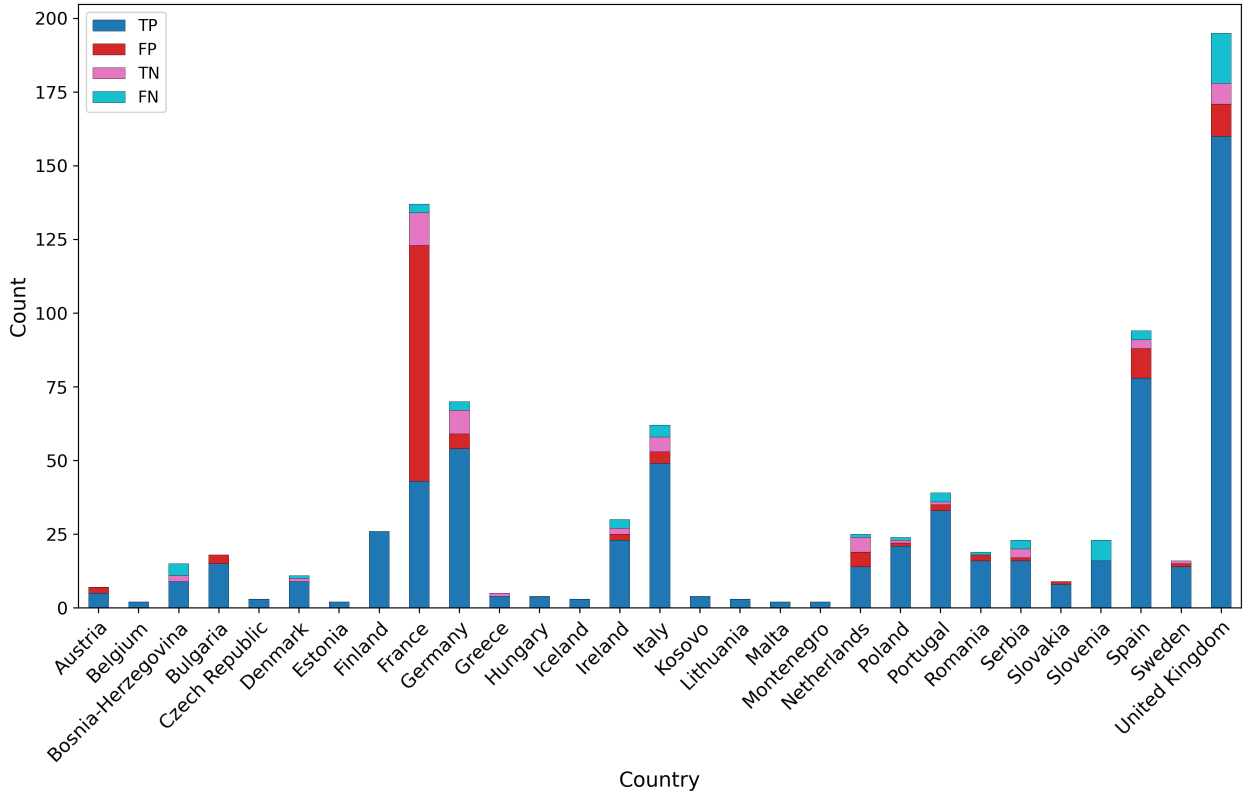


Figure 5. Count of U-Net test set predictions per country. The high number of false positives can be seen in this plot in red for France whilst the highest rate of false negatives is in the U.K.

The test set of landfill locations was overrepresented by the U.K. (n=195), France (n=137) and Spain (n=94). Some countries displayed no misreporting or misattribution by our model such as Finland with a 100 % detection accuracy across its 24 sites and, across a smaller sample size, Hungary (4 sites), Kosovo (4 sites) or Iceland (3 sites). However, the model displayed a

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

high number of false positives for some countries, whereby it detected a landfill where there was none. This is particularly the case in France (n=80) and to a lesser extent in the United Kingdom (n=11).

3.2. Error detection in France and the United Kingdom

3.2.1. France

As previously seen in Fig.5, there is a high number of false positive detections in France, representing the majority of model outcomes for the country (58 %). The FP detections in France are a consequence of a systemic misreporting issue whereby reported location coordinates were rounded at one to two significant figures. For the 2020 IEPR version for the 2018 year, this is the case for 87 unique sites in France. In some cases, this is detected by our model and is accounted for in the TNs count (Fig. 3), particularly in urban areas where our model performs well (Fig. A2a). In other cases, the model tends to segment other pixels with a similar spectral reflectance as a landfill when it does not find one (Fig. A2b). As can be seen in the European outcomes map in Fig. 3, the errors in coordinates are not concentrated in one specific French region or département and therefore cannot be attributed to a specific local authority. It is also spread throughout the country irrespective of company or location, although it tends to occur in small towns and some mid-size cities (such as Rennes). This points to a systematic (unintentional) error in the French reporting throughout the years, which has been partially amended in the newest IEPR versions. As per the latest version (v15) for the reporting year 2025, there were only 29 sites remaining with 2 significant figures issues in France, all dating back from the 2016 reporting year and before, as IEPR submissions are not retroactive by design. It is also important to note that some of the reported cases can be either landfills or local waste disposal centers ('déchetterie') which also fall outside of the scope of our training set but potentially within the scope of IEPR. Interestingly, in cases where no landfill is present in the Sentinel-2 imagery at a given coordinate, upon visual inspection, false positives tend to be associated with either locations with waste deposits (tires or construction material) in an urban-dominated or with croplands in cropland-dominated tile. This last confusion is explored in Fig. 6. Figure 6a shows the contribution of each Sentinel-2 band to landfill predictions, computed using gradient-based saliency (Sect. 2.7).

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

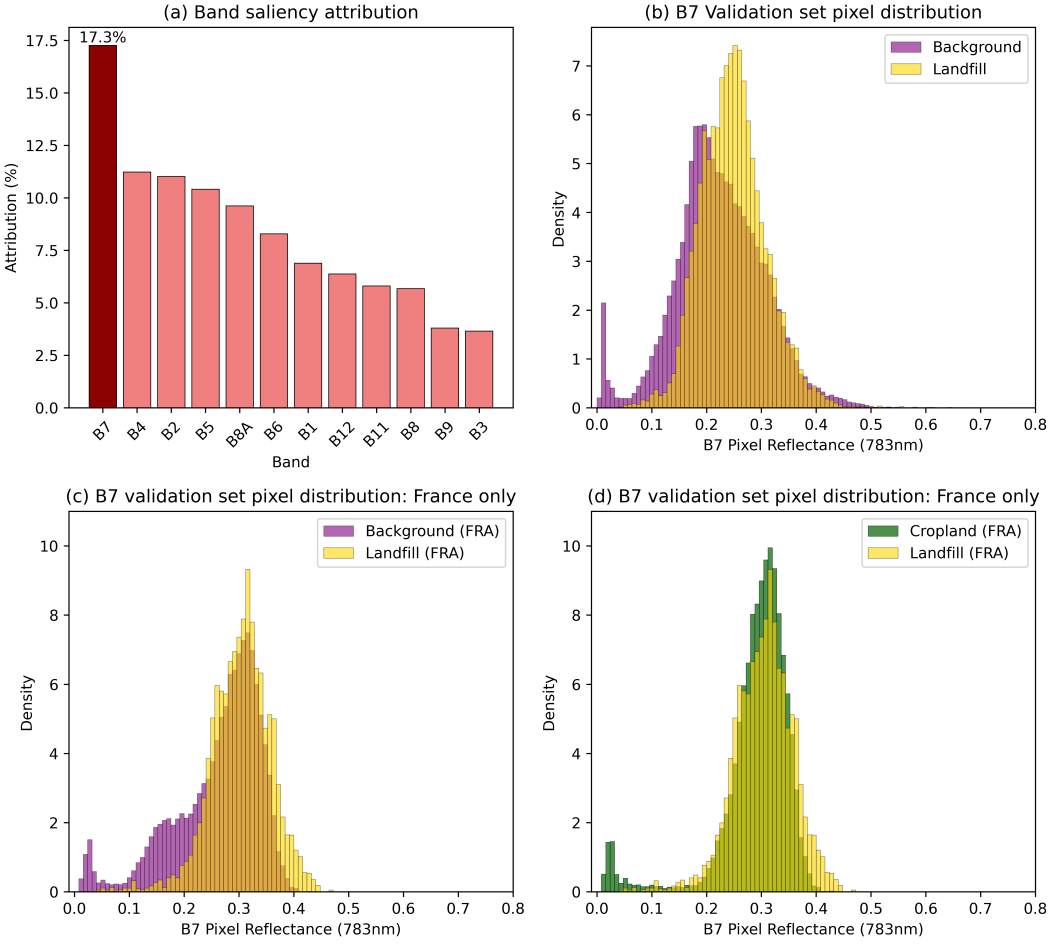


Figure 6. Saliency test on Sentinel-2 bands used in the U-Net validation set for band attribution and corresponding density plots (0.008 reflectance units per bin). **(a)** Saliency test for all 12 bands used in the validation set. **(b)** Pixel distribution of red edge B7 surface reflectance for the entire European validation set between background and landfills (median full year composites). **(c)** Pixel distribution of red edge B7 surface reflectance for only French landfills and French background (median full year composites). **(d)** Pixel distribution of red edge B7 surface reflectance for only French landfills and majority cropland areas (median full year composites). Landfills and background were defined using the ground truth masks from the validation set, whilst croplands were defined based on visual inspection.

This gradient-based saliency analysis shows that B7 (783nm) contributes to 17 % of the predictions for landfills for our U-Net followed by 11 % for B4. B7 sits in the red edge and is typically used for vegetation monitoring particularly leaf area index (LAI) and chlorophyll content particularly for agricultural monitoring (Delegido et al., 2011). Analysis of the pixel intensity of B7 distributions across the entire validation set (Fig. 6b) shows overlap between landfill and background class pixels although the majority of landfill pixel reflectance has a higher mode (0.26) than background pixels reflectance (0.19),

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

potentially allowing for pixel discrimination in the segmentation. When restricting the same set for France only, (Fig. 6c) red edge background pixels exhibited a bimodal distribution, with the mode at 0.31 pixel reflectance overlapping with the landfill pixel distribution with a similar major mode at 0.30, originating from cropland-dominated imagery. The second mode at 0.19 pixel reflectance shows little overlap with the landfill pixel distribution and originates from urban-dominated imagery. When separating for only cropland pixel reflectance in France (Fig. 6d), there is an almost perfect overlap between the modes of the landfill and the croplands at 0.30 pixel reflectance and 0.31 respectively, thus indicating that the most attributive band does not provide clean spectral separability between landfill and cropland classes in France. As such, we can attribute some of the high number of French cropland false positives to a confusion in pixel reflectance due to the most attributive red edge band. Moreover, landfill pixels in the whole validation set's B7, throughout European countries, had a similar pixel reflectance distribution and mode as the landfill pixels in France. This suggests that there is consistency across landfill reflectance in European countries thus indicating that this red edge band could be a useful band for the detection of landfills. However, despite the red edge being the most attributive, no single band clearly dominated the detection of landfills suggesting the model draws on multi-band information from all 12 bands rather than any single spectral feature.

3.2.2. United Kingdom

From the segmentation of the U-Net on the test set, we can also see the highest number of FNs in the U.K. (n=17) (Fig.5). Upon visual inspection, the main reason for landfills being missed by the model itself stems from the fact that many landfills reporting to IEPR seem to be either inactive or in the after-care phase. As such, 27 % of reported landfills in the U.K. in the inventory seem to be covered by vegetation at different succession phases namely grassland, shrubland or forests, or a mix (Fig. A3a). This therefore hinders the detection of landfills underneath. As can be seen in Fig. A3, our model trained on active cells in landfills therefore misses inactive landfills whilst successfully detects active landfills. Active landfills are defined as sites whereby waste is being actively deposited, including sites with complex treatment such as leachate treatment and geomembrane or daily cover usage. For example, the blue colour, in the visible, on the landfill in Verwood (Fig. A3, b) may be a sign of the active cell of a landfill facility using a daily cover to mitigate greenhouse gas emissions, waste from flying or rainwater from percolating at night when the landfill is not being worked on. On the other hand, the old access road to the landfill, the structural lines and vegetation cover on the Bexhill and Hastings landfill (Fig. A3a) are characteristic of an inactive landfill.

In opposition to the U.K., landfills in the France test set were open, active landfills and with limited potentially inactive sites (only 3 sites assessed as inactive). This is also the case in the rest of Europe and makes the U.K. a unique case for the IEPR reporting. Therefore, the afore-mentioned spectral overlap observed in France is not due to the presence of vegetated cover from natural succession but a potential cropland and landfill confusion on the most attributive band. One example of successful TN detection by the model, however, is the erroneous presence of wastewater treatment plants as solid waste landfills in the

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Netherlands due to a classification error in the inventory: 5 (c) 'Installation for the disposal of non-hazardous waste' which has now been amended to 5 (f) 'urban wastewater treatment plants' (European Commission, 2006). At the time of writing this is no longer present in the IEPR v.15 reporting and the inventory. Other errors were also detected in the test set (TNs) thanks to our model, however, due to limited knowledge as to the reason for their inclusion by national inventories, they will not be discussed further. In total, 45 % of the TNs could be explained by the aforementioned issues in France, the U.K. and the Netherlands.

3.3. Performance against ESA CORINE 2020

Performance evaluation of our model was done against a more classical land cover classification map to assess the quality of the image level classification (Fig. 7).

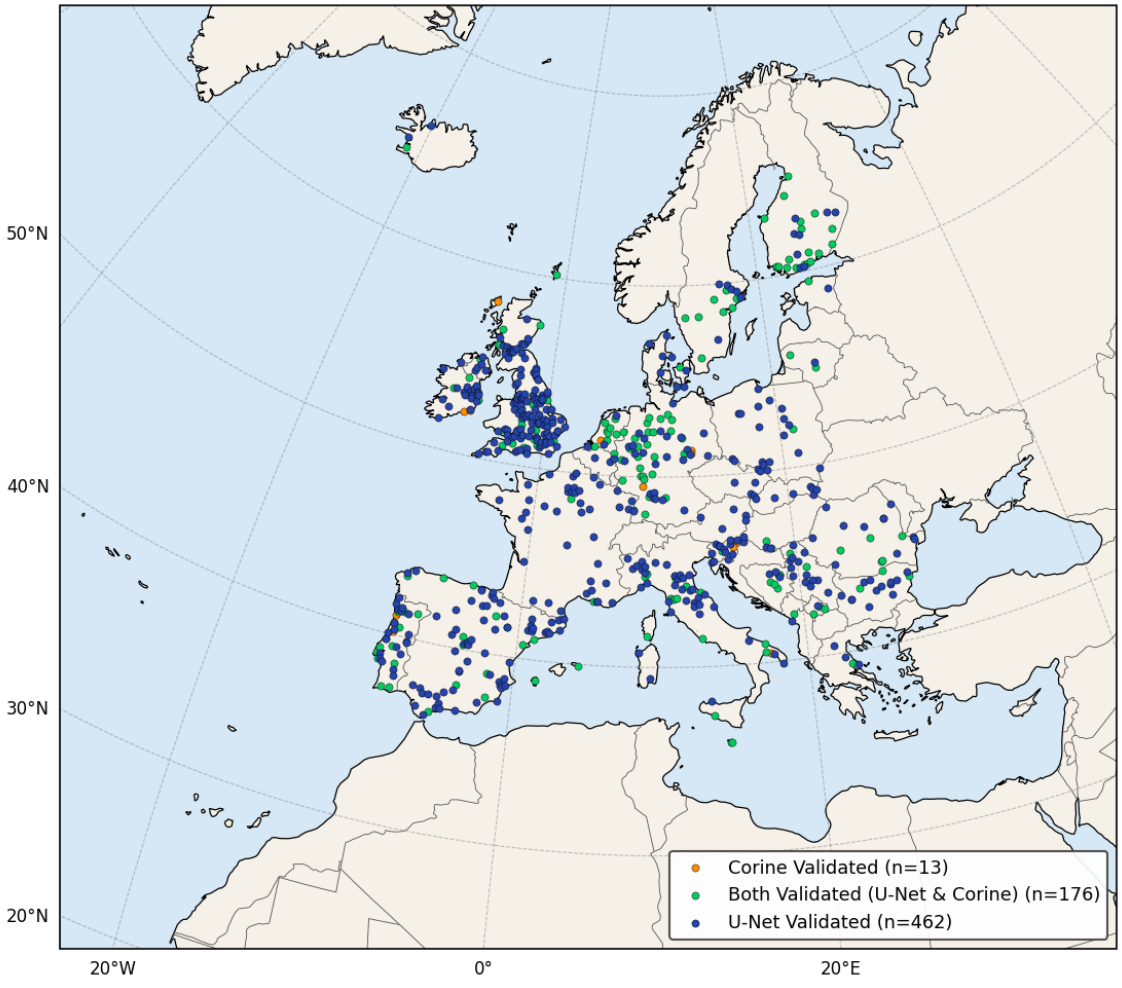


Figure 7. Comparison of U-Net based and CORINE-based validated inventory locations (true positives only), made with Natural Earth.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Both the model and the CORINE dataset found 176 reporting locations that both had polygons of landfills, thus increasing confidence in these data points. This is particularly the case in Germany (n=29) and the U.K. (n=31) as well as Finland (n=18). There were also 13 points validated in CORINE that were not validated with our U-Net, with no particular geographical pattern. However, some landfills were likely not detected due to the presence of vegetation cover on inactive landfills or geomembranes but also owing to a model error. In France, there were notably only three landfills validated by CORINE. 94 landfills were neither validated by CORINE or the U-Net, thus reaffirming the high number of errors in the reporting locations highlighted by the U-Net in France and emphasising the difficulty in detecting landfills in France. Another error agreement between CORINE and the U-Net was the case for the misreported wastewater treatment sites in the Netherlands, as mentioned in Section 3.2. In both cases, the CORINE dataset did not have landfill polygons and the U-Net did not detect the locations as landfills, showcasing the usefulness of cross-comparison. Finally, our U-Net detected and segmented 462 landfills from the inventory that were not present in the CORINE dataset. This makes the case for the use of this model as an improvement in detecting landfill locations compared to general land cover classification for inventory validation, although much is to be gained from a cross-comparison of the two methods, particularly in relation to TNs.

4 Discussion

4.1. Model Performance

The evaluation in Section 3 shows that our U-Net segmentation model performs well in detecting active landfills at the time of the study. This echoes another study using U-Net on UAVs (Unmanned Aerial Vehicles) which also found it to be effective in segmenting landfills that contained plastic waste in SWIR hyperspectral imagery (Bouchelaghem et al., 2026). In addition, the use of the red edge B7 by our model to discriminate a landfill from a background is still unknown. Studies investigating plastic waste detection inland have used the SWIR (Moroni et al., 2025) as well as NIR and SWIR in the case of microplastics, specifically for plastic polymers detection (Serranti et al., 2024). In marine environments, the NIR shows a peak in reflectance for marine plastic, acting as a differentiator with organic material such as seaweed, reflecting more in the green (560 nm) and red edge (700-780nm) (Bierman et al., 2020). As our training set is based on landfills that have emitted methane from Carbon Mapper satellite detections and estimates from Climate TRACE, this finding may point to a specific organic composition reflectance relevant for methane emissions from waste (in opposition to plastic-dominated landfills). However, further studies should investigate this specific red edge band for organic waste detections and in particular decomposition stages (Albrecht et al., 2008), particularly as new methane point source detections from landfills are being acquired by Carbon Mapper or GHGSat which can be used to improve this model further. The latter, as a hyperspectral imager, will be particularly useful for a detailed investigation of waste or landfill cell characterisation for methane detection. However, as the red edge pixels helped to differentiate only 17 % of the landfill pixels from the background, not one band seems to dominate the attribution of landfills.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

This indicates that the good performance on generalisation stems from the U-Net learning not only spectral reflectance combinations but also fine-grained spatial details (Bouchelaghem et al., 2026), such as landfill boundaries and patterns. This therefore emphasises the importance of U-Nets in segmenting the complexity of landfills from space.

4.2 Ablation Studies

As mentioned in the previous section, an edge case in our model was France, whereby the cropland and landfill confusion occurred. To remediate this, we tried to add hard negative cropland samples in France to the training set to teach our model to distinguish between cropland and landfills in Sentinel-2 imagery. With the addition of 30 cropland S-2 imagery throughout the European domain, including 18 in France, we ran an experiment to assess the performance difference using this new training set with a 54 % landfill and 46 % background image percentage in the training set, an increase from the aforementioned 2:1 landfill:background ratio. Despite some positive results in France with a reduction of false positives (-11 FP) and an increase in true negatives (+13 TN), it reduced the overall generalisation capability of the model. It became more conservative and displayed more false negatives throughout (recall=0.81), essentially missing out more correct landfills than our previous training set by 12 % (recall). However, the segmentation boundary precision was improved, suggesting the negative mining forced the model to learn more discriminative spatial features. As, the aim of this study is the validation of inventory points, it is recommended to use the baseline setup for Europe. Depending on the task, however, careful hard negative mining may improve the segmentation of the model, for example if segmentation accuracy of landfill waste size, capacity estimates, or active cell detection is key to the assessment. Specifically, negative mining should be done in a general manner and not be concentrated to one country to avoid skewness. Efforts should also focus on gathering as much of a diverse set of positive landfill imageries in the training set as possible.

Another ablation study was on improving the training set itself using data augmentation per image throughout bands. Using our baseline model, we performed individual trials on the following data augmentation techniques, both individually and as combinations: horizontal and vertical flips, 90° rotations, brightness and contrast enhancement. The results were evaluated according to the validation set IoU of either the landfill or background, the mean IoU and qualitatively. The combined augmentation achieved slightly lower mean IoU to the non-augmented baseline (Δ = -0.06) and the model training rose slower and achieved an IoU for landfills lower than the baseline (Δ = -0.14) at epoch 10, indicating that augmentation increased the difficulty of the training task but may have been too aggressive for the task at hand. Regarding each augmentation technique, however, it was found that the data rotations, flips and image enhancement were all detrimental to the IoU metrics, consistently under-performing both the baseline and combined data augmented baseline. This stresses the importance of data augmentation as a task-dependent and dataset size-dependent problem. This is consistent with prior work, investigating multispectral Sentinel-2 data augmentation techniques. Illarionova et al., (2021) found that geometrical transformations (data rotations, flips)

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

offered limited benefit for satellite images with medium resolution (10 m per pixel) due to insufficient introduced variability. They also further noted that brightness augmentation risks pushing samples outside the original distribution without associated gains in generalisation.

4.3 Limitations

Since one of the study's aims is to provide methane sensing satellites with locations of active landfills, the recall is the most important metric as it tracks our model's ability to detect true landfills and minimise missing an occurrence. A recall of 92 % thus means that 92 % of landfill coordinates in the reporting were detected and thus validated. However, for every landfill correctly detected, the model missed approximately 1 in 12 locations. Whilst our model generalises well to new locations, except for French croplands, and offers opportunities for validation at the European scale, we acknowledge that its overall performance could benefit from a larger training dataset. However, this was not possible at the time of the study as little training data in Europe was available from Carbon Mapper measurements and duly checked Climate TRACE locations. It is expected that more data will be produced in the upcoming years from these initiatives and could play a key role in improving the performance of our method. Regarding the MCC of 0.25, parts of this low score can be attributed to the number of FP images investigated in France, accounting for 44% of the FPs in the test set. When accounting for this error by experimentally removing the attributable FPs, the MCC climbed from 0.25 to 0.36, showing potential for improvement in our model. However, this study prioritises increasing the confidence of actual emission inventory landfill point sources by validating true positives rather than negatives. The reduction of false positives could nonetheless be improved by adding more discriminating information from terrain data such as NASA DEM. Regarding the cropland and landfill confusion observed in France in the red edge, it is therefore important to note that, as our model relies on both reflectance and the influence of high-resolution edges and landfill patterns, further confusion may have arisen from similar structural characteristics between French croplands and landfills which could be investigated by further work, for example by implementing Synthetic Aperture Radar (SAR) information from Sentinel-1 (Papale et al., 2023).

Regarding model parameters, particularly loss functions, a study on a limited set of imagery from Sentinel-1 for sea ice detection using U-Net similarly found good performance from the CE loss, pointing to good performance on a limited training set (Pires de Lima et al., 2023). This however contrasts with Bouchelaghem et al. (2026) who used a similar number of patches to our study (17,000) but potentially richer class representation from hyperspectral data leading to a good response to their combined focal-Dice loss setup. The insensitivity of our model to loss functions dealing with class imbalance such as Dice loss or combined CE-Dice loss demonstrated earlier, may point to a data ceiling whereby the model did not reach the point where loss function design was a binding constraint. It is therefore advised that increasing the training set with a greater variety of landfill patches across the European domain will call for a testing of loss functions.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

The presented training set may also show geographical bias as, despite it spanning 23 countries on the European continent, some countries were more represented such as France (n=14), Germany (n=11), Italy (n=10) or Sweden (n=10). For Germany, Italy and Sweden this representation also means that the model was effective at detecting the presence of landfills, regardless of their more complex on-site waste treatment compared to other countries in the test set, such as daily cover, leachate treatment and/or geomembrane. Therefore, this model is suitable for future advancements in complex waste management facilities over the European continent. For countries with open dumping and/or that receive a high proportion of municipal organic waste, it is therefore important to compile a training dataset that covers the whole variety of infrastructure types that are in use across the continent. This is particularly crucial for countries lacking waste management infrastructure, typically non-EU countries such as Serbia or Bosnia-Herzegovina, but also in some cases in Spain and Romania. This would otherwise be at the risk of including a bias towards well-managed landfills that would typically emit less methane than poorly managed landfills where a lot of efforts in both methane abatement and waste management are yet to be done (Mor and Ravindra, 2023). This structural bias could be prevented by using upcoming methane emission detection data from Carbon Mapper, GHGSat and/or scientific research campaigns to feed into a representative training set of the locations of methane emissions from landfills in Europe.

5. Conclusions

Recent advancements in satellite-based methane detections have provided a unique opportunity to leverage these observed locations as a training set for detecting methane-emitting facilities. Particularly in the case of Europe, this study's supervised application to a deep learning U-Net model using Sentinel-2 data shows potential for validation of point source inventories. With a recall of 92 % and F1 of 87 %, the high number of correctly detected landfills across the test set showcases that this model generalised well for both the detection and segmentation of the landfills across 29 countries on the European continent. Countries with the most intricate landfill management facilities such as Sweden, Germany and in some cases Italy, were also accurately segmented by our model, showcasing capabilities to adapt and detect to any progress in landfill management infrastructure, future proofing this model as countries are pushed to invest in better waste management. From an inventory improvement point of view, the identification and subsequent analysis of the true negatives (TN) is a major asset. In our case the 51 TN, representing ~6 % of the landfill locations, can be corrected with limited effort to their true coordinates. This immediately results in a higher quality database. However, model edge cases such as a confusion between croplands in France and landfills in the red edge triggered by a systemic coordinate reporting error as well as a significant number of inactive landfills in the U.K. showed potential for improvements in the training set itself as well as a good understanding of reporting intricacies across all inventory compilers. Beyond this, the model can be used to support the remote sensing community through satellite tasking, using this method as a prior QA/QC check, or for source attribution of medium resolution satellites such as TROPOMI (7 x 3.5 km) or the upcoming ESA TANGO mission (300 x 300 m). This method could equally support ground teams in campaign preparation, for example by determining the status of the landfill as active or inactive. Therefore, a

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

U-Net using Sentinel-2 for landfill segmentation and detection could be used by policy makers, emission reporting agencies as well as remote sensing companies to drive accurate quantification and help mitigate the rise in methane emissions globally.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix A

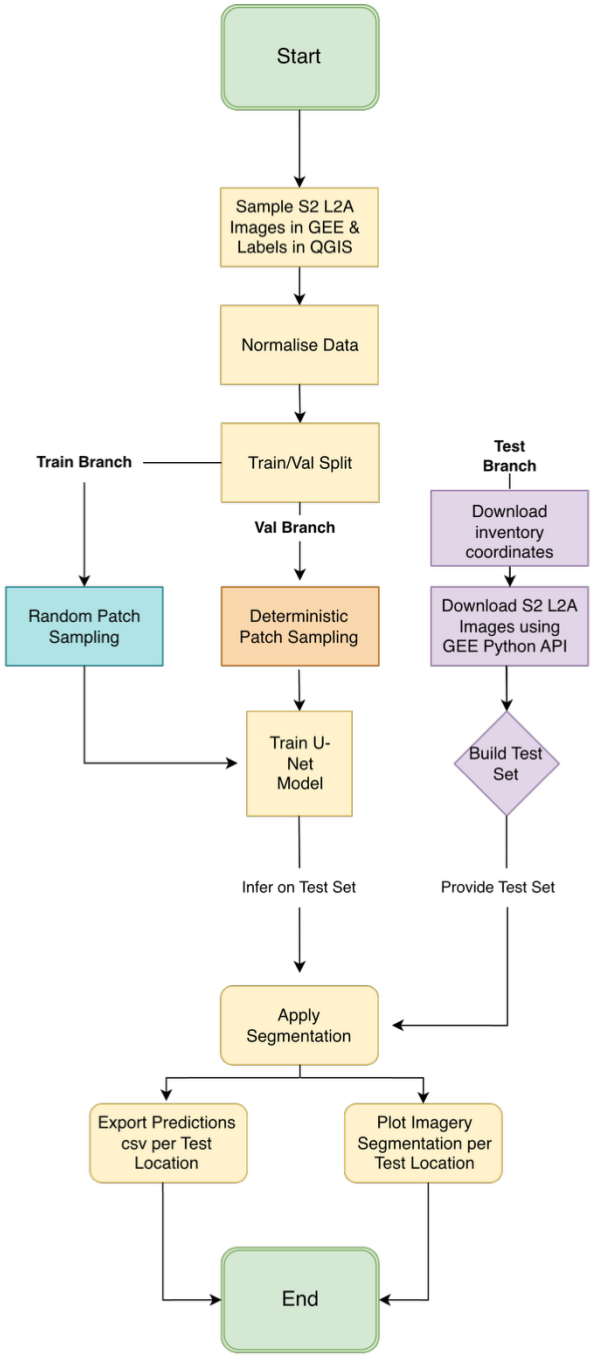


Figure A1. Methodological workflow for the landfill segmentation model on inventory locations. Glossary: GEE = Google Earth Engine; S2 = Sentinel-2 imagery, val = validation.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

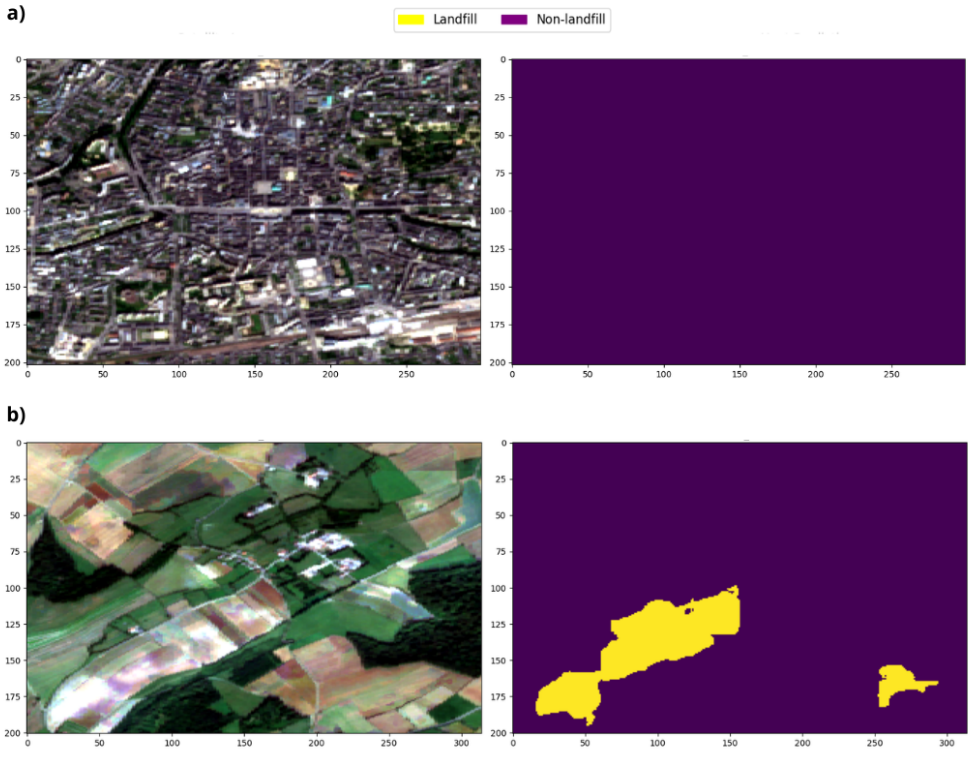


Figure A2. (a) Urban U-Net segmentation outcome (True Negative) against RGB in Rennes, France. **(b)** Cropland U-Net segmentation outcome (False Positive) against RGB in Bimont, France. The misreported coordinates in Bimont show model confusion of croplands with a landfill.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

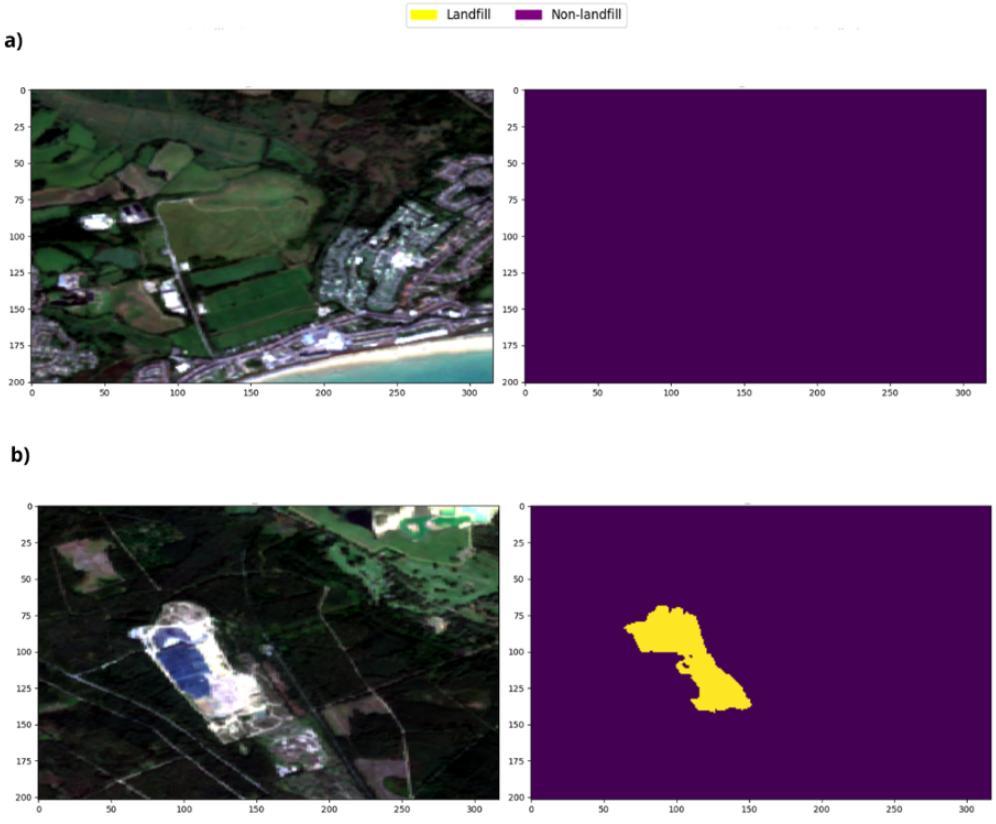


Figure A3. (a) U-Net segmentation outcome (True Negative) against RGB in Bexhill and Hastings Landfill, United Kingdom. **(b)** U-Net segmentation outcome of correctly segmented landfill (True Positive) against RGB in Verwood, United Kingdom. The landfill in Verwood has a distinctive blue reflectance showing the use of a landfill cover.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

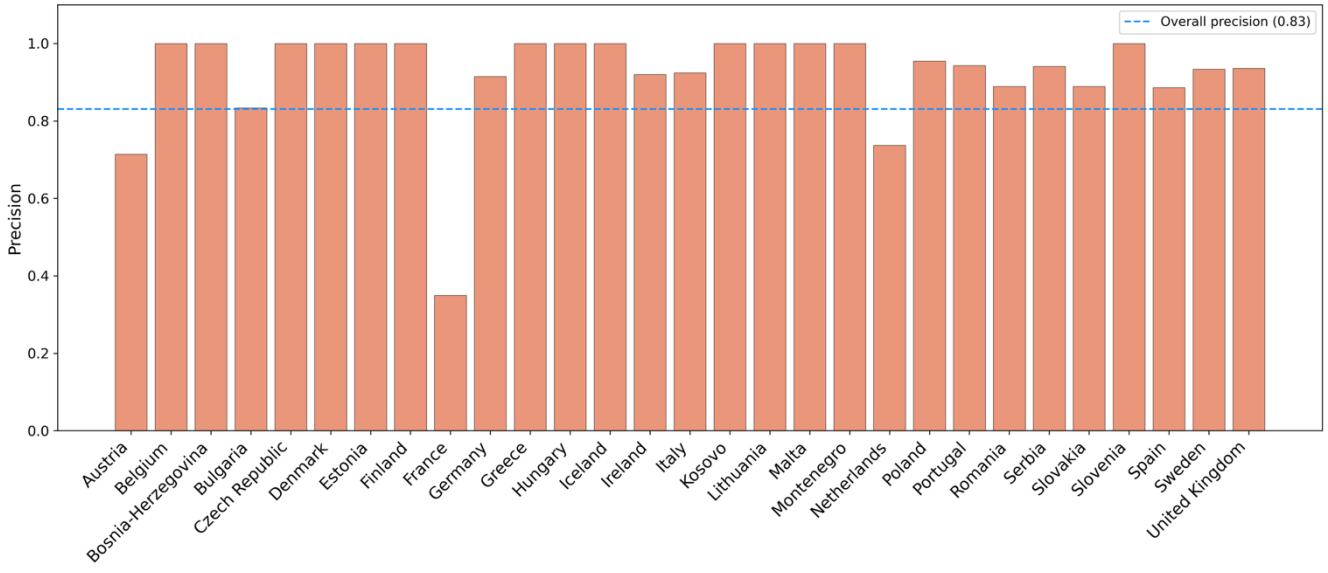


Figure A4. Precision metrics on TNO-GHGco v5 test set per country.

665

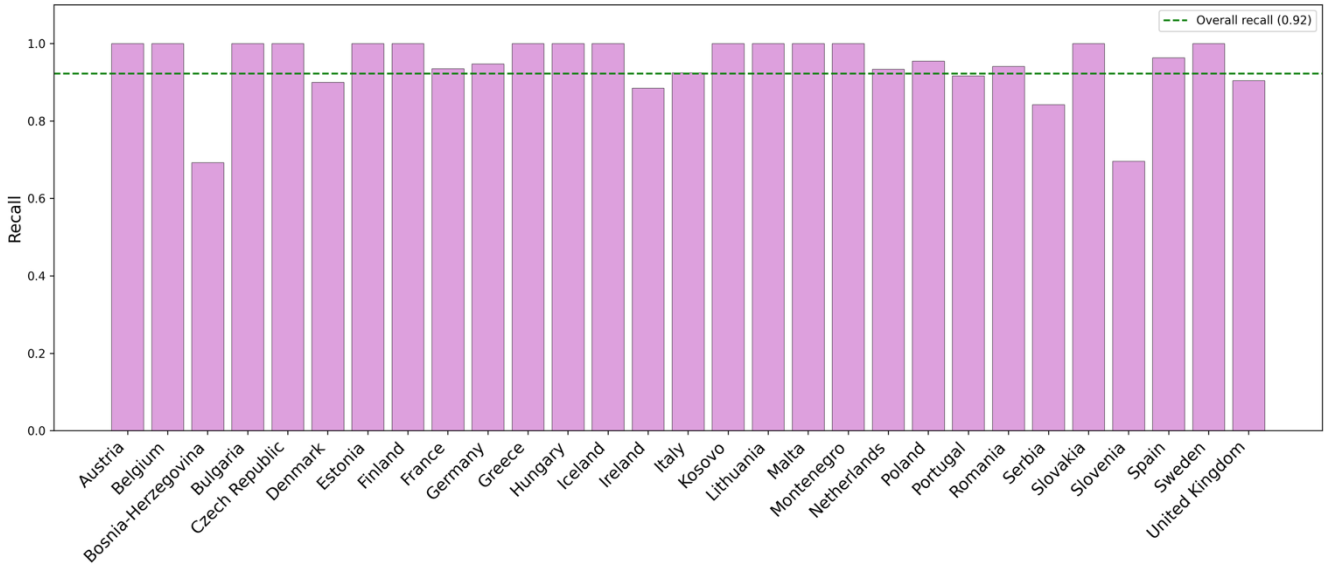


Figure A5. Recall metrics on TNO-GHGco v5 test set per country.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Data availability

Methane plume location data from landfills can be obtained from Carbon Mapper (https://carbonmapper.org/data). Landfill locations can be obtained from Climate TRACE (https://climatetrace.org/data). CORINE land cover data can be obtained from https://land.copernicus.eu/en/products/corine-land-cover/clc2018#download. Sentinel-2 Level-2A data can be downloaded from the Google Earth Engine (GEE) catalogue (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED) and using the Python ee API (https://github.com/google/earthengine-community/blob/master/guides/linked/ee-api-colab-setup.ipynb).
TNO-GHGco data available upon request. Access can be requested from Hugo Denier van der Gon: hugo.deniervandergon@tno.nl.

Author contributions

ES, AAN, AC and HD conceptualised the study. ES designed the experiments and methodology. ES and AAN developed the model code and performed the simulations. ES performed the formal analysis, investigation and visualisation for this manuscript. HD and SH supervised, reviewed and commented on the research activity. ES prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare no competing interest.

Statement on inclusion in global research

We acknowledge that parts of the work in manual sampling were done by AAN, as part of his internship at TNO under the supervision of ES. This included a remuneration for this work. Owing to the diverse positionality of researchers involved in this work (birth countries, lived experiences, class, education), it is the aim of this paper to minimise the bias in the manual sampling fed into the model and we aim for the experiences of researchers with regards to landfills to enrich the accuracy and understanding of our dataset and model, including their outcomes.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Moreover, the boundaries and designations used in figures do not imply official endorsement or acceptance by the authors or their institutions.

Acknowledgements

We thank Stijn Dellaert (TNO) for the development of the point source data in the TNO-GHGco dataset. During the preparation of this work the authors used Anthropic's Claude Sonnet 4.6 to improve some plot editing and code development. All generated code and figures were reviewed, tested and modified. All the manuscript writing, scientific analysis, interpretation of results, conclusions and reviews were solely performed by the authors, who take full responsibility for the content of the manuscript.

Financial support

This study is funded by the Horizon Europe programme Investigating Methane for Climate Action (IM4CA) (Grant Number: 101183460).

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Albrecht, R., Joffre, R., Gros, R., Le Petit, J., Terrom, G., and Périssol, C.: Efficiency of near-infrared reflectance spectroscopy to assess and predict the stage of transformation of organic matter in the composting process, Bioresource Technology, 99, 448–455, https://doi.org/10.1016/j.biortech.2006.12.019, 2008.

Biermann, L., Clewley, D., Martinez-Vicente, V., and Topouzelis, K.: Finding Plastic Patches in Coastal Waters using Optical Satellite Data, Sci Rep, 10, 5364, https://doi.org/10.1038/s41598-020-62298-z, 2020.

Bouchelaghem, S., Balsi, M., and Moroni, M.: Attention-Gated U-Net for Robust Cross-Domain Plastic Waste Segmentation Using a UAV-Based Hyperspectral SWIR Sensor, Remote Sensing, 18, https://doi.org/10.3390/rs18010182, 2026.

Chicco, D. and Jurman, G.: The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification, BioData Min, 16, 4, https://doi.org/10.1186/s13040-023-00322-4, 2023.

Climate TRACE, available from https://github.com/climatetracecoalition/methodology-documents/tree/main/2023, last access: 1 December 2023, 2023.

CORINE, European Union's Copernicus Land Monitoring Service, https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac, last access: March 2024, 2020.

Delegido, J., Verrelst, J., Alonso, L., and Moreno, J.: Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content, Sensors (Basel), 11, 7063–7081, https://doi.org/10.3390/s110707063, 2011.

Denier van der Gon, H. and CoCO2 WP2 team: D2. 1 Prior Emission data 2018 documentation report, https://coco2-project.eu/sites/default/files/2022-03/CoCO2-D2.1-V1-0.pdf, 2022.

Dogniaux, M., Maasakkers, J. D., Girard, M., Jervis, D., McKeever, J., Schuit, B. J., Sharma, S., Lopez-Noreña, A., Varon, D. J., and Aben, I.: Global satellite survey reveals uncertainty in landfill methane emissions, Nature, 647, 397–402, https://doi.org/10.1038/s41586-025-09683-8, 2025.

Duren, R., Cusworth, D., Ayasse, A., Howell, K., Diamond, A., Scarpelli, T., Kim, J., O'neill, K., Lai-Norling, J., Thorpe, A., Zandbergen, S. R., Shaw, L., Keremedjiev, M., Guido, J., Giuliano, P., Goldstein, M., Nallapu, R., Barentsen, G.,

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Thompson, D. R., Roth, K., Jensen, D., Eastwood, M., Reuland, F., Adams, T., Brandt, A., Kort, E. A., Mason, J., and Green, R. O.: The Carbon Mapper emissions monitoring system, Atmospheric Measurement Techniques, 18, 6933–6958, https://doi.org/10.5194/amt-18-6933-2025, 2025.

EEA: https://www.eea.europa.eu/en/european-zero-pollution-dashboards/indicators/reaching-the-2030-residual-municipal-waste-target-indicator-1, last access: 10 May 2026, 2025.

EEA: https://climate-energy.eea.europa.eu/topics/climate-change-mitigation/greenhouse-gas-emissions-inventory/data, last access: 10 May 2026, 2026a.

EEA: Annual European Union greenhouse gas inventory 1990–2024 and inventory document 2026, EEA Report No 04/2026, European Environment Agency: https://www.eea.europa.eu/en/ghg-knowledge-hub/resources-events/publications/annual-european-union-greenhouse-gas-inventory-1990-2024-and-inventory-document-2026, last access: 11 August 2026, 2026b.

EEA: https://industry.eea.europa.eu/, last access: 6 June 2026, 2026c.

European Commission: https://eur-lex.europa.eu/eli/reg/2006/166/2020-01-01/eng, last access: 9 July 2026, 2006.

European Commission: https://environment.ec.europa.eu/topics/industrial-emissions-and-safety/industrial-emissions-portal-regulation-iepr_en#objectives, last access: 16 May 2026, 2024.

Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Municipal_waste_statistics, last access: 10 May 2026, 2026.

Eleazer, W. E., Odle, W. S., Wang, Y.-S., and Barlaz, M. A.: Biodegradability of Municipal Solid Waste Components in Laboratory-Scale Landfills, Environ. Sci. Technol., 31, 911–917, https://doi.org/10.1021/es9606788, 1997.

Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13-15 May 2010, 249–256, 2010.

Grégoire, D. S., George, N. A., and Hug, L. A.: Microbial methane cycling in a landfill on a decadal time scale, Nat Commun, 14, 7402, https://doi.org/10.1038/s41467-023-43129-x, 2023.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Jackson, R. B., Saunois, M., Martinez, A., Canadell, J. G., Yu, X., Li, M., Poulter, B., Raymond, P. A., Regnier, P., Ciais, P., Davis, S. J., and Patra, P. K.: Human activities now fuel two-thirds of global methane emissions, Environ. Res. Lett., 19, 101002, https://doi.org/10.1088/1748-9326/ad6463, 2024.

Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I., and Denier van der Gon, H.: CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling, Earth System Science Data, 14, 491–515, https://doi.org/10.5194/essd-14-491-2022, 2022.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal Loss for Dense Object Detection, arXiv [preprint], https://doi.org/10.48550/arXiv.1708.02002, 7 February 2018.

Maas, A. L., Hannun, A. Y., and Ng, A. Y.: Rectifier nonlinearities improve neural network acoustic models, in: Proceedings of the 30th International Conference on Machine Learning (ICML), vol. 28, Atlanta, GA, USA, 3, 2013.

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

Madhuanand, L., Sadavarte, P., Visschedijk, A. J. H., Denier Van Der Gon, H. A. C., Aben, I., and Osei, F. B.: Deep convolutional neural networks for surface coal mines determination from sentinel-2 images, European Journal of Remote Sensing, 54, 296–309, https://doi.org/10.1080/22797254.2021.1920341, 2021.

Moroni, M., Balsi, M., and Bouchelaghem, S.: Plastics detection and sorting using hyperspectral sensing and machine learning algorithms, Waste Management, 203, 114854, https://doi.org/10.1016/j.wasman.2025.114854, 2025.

Naus, S., Maasakkers, J. D., Gautam, R., Omara, M., Stikker, R., Veenstra, A. K., Nathan, B., Irakulis-Loitxate, I., Guanter, L., Pandey, S., Girard, M., Lorente, A., Borsdorff, T., and Aben, I.: Assessing the Relative Importance of Satellite-Detected Methane Superemitters in Quantifying Total Emissions for Oil and Gas Production Areas in Algeria, Environ. Sci. Technol., 57, 19545–19556, https://doi.org/10.1021/acs.est.3c04746, 2023.

Papale, L. G., Guerrisi, G., De Santis, D., Schiavon, G., and Del Frate, F.: Satellite Data Potentialities in Solid Waste Landfill Monitoring: Review and Case Studies, Sensors (Basel), 23, 3917, https://doi.org/10.3390/s23083917, 2023.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Rachor, I. M., Gebert, J., Gröngröft, A., and Pfeiffer, E.-M.: Variability of methane emissions from an old landfill over different time-scales, European Journal of Soil Science, 64, 16–26, https://doi.org/10.1111/ejss.12004, 2013.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

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, Environ Sci Technol, 55, 16573–16580, https://doi.org/10.1021/acs.est.1c03976, 2021.

Scharff, H., Soon, H.-Y., Rwabwehare Taremwa, S., Zegers, D., Dick, B., Villas Bôas Zanon, T., and Shamrock, J.: The impact of landfill management approaches on methane emissions, Waste Manag Res, 42, 1052–1064, https://doi.org/10.1177/0734242X231200742, 2024.

Serranti, S., Capobianco, G., Cucuzza, P., and Bonifazi, G.: Efficient microplastic identification by hyperspectral imaging: A comparative study of spatial resolutions, spectral ranges and classification models to define an optimal analytical protocol, Science of The Total Environment, 954, 176630, https://doi.org/10.1016/j.scitotenv.2024.176630, 2024.

Simonyan, K., Vedaldi, A., and Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, arXiv [preprint], https://doi.org/10.48550/arXiv.1312.6034, 19 April 2014.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M.: Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 240–248, https://doi.org/10.1007/978-3-319-67558-9_28, 2017.

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

Wang, Y.-R. and Li, X.-M.: Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning, Earth System Science Data, 13, 2723–2742, https://doi.org/10.5194/essd-13-2723-2021, 2021.

https://doi.org/10.5194/egusphere-2026-6367
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O.: ESA WorldCover 10 m 2021 v200, https://doi.org/10.5281/zenodo.7254221, 2022.

Zhang, Z. and Sabuncu, M.: Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels, in: Advances in Neural Information Processing Systems (NeurIPS), Montreal, Canada, 2-8 December 2018, 2018.