Geomatica 77 (2025) 100048

Contents lists available at ScienceDirect

Geomatica

journal homepage: www.sciencedirect.com/journal/geomatica

Sensitivity of land-type variations across Canada using S-5p products

Saheba Bhatnagar [a,c,*], Mahesh Kumar Sha [b], Mariana Silva [a], Laurence Gill [a,**], Bavo Langerock [b], Bidisha Ghosh [a]

[a] School of Engineering, Trinity College, Dublin, Ireland
[b] Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[c] BeZero Carbon Ltd, London, United Kingdom

ARTICLE INFO

Keywords:
Land change
S5-p
Machine learning

ABSTRACT

Methane (CH_4), a potent greenhouse gas, traps heat in the atmosphere and significantly contributes to global warming. It is unclear whether CH_4 emissions from various land-types and other natural sources have increased substantially in the last decade linked, for example, to global warming and uncertainties remain regarding sources and their spatial extent causing discrepancies between emission estimates from inventories/models and estimates inferred by an ensemble of atmospheric inversions. Here we compared remotely sensed CH_4 total column data, along with surface albedo from the Sentinel-5 Precursor (S-5p) satellite against six main temperate zone land types (marsh, swamp, forest, grassland, cropland, and barren-land across Canada over a four-year period (2019–2022). The study developed a machine learning based algorithm that can be used to classify between such different land types using S-5p products. From 2019 to 2022, the average producer's accuracy (PA) across all land types ranged from 50.8 % to 98.4 %, while the average user's accuracy (UA) ranged from 69.9 % to 95.4 %. Although the methodology presented does not directly differentiate the methane fluxes from different land types, it does provide a foundation that with better ground truth monitoring and higher resolution imagery, could lead to a being able to differentiate methane emissions between land types with increased confidence, as well as determining whether significant changes are occurring over time. This would yield valuable insights for climate scientists and policy makers at both national and international levels.

1. Introduction

Methane (CH_4), after carbon dioxide (CO_2), is the second most important anthropogenic greenhouse gas contributing to climate change. Compared to CO_2, it has a shorter atmospheric lifetime of about 9 years [25], making it a favourable target for climate change mitigation. Atmospheric emissions and concentrations of CH_4 have increased continuously over the last decade [27]. Wetlands are known to be the largest natural source of CH_4, with an estimated average global emission, from "bottom-up" inventories/modelling approaches, of 149 Tg CH_4 yr^{-1} (range 102–182) during the past decade (2008–2017) [27]. About 5 % of the atmospheric CH_4 uptake is by the methanotrophic bacteria present in unsaturated oxic soil, with the main sink being chemical reactions in the atmosphere [27]. The CH_4 emission contribution from land types is calculated as the product of emission flux density and the surface extent of CH_4 source/sink area [6,21]. The seasonal and inter-annual variability of these land types' areal extent is considered the main cause of uncertainty in calculating their absolute flux of CH_4 emissions, which is significant for the global CH_4 budget [6,9,24].

Satellites can be utilised to gain insights into CH_4 dynamics at a global scale. Greenhouse gases Observing SATellite (GOSAT) and Envisat onboarded instrument SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY (SCIAMACHY) provide measurement of trace-gases at a resolution of (10.5 ×10.5) km, (60 ×30) km at best, respectively. The observations from these satellites/instruments show a larger concentration of CH_4 and other gases in highly populated areas affected by human activities (as hotspots). Webb et al. [36] verified the agreement between methane concentrations measured by GOSAT and in-situ analysis over the wetland ecosystem at Amazon Basin. Despite the growing availability of satellite-based CH_4 measurements, important research gaps remain. Previous studies have largely

* Corresponding author at: BeZero Carbon Ltd, London, United Kingdom.
** Corresponding author.
E-mail addresses: saheba.bhatnagar@bezerocarbon.com (S. Bhatnagar), laurence.gill@tcd.ie (L. Gill).

https://doi.org/10.1016/j.geomat.2025.100048
Received 2 October 2024; Received in revised form 23 December 2024; Accepted 27 January 2025
Available online 28 January 2025

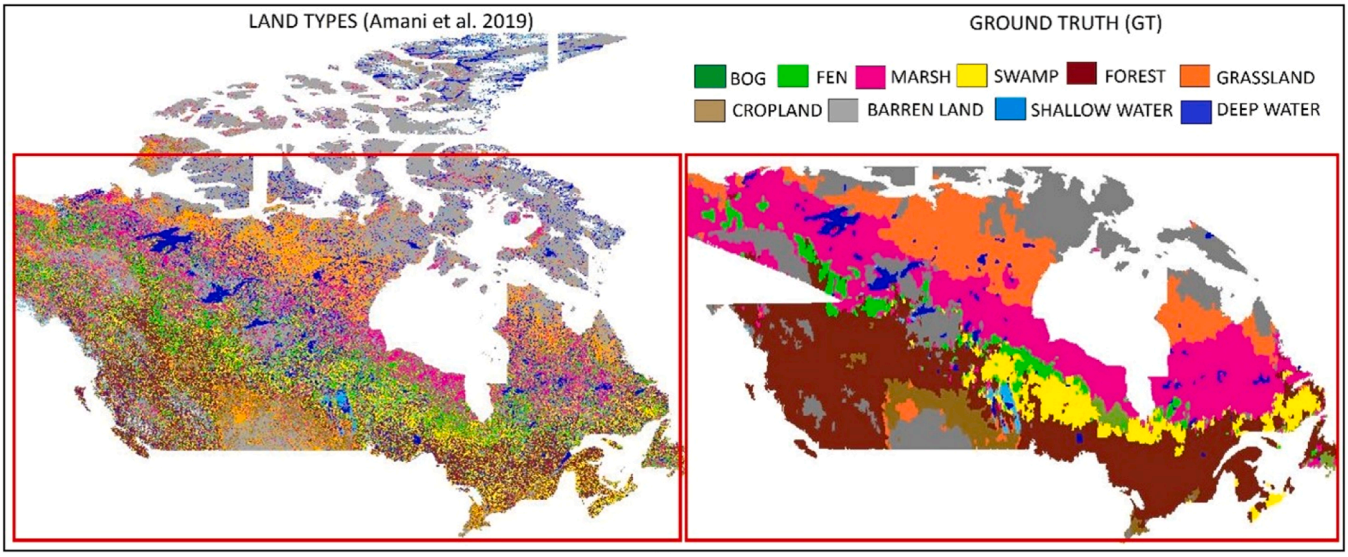

Fig. 1. Land type classification map (ground truth) creation for Canada. (left) Land types in Canada as described in Amani et al. [1] (10 classes) at 30 m spatial resolution and (right) ground truth (GT) created using the MODIS NDVI product (Didan [10]) and graph cut segmentation at 0.05° spatial resolution. The maps were generated using Matlab v.2019b software. The boundary of the map was taken from the open-source website https://www.igismap.com/canada-shapefile-download-free-adminstrative-boundaries-provinces-and-territories/ (last accessed on 15 June 2020).

focused on either point-scale in-situ measurements or global-scale model assessments, often without a nuanced understanding of how surface albedo variations and land-type differences might affect satellite CH_4 retrievals [16]. A critical gap emerges in integrating these observations and modelling frameworks to systematically assess the interplay among land-cover change, CH_4 flux variability, and the albedo-dependent sensitivities inherent in satellite retrievals. This gap is particularly evident for regions with dynamic wetland extents or intensive agricultural activity, where shifts in land use can substantially alter the magnitude and distribution of CH_4 sources and sinks. Wetlands, agriculture, and forests (afforestation, deforestation) significantly affect methane emissions. Hence, some correlation between land-types and methane emission at a large scale, although difficult to quantify without in-situ measurements, is expected. To quantify this at a global scale, an extensive database of methane sources and sinks is required. There are further uncertainties due to the presence of methane in the atmosphere for a long time and its transportation via wind. However, the quantification is also dependent on the actual areal extent of different land-type, which is still very approximate and in many parts of the world and needs to be improved, such as the extensive areas of marsh, bog, swamp and fen in Canada [8,14].

The objective of this study is to show the sensitivies and albedo dependence to variations in land types to a CH_4 product from a space instrument - i.e., TROPOspheric Monitoring Instrument (TROPOMI) - which is onboard the European Space Agency's (ESA) Sentinel-5 Precursor (S-5p) satellite measuring daily global total column concentrations of atmospheric CH_4. Our primary hypothesis posits that the total column methane (CH_4) measurements from Sentinel-5 Precursor (S-5p) products, combined with machine learning algorithms, exhibit unique sensitivity to large land types such as marsh, swamp, forest, and grassland, enabling effective differentiation despite the coarse resolution of satellite imagery.

2. Data and study region

2.1. Satellite-based CH_4 total column data used in this study

This study uses CH_4 total column measured by the satellite Sentinel-5 precursor (S-5p) [2,34]. The CH_4 total column measured by the satellite is a combination of CH_4 production, oxidation in the atmosphere (or soil uptake), and transport. The spatial resolution of the operational level 2 SWIR product was originally 7×7 km^2 in exact nadir and was increased to 5.5×7 km^2 on 6 August 2019. The operational processing to retrieve the column averaged dry air mixing ratio of CH_4 is performed by RemoTeC S5 algorithm [15]. The operational CH_4 total column product consists of a standard product and a bias-corrected product. The standard methane product (operational and scientific) suffers from albedo related methane biases and often elevated methane features were related to such dependencies on surface albedo [3,12,20,28,29]. The details of the bias correction and the albedo dependence are described in the Algorithm Theoretical Baseline Document (ATBD) (Hasekamp et al., 2022). The latest operational product version 02.04.00 of the S-5p CH_4 total column data from Jan 2019 until Dec 2022 has been used in this study. This was used for the full mission reprocessing of data on or before 25 July 2022 and after that from offline data The quality of the S-5p CH_4 data has been verified by the ESA mission performance centre (MPC) by performing validation against reference ground-based remote sensing networks of the Total Carbon Column Observing Network (TCCON) and the Infrared Working Group (IRWG) of the Network for the Detection of Atmospheric Composition Change (NDACC) [18,19,29]. The reported systematic uncertainty of the bias corrected methane product validated against 24 TCCON stations is 0.29 ± 0.27 % and the random uncertainty is 0.68 ± 0.17 % [18]. These stations are located in different parts of the world representing different surface conditions (land types and corresponding surface albedos) and atmospheric conditions. However, the network is relatively sparse with gaps in many regions of the world and do not cover the complete range of surface conditions. So, the uncertainty values stated corresponds to only the characteristics of the sites where the TCCON stations are located and similar situations elsewhere. As S-5p records solar absorption measurements reflected by the Earth's surface and the atmosphere, measurements are not possible over larger parts of Canada during the winter months (Nov-Jan) and poor coverage during the early spring and late autumn periods.

The S-5p bias-corrected CH_4 total column values along with the retrieved surface albedo (SA) for quality assurance (qa) value greater than 0.5 were selected and binned on a regular 0.05° grid to form the level 3 (L3) data. The HARP toolkit component of the ESA atmospheric toolbox (https://atmospherictoolbox.org) was used to perform the latitude longitude regridding where each S-5p pixel contributes to the

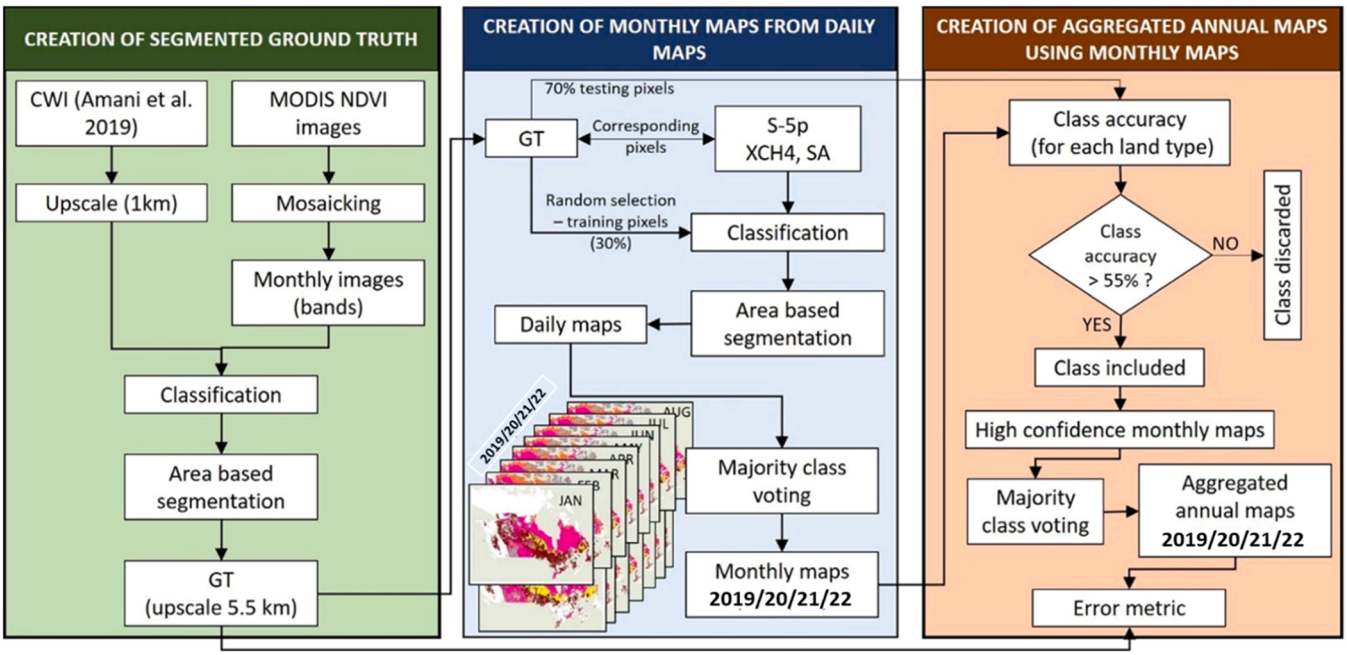

Fig. 2. Flowchart showing the complete process from the creation of the ground truth (GT) to the creating of annual land type classification maps. The maps were generated using Matlab v.2019b software.

regridded CH_4 value of the target grid cell if there was an overlap of the pixel and the grid cell. In case when multiple pixels overlap, a grid cell weighted average was taken using the overlap area as the weight.

2.2. Region of study

2.2.1. Selection of the region

Wetlands cover approximately 5.5 % of the global land surface with an average areal extent of 8.0 to 8.4 million km^2. Apart from the ecological significance, wetlands store atmospheric carbon and act as a net carbon sink but also emit methane due to anaerobic substrate conditions. Peatland wetlands, for example, cover 3 % of the Earth's land surface but store approximately 25 % of the global soil carbon [38]. The CH_4 production in wetlands is influenced by the spatial and temporal extent of anoxia (water level in the soil), temperature, availability of substrate, and plant ecology [32,33,35,37]. Monitoring these wetlands using remote sensing is a resource and time-efficient endeavour with significant ecological and environmental importance. A large section (~25 %) of the world's remaining wetlands are located in Canada, covering 12.9 % of Canada's terrestrial area [11,23]. Therefore, we have chosen Canada as our study region due to the presence of large wetland areas (and other land types) which are known to emit differing quantities of CH_4 and the availability of a land type map for Canada (described in Section 2.2.2) for verifying our results.

2.2.2. Canadian Wetland Inventory

In 2019, Amani et al. [1] created the first Canada wetland inventory (CWI) using a composite of approximately 30,000 Landsat-8 surface reflectance images collected from 2016 to 2018. This method allows monitoring and mapping wetlands every three years with 66 % producer and 63 % user accuracy. The producer accuracy is accuracy that relates to correct classification of a class by the algorithm whilst the user accuracy depicts the reality on the ground. The CWI map included five wetland classes defined by the Canadian Wetland Classification System (CWCS) – bog, fen, marsh, swamp, and shallow-water – as well as other land types – forest, grassland, cropland, barren (rocks, gravel, built-up areas, non-vegetation), deep-water, and snow (Fig. 1a left inset). The S-5p data being used for this study matches the timeframe of creation of the first CWI; therefore, it was used to generate the S-5p resolution specific ground truth labels described in the next section.

3. Methods

A brief description of the machine learning (ML) algorithm utilised to create the labels and analyse the satellite data is described here. The ML algorithm used was initially developed to identify vegetation communities within wetlands using remote sensing, and the steps for customising the algorithm for detecting the sensitivities of land types to the S-5p products (methane and SA) are as follows.

3.1. Creating ground truth labels from CWI

The CWI map is available at a significantly higher spatial pixel resolution of 30 m compared to the binned S-5p resolution at 0.05° grid (~5.5 km). The CWI map was therefore upscaled to a lower resolution map combining additional Moderate Resolution Imaging Spectroradiometer (MODIS) normalised difference vegetation index (NDVI) product. The MODIS NDVI (Normalized Difference Vegetation Index) (MOD13A3) produces monthly NDVI maps at 1 km resolution with about 15 tiles covering the area of interest (AOI) in Canada marked with a red rectangular box in Fig. 1. All of the NDVI images were mosaicked using the mean value (for the overlapping areas) for the years 2018 and 2019 to have consistency with first CWI. Therefore, a 3-dimensional image with 48 bands was created using layer stacking for the AOI. To create proper segments, the CWI map was also upscaled to 1 km resolution such that it is compatible with the MODIS NDVI product. The map was upscaled to 5.5 km spatial resolution using nearest neighbour interpolation for upscaling it to the same resolution as the L3 S-5p data.

Some of the islands (far north) were not considered due to poor or insufficient availability of S-5p data during long winter periods, in which the area was covered in snow/ice and/or clouds, limiting the satellite's view. The smoothed segmented map created for the selected area (Fig. 1a right inset) was used as the ground truth (GT) in this paper. The conventional remote sensing analysis is performed against manually collected field data. Other ready to use products like CORINE land cover (CLC 2018), MODIS land cover map (MCD12Q2: https://lpdaac.usgs.

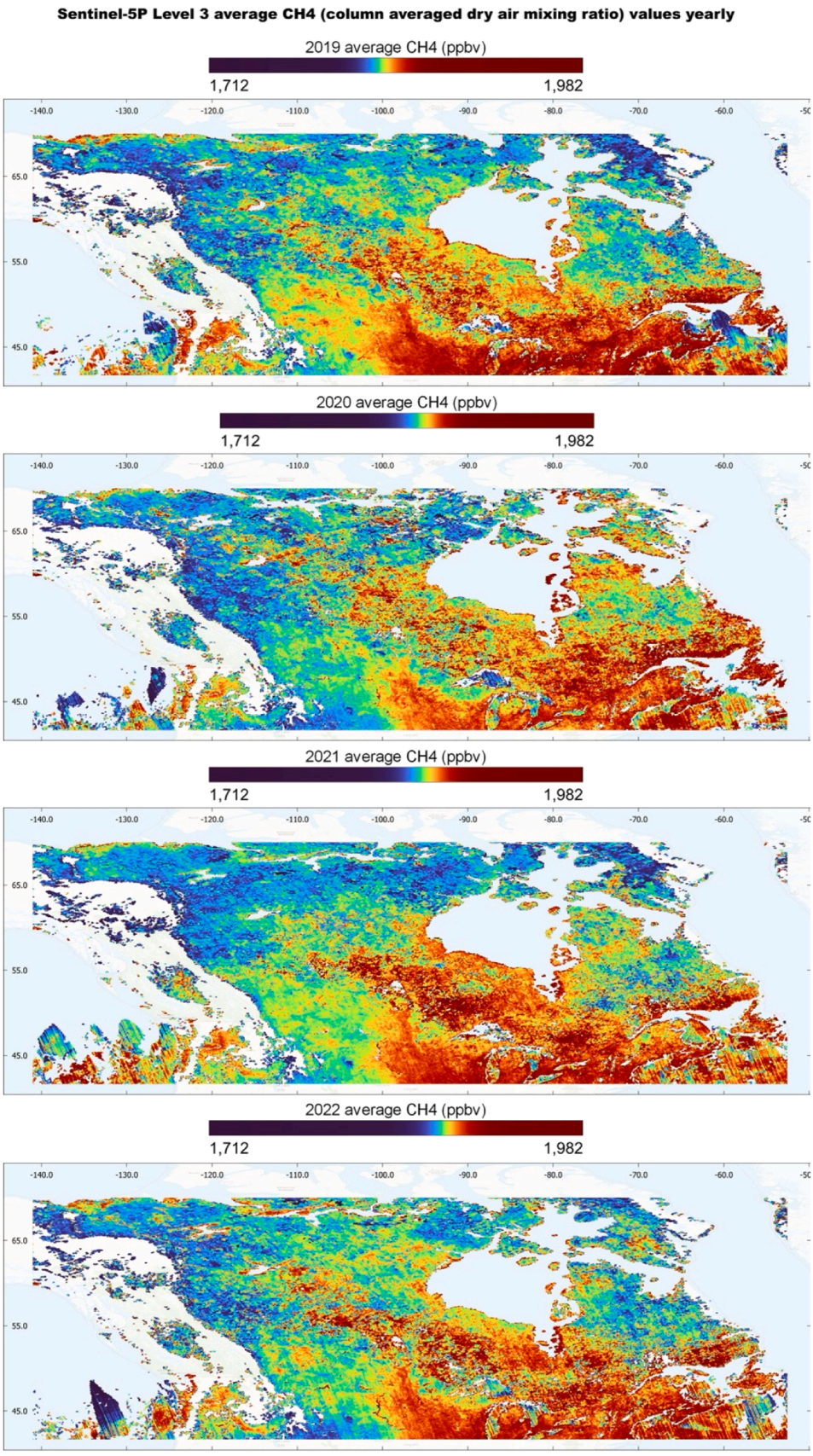

Fig. 3. Annually averaged CH_4 maps (2019–2022) based on continuous colour ramp. The maps were converted from NetCDF using Matlab v.2024a and formatted in QGIS v. 3.24.3.

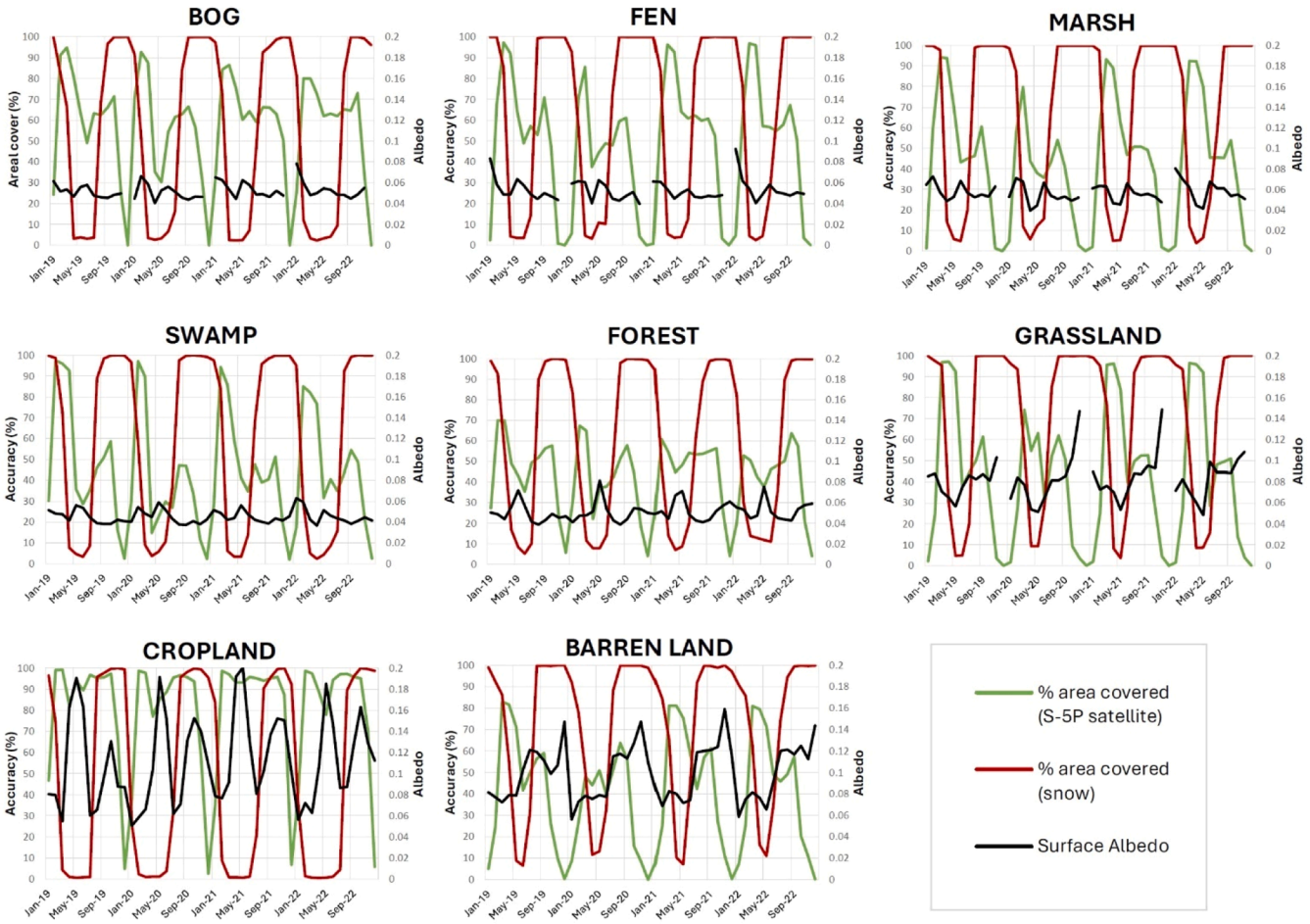

Fig. 4. Time series plots of relevant data for comparison against class accuracies for different land types. Representation of % area covered by snow (deep red), S-5p (green), and surface albedo (black; right y-axis) are shown for each land type over the time period Jan 2019 – Dec 2022 (x-axis). The % area covered by snow is obtained using MODIS daily snow cover product (MOD10A1); the area covered by S-5p is the monthly average of the area captured by S-5p.

gov/products/mcd12q2v006/), etc., can also be used as initial ground truth for validating a similar study. It has to be noted that in this study, the upscaling was due to the absence of field-based ground truth. The MODIS land cover map also had a limitation on the wetlands types; therefore, CWI was the best product available to conduct this study.

3.2. Classification-segmentation machine learning algorithm

The level 3 regridded S-5p CH_4 total column and SA over Canada were analysed utilising a ML algorithm described in Bhatnagar et al. [4]. The workflow, including the development of ground truth (GT) maps, creation of monthly and yearly maps, and performance evaluation of the algorithm, is described in Fig. 2. An example of CH_4 data, annually averaged for the time period 2019–2022, within the study period is provided in Fig. 3.

The first step of the analysis was creating a GT map for evaluating the sensitivity of S-5p CH_4 total column measurements to certain land types, especially wetlands. Next, the S-5p data was used for classification using a segmentation model using random forest classification followed by graph cut segmentation based on posterior probability; a detailed description of the model can be found in Bhatnagar et al. [4]. Random forest classification was implemented with the number of trees = 100, as this value provided a balance between computational efficiency and classification accuracy. The model parameters included a maximum depth of the trees, minimum samples per leaf, and minimum samples required to split an internal node, all optimized through cross-validation. Both pixel-based intensity and contextual information (area-based segmentation) were utilised to enhance classification performance. The dependent variable in this study was the land type classification for each pixel, derived from the GT map. The training process involved stratified random sampling, with 30 % of pixels from the GT map selected for training and the remaining 70 % used for validation. Collins et al. [7] suggest that using random training samples with equal representation of each class is necessary to avoid classifiers' bias. The manual selection of training points may produce clustered training points, thereby increasing the inherent spatial autocorrelation [22]. Therefore, stratified random sampling of training data with equal weightage to each class was selected for this study. Stratified random sampling is advantageous as it usually yields more accurate estimations [31].

To detect the areal extent of land types, the methodology extended beyond pixel-based (PB) classification by incorporating area-based segmentation using a graph cut algorithm. The graph cut segmentation approach, based on the maximum a-posteriori (MAP) principle, aimed to minimize energy (E) by reducing the data cost ($D(p,np)$) and smoothness cost ($Vp,q(np,nq)$) given in Eq.1:

$$E = p\sum_{p} D(p, np) + \sum_{p,\ q} V_{p,q}(np, nq) \quad (1)$$

Where (p,np) is the posterior probability of pixel p belonging to class np, represented as a confidence score vector of size p × n. $Vp,(np,nq)$ is the smoothness cost, which assigns the same labels to neighbouring pixels to maximise the gradient between differing classes. This matrix was of size n × n.

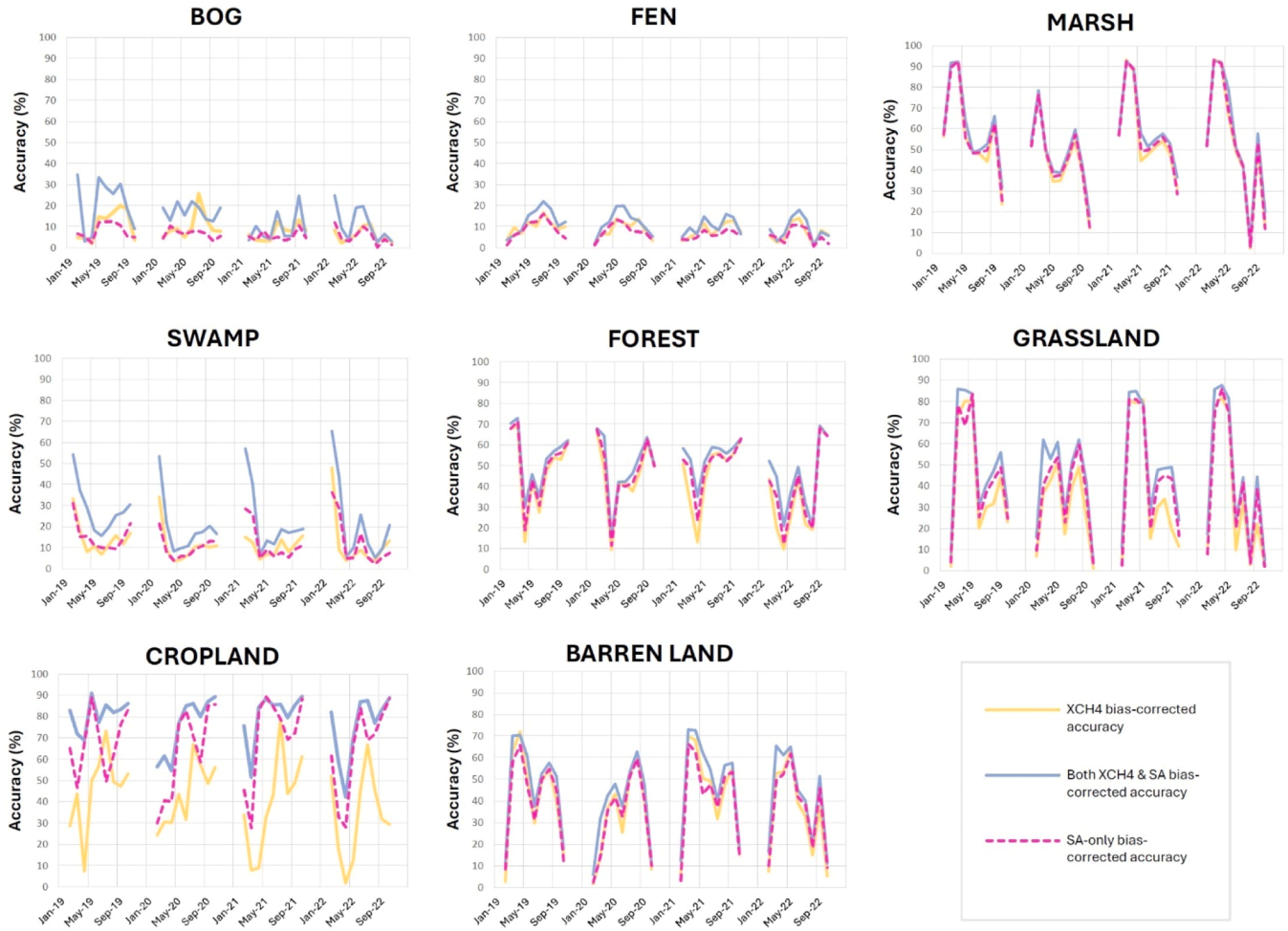

Fig. 5. Time series plots of class accuracies (in %) achieved for different land type classification using proposed methodology using S-5p bias-corrected XCH4 (yellow), bias-corrected XCH4 + SA (blue) and surface albedo (SA – dashed pink) for each land type over the time period Jan 2019 – Dec 2022 (x-axis).

For temporal mapping, using the segmentation model mentioned above, every pixel under AOI was mapped at least once every month. Therefore, a total of 1461 daily maps were created for the years 2019 - 2022. It has to be noted that this study does not use conventional time-series analysis. Here, every image was treated individually with equal realisation, i.e., all the images under consideration had equal importance. The majority voting was done for each pixel in the daily maps to create the monthly maps using Eq.2 (Jimenez et al. [17]), i.e., for every pixel $p \epsilon N$ a class $x \epsilon n$ would be assigned if,

$$\sum_{p=1}^{N} \widehat{F_p}(x) \quad = \quad \max_{x=1}^{n} \sum_{p=1}^{N} F_p(x) \tag{2}$$

where N are the total number of pixels, and $\widehat{F_p}(x)$ is the majority voted map at the end of each month for the years. Pixels that were not mapped for any given day for that month were removed. Furthermore, only the covered/mapped area was used for further accuracy analysis for each land type (class). Class Accuracy (CA) is the ratio of the diagonal vector of the class under consideration for true positives (TP) (i.e. the total number of pixels belonging to the same class) and false negatives (FN), shown in Eq.3.

$$CA = \frac{TP}{TP + FN} \tag{3}$$

The monthly high-confidence maps were again combined (using majority voting) to form the final aggregated map for a year, which has been used for performance evaluation of the algorithm. This map gives an idea of the difference in land types and opens an area of application of S-5p products for this purpose. For every year, only the pixels with CA$\geq$ 55 % were selected to form a high-confidence yearly aggregated map; classes like bog, fen, deep water, shallow water were omitted in these maps due to low CA values.

The accuracy of areal detection should not only be checked by typical pixel-based (PB) evaluation metrics such as producer and user accuracy [30]. For change detection, PB comparison may not provide the true extent of error; therefore, segment-based comparison based on the area's geometry was needed [5]. Hence, a set of error metrics linked with location and extent of land type detection was calculated for the annual maps for both years while comparing it with the GT of the captured region [5].

3.3. Error metric

The set of error metrics specific to measuring spatial changes to identify differences between CH_4 identification, used in this study, is described below. The spatial error metrics treat each class as a segment, and accuracy is determined using the intersection of the segments with the ground truth. The segment-based approach improves the interpretation of the data, as opposed to the conventional pixel-based confusion matrix, by computing error diagnostics based on the distance of the segments. In reality, the land type can be changed/shifted/grown. This change is penalised heavily in a pixel-based approach, whereas, in a segment-based approach, such changes are weighted and can be forfeited based on the amount of change and interpretation of it.

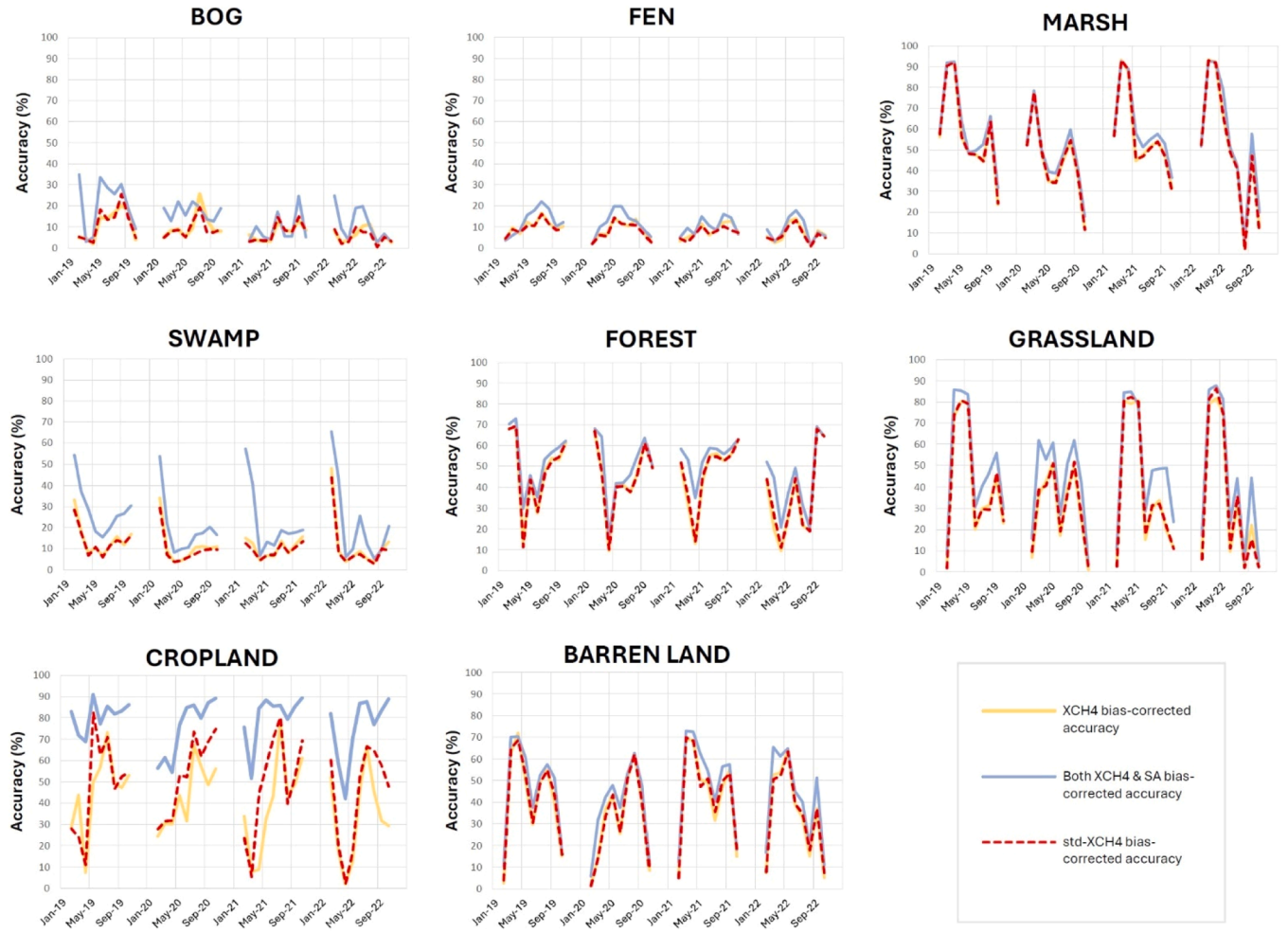

Fig. 6. Time series plots of class accuracies (in %) achieved for different land types classification using proposed methodology using S-5p bias-corrected XCH4 (yellow), bias-corrected XCH4 + SA (blue) and standard XCH4 (std-XCH4 – dashed red) for each land type over the time period Jan 2019 – Dec 2022 (x-axis).

Table 1
Geometrical error metric and accuracies for the union of the area covered in aggregated maps of 2019/20/21/22 (majority voted across year). The % Δ signifies the absolute change in the parameter value being identified in comparison to the original value of the parameter in the GT. ΔArea signifies the change in the geometrical area. Producer and User accuracy on the GT made using [1].

Aggregated Maps (2019/20/21/22) Land types	Identified region Area (1000 km^2)	Yr. 2019/20/21/22 % ΔArea	Yr. 2019/20/21/22	
			Producer Accuracy (%)	User Accuracy (%)
MARSH	602.22	19.5 /16.4 /24.5 /32.6	96.2 / 95.3 / 96.8 / 97.0	77.4 / 77.7 / 74.4 / 69.9
SWAMP	131.63	44.8 / 52.8 / 42.1 / 44.6	50.8 / 44.3 /52.0 / 52.2	92.0 / 93.8 / 89.2 / 92.9
FOREST	544.98	3.8 / 0.3 / 12.3 / 13.7	97.7 / 98.4 / 95.6 / 94.3	79.2 / 75.8 / 83.9 / 84.1
GRASSLAND	354.01	4.7 / 8.9 / 5.1 / 3.1	89.3 / 87.7 / 88.7 / 88.5	93.0 / 93.1 / 92.1 / 90.2
CROPLAND	75.81	1.1 / 0.3 / 5.8 / 6.1	87.6 / 89.1 / 89.5 / 90.0	86.5 / 88.4 / 84.1 / 84.4
BARREN LAND	334.86	20.1 / 27.4 / 19.4 / 27.0	87.4 / 83.0 / 88.3 / 81.5	94.2 / 95.4 / 93.3 / 95.2

- Jaccard Similarity Index (J) measures the similarity between the members of the two sets and reports the amount of similarity and distinction [26].
- Area (A) estimates the total area of the selected land type. The area of every individual pixel is determined by looking at its 2 × 2 neighbourhood. Each pixel is part of four different 2 × 2 neighbourhoods, which indicates the change in the overall growth/shrinkage of the community.
- Orientation (O) gives the angle between the x-axis and the ellipse's major axis (covering the entire land type). It can range from −90 to + 90 degrees, indicating the direction of the land type change.
- Extent (E) indicates the ratio of total pixels present in the bounding box to the total pixels present in the image. The bounding box represents a box (rectangle/square) covering the major cluster of pixels present for a land type in an image.

4. Results and discussion

For understanding the relationship between land-type emissions, we used S-5p products for each month to deduce presence of sensitivity in each season. The S-5p CH_4 total column and SA from gridded pixels were analysed together and separately using a classification-segmentation algorithm for each available day. Algorithm steps, training data (30 %) and all details were described previously in Section 3. The analysis generated daily maps from the testing data (70 %) showing the extent of 8 different land types over 365 days in 2019, 2021, 2022 and 366 days in 2020. These values were available for all key land types as

Table 2
Confusion Matrix for 2019–2022. The x-axis describes the Predicted Class, and the y-axis describes the True Class.

	BOG	FEN	MARSH	SWAMP	FOREST	GRASS LAND	CROP LAND	BARREN LAND
	2019							
BOG	521	0	2609	234	1296	0	4	0
FEN	0	493	10579	156	5119	0	0	36
MARSH	6	14	106991	107	2448	1403	0	181
SWAMP	0	0	3340	15246	11376	0	0	15
FOREST	29	7	1539	451	107608	2	417	62
GRASS LAND	0	0	2400	33	597	60815	1410	2792
CROP LAND	3	0	4	0	1669	9	14052	287
BARREN LAND	0	0	3872	5	3314	2215	288	67337
	2020							
BOG	790	0	2162	72	1608	0	1	0
FEN	32	1266	9050	323	5521	0	0	14
MARSH	8	51	104587	103	3872	935	5	125
SWAMP	1	4	2033	13282	14587	0	0	11
FOREST	8	12	1084	130	107415	10	369	70
GRASS LAND	0	1	3761	28	967	58191	1174	2212
CROP LAND	0	0	0	3	1411	9	14235	303
BARREN LAND	0	13	6422	1	3526	2460	256	61995
	2021							
BOG	362	0	2792	31	1290	1	0	0
FEN	0	566	12154	337	3159	0	0	136
MARSH	4	12	107104	87	1570	1518	0	283
SWAMP	1	12	4884	15500	9255	0	0	101
FOREST	11	12	2808	851	103848	14	790	286
GRASS LAND	0	0	2940	34	371	59965	1534	2758
CROP LAND	2	0	14	0	1294	24	14291	326
BARREN LAND	0	1	4014	16	1977	2619	265	67352
	2022							
BOG	324	0	2467	47	1705	1	0	0
FEN	0	376	12634	267	3046	5	0	48
MARSH	1	1	107323	105	1568	1495	0	149
SWAMP	11	19	6381	15450	7700	4	0	25
FOREST	10	1	4816	574	102417	21	618	105
GRASS LAND	0	0	3451	13	586	59978	1647	2060
CROP LAND	0	0	22	9	1260	20	14373	282
BARREN LAND	0	2	6935	11	2797	4046	316	62177

classified in CWI map [1]. Area covered by snow, as obtained using MODIS snow product [13], interfered with the capture and visibility of land types decreased in all cases with increased snow cover (Fig. 4).

The time-series of the class accuracy (CA) values for each land type calculated compared to the GT maps are shown in Figs. 5 and 6, along with the SA for the respective land types in Fig. 4. Taken together, the figures show that accuracy is affected by S-5p coverage and changes drastically even when the SA is almost constant throughout the years. The % accuracy depicts the CA (Eq. 3), and the SA (grey line) is the spatial mean for each class each day. Using the example of Bog land type, we can see the SA remains approximately constant over the year (which is expected). However, there is varying accuracy when using SA as a variable in the classification algorithm (blue line). This therefore shows some independent effect of methane on the land-type variability. The cropland class, in contrast, has a more variable SA through the years – this is also due to the presence of higher levels of snow in the area and eventual residuals. The addition of SA to methane, in this case, definitely complements the land-type analysis. However, looking at the variation in the methane-only accuracy of the algorithm (yellow line), such high dips also concur with our previously stated speculation. The performance metrics are further discussed in detail in the following paragraphs.

The performance of the algorithm for detecting area of each land type showed the sensitivity of S-5p CH_4 total column and SA to the land type in 2019, 2020, 2021, and 2022 as presented in Table 1, and the confusion matrix in Table 2. Table 1 depicts the six major land types. Since the overall CA of bog and fen were less than 55 %; these land types were not considered in the final creation of annual maps (Fig. 7). Table 2 (confusion matrix) gives pixels for all eight land types for better understanding. The CA improved slightly (~2 %) with the inclusion of SA data for most land types, and ~10 % for cropland compared to the analysis considering only S-5p CH_4 data (Fig. 5).

Using S-5p products, the land types with large areal extent, such as marsh, forest, grassland, swamp, and cropland, showed high detectability (CA>60 %), while wetland types such as bog and fen showed low CAs due to low areal extent and proximity to other dominant land types. Bog, fen, and swamp were often misclassified due to their intermixed land distribution. In the winter months, CA decreased due to the lack of S-5p data and were omitted from the plots. Marsh was detected with the highest CA, with variations in accuracy linked with a lack of available pixels. This might be expected due to marshes being significant emitters of methane – in particular, the wetland land types (marsh, bogs and fen) due to the anaerobic conditions within the highly organic waterlogged environment as well as the presence of certain aerenchymatic or "shunt-species" vegetation enabling gases generated beneath the surface to be vented / released directly to the atmosphere. Similarly, for grassland and barren-land, CAs were linked with the area covered by S-5p, with grassland showing better detectability. The other key wetland-type, swamp, showed better detectability in spring and autumn than summer when it was misclassified as forest during the growing season. Forest was detected with reasonable OA, which slightly reduced with the melting of snow cover. The inclusion of SA improved the detectability of cropland significantly. The detectability of most dominant land types utilising CH_4 data thus indicates a significant difference and sensitivities of CH_4 emissions between land types.

The seasonal variations of the CH_4 data and SA from land types are illustrated by delineating different land types during different seasons of 2019 (Fig. 7a), 2020 (Fig. 7c), 2021 (Fig. 7e) and 2022 (Fig. 7 g). Land types detected with all CA values were plotted in monthly maps, with any missing or non-detectable pixels shown as white in Fig. 7. The sensitivities of the method for distinguishing between marsh, swamp, grassland, and barren-land were strongest during March-May, due to the

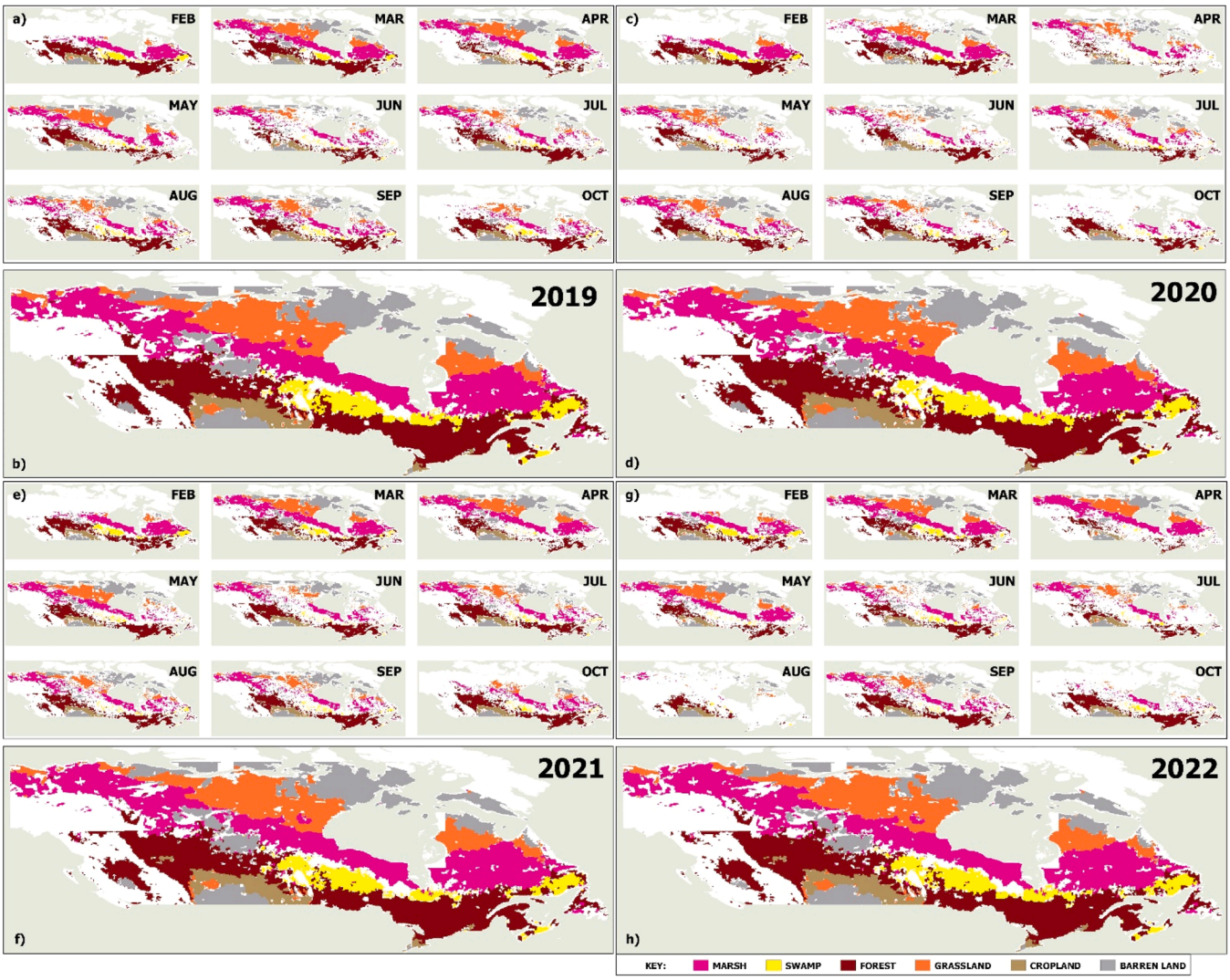

Fig. 7. Land type maps created using CH4 and SA data as input. a/c/e/g, 2019/20/21/22 – seasonal classified maps created using combining the daily images obtained from S-5p, the missing area (white/blank) was not covered by S-5p for that month. b/d/f/h, 2019/20/21/22 – aggregated maps created using the pixels with class accuracy $\geq$ 55 % over the months (majority voted) for each year separately. These represent the area with high confidence for 2019–2022. The maps were generated using Matlab v.2019b software. The boundary of the map was taken from the open-source website https://www.igismap.com/canada-shapefile-download-free-adminstrative-boundaries-provinces-and-territories/ (last accessed on 15 June 2020).

warming soil (coming out of very cold winters in this part of the world) activating the anaerobic microbiological metabolic pathways in the spring in the wetlands generating a more significant and distinguishable methane signal at that time of year compared to the other more aerobic soils in the other land types.

The yearly maps for 2019–2022 were created using majority voting of the monthly maps during the calendar year, only including land types with CA> 55 % (Fig. 7b,d,f&h). The CA was calculated considering the classified map provided by Amani et al. [1] and not validated against independent field assessment (due to unavailability of this information). The areal extent of marsh, forest, and grassland was identified with high confidence (Figs. 5 and 6). The land types with large areal extent were generally detected well with high accuracy. The bog and fen wetlands with lower areal extent were misclassified as marsh (Table 2). Similarly, some pixels in the swamp were misclassified as forest during summer periods (June-August). This is mainly due to the proximity of the land types, which leads to pixel-mixing effects at this coarse resolution. Land types other than wetland (cropland and barren-land) were best identified in summer with good boundary delineation, and grassland, although adjacent to marsh, was well distinguishable throughout the year, indicating the sensitivity to the difference CH_4 for these land types.

It is possible that SA at 2.3 μm is particularly sensitive to some land types, leading to enhancement of identification (additional tests done using just SA information are shown in Fig. 5). Also, as seen in Fig. 4, the land types show very small changes in the SA retrieved in the 2.3 μm over time during the year.

In Fig. 5, the dashed pink line shows the SA case, plotted along with the bias-corrected XCH4 case (yellow line) and the bias-corrected XCH4 + SA case (blue line). When we use CH_4 and SA combined, we see the best performance in identifying the sensitivities of land types as in the proposed algorithm. Therefore, we find that the best choice is the combined usage of the bias-corrected XCH4 + SA case. In addition, the effect of the bias-correction applied to the XCH4 data was seen to be similar to standard XCH4 (std-XCH4) data for most land types is shown in Fig. 6, where the dashed red line shows the std-XCH4 case, plotted against the same bias-corrected XCH4 and bias-corrected XCH4 + SA cases in the same colours. The largest difference in accuracies between the standard and the bias-corrected XCH4 runs can be seen for cropland. This is because the maximum change in SA was observed for cropland; the other land types show smaller changes in the SA retrieved in the 2.3 μm over time during the year (Fig. 4). Therefore, the effect of the bias-correction applied to the XCH4 data is similar for those land types

Table 3
Geometrical error metric and accuracies for the union of the area covered in aggregated maps of 2018, 2019 (confident land types). Jaccard represents the 2D similarity when maps are overlapped directly. ΔExtent change in extent, and ΔOrientation change in orientation.

Aggregated Maps (2019/20/21/22)	Present study				
	2019/20/21/22	Identified region	2019/20/21/22	Identified region	2019/20/21/22
Land types	Jaccard	Orientation (°)	% ΔOrientation	Extent	% ΔExtent
MARSH	0.73/0.72/0.70/0.66	−11.24	0.55/0.36/0.50/0.56	0.148	−0.04/−0.92/0.10/−1.24
SWAMP	0.49/0.43/0.49/0.50	−5.49	−1.96/−0.49/−2.32/−1.35	0.073	3.57/3.54/3.73/4.17
FOREST	0.64/0.61/0.65/0.64	−8.01	1.36/0.87/1.29/1.17	0.167	0.74/1.77/4.02/3.34
GRASSLAND	0.83/0.80/0.81/0.80	−9.23	0.96/0.67/1.05/1.12	0.099	−0.17/0.76/2.80/0.97
CROPLAND	0.77/0.80/0.76/0.77	−15.19	0.45/0.24/−0.09/−0.29	0.032	−1.29/1.41/0.20/−0.45
BARREN LAND	0.72/0.67/0.71/0.67	2.86	−2.17/−2.55/−3.80/−3.36	0.101	2.30/2.95/2.18/3.11

with similar SA conditions.

The differences in the sensitivities based on S-5p CH_4 total column combined with SA, for different land-types, was carried out utilising the proposed machine learning (ML) algorithm, where the efficiency of detection was investigated using a set of areal error metrics. Jaccard similarity index (J), area (A), orientation (O) and extent (E) (Table 3). Good detection was seen for all six key land types, while the variability of these metrics mainly was attributed to the lack of availability of S-5p pixels, which were often due to inimical meteorological conditions.

Lastly, this study aims to highlight the potential of S-5p products for understanding the variation in land types. Due to the absence of in situ measurements, the study could not verify the correlation of sensitivities of the land types as seen from space and actual ground coverage. The other limitation was the lack of field measurements in terms of land cover. This study is highly dependent on prior knowledge about the ground truth/land type locations. Furthermore, at least coarse information of land type location is essential to conclude the sensitivity variations present. Whilst this methodology presented does not actually quantify the methane fluxes from the different land types, with ever more focus and field studies globally on greenhouse gas emissions from different land types providing ground truth data, as well as advancing knowledge about atmospheric physics, it may soon be possible to fractionate the remotely-sensed net CH_4 signal via further modelling to be able to start to differentiate CH_4 emissions between the land types with more confidence using this product.

5. Conclusions

This work demonstrates that Sentinel-5p products with a machine learning algorithm can reveal unique sensitivity to certain land types, especially between large areas of marsh, forest and grassland. However, due to the coarse resolution of the satellite, the methodology was not able to distinguish between the land types smaller than 10 % of the total land area (such as fens, bogs). Additionally, the unavailability of data during winter months posed a significant constraint, limiting the temporal coverage and potentially affecting the accuracy of the seasonal analysis. By analysing such CH_4 data along with derived surface albedo, the areal extents of six land types (following CWI), including two major wetland types (marsh and swamp) covering ~60 % of the total wetland area of Canada, were identified for four consecutive years 2019–2022. We also establish that there is some correlation of land-types with CH_4 total column data, independently, irrespective of high-dependence of surface albedo on the land-type. The CWI generated using S-5p data in this study is complementary to the traditional methods of land type identification showing daily, monthly, seasonal, and yearly changes. These maps can be used by the WAD2M (Wetland Area Dynamics for Methane Modeling) to either verify or complement their data where measurements from other sources are not available. Despite the coarse resolution, the study reveals the subtle variations in total column methane based on different land-types and lays a foundation for better sensitivity analysis using higher spatial-resolution datasets. Such higher resolution imagery alongside better ground truth monitoring could lead to a being able to differentiate and quantify methane emissions between land types using satellite data with increased confidence, as well as determining whether significant changes are occurring over time. This would yield valuable insights for climate scientists and policy makers at both national and international levels.

Code availability

The classification [4] code used for the study can be accessed via the GitHub repository https://github.com/saheba92/Mapping-vegetation-communities.

CRediT authorship contribution statement

Ghosh Bidisha: Writing – review & editing, Supervision, Methodology, Conceptualization. **Langerock Bavo:** Writing – review & editing, Supervision. **Gill Laurence:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Silva Mariana:** Writing – review & editing, Formal analysis, Data curation. **Sha Mahesh Kumar:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Bhatnagar Saheba:** Writing – original draft, Investigation, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M. Amani, S. Mahdavi, M. Afshar, B. Brisco, W. Huang, S. Mohammad Javad Mirzadeh, L. White, S. Banks, J. Montgomery, C. Hopkinson, Canadian wetland inventory using google earth engine: The first map and preliminary results, Remote Sens. 11 (7) (2019) 842, https://doi.org/10.3390/rs11070842.
[2] Apituley, A., Pedergnana, M., Sneep, M., Veefkind, J.P., Loyola, D. and Hasekamp, O. Sentinel-5 precursor/TROPOMI level 2 product user manual methane. Source: SRON, 2017.
[3] J. Barré, I. Aben, A. Agustí-Panareda, G. Balsamo, N. Bousserez, P. Dueben, R. Engelen, A. Inness, A. Lorente, J. McNorton, V.-H. Peuch, G. Radnoti, R. Ribas, Systematic detection of local CH4 anomalies by combining satellite measurements with high-resolution forecasts, Atmos. Chem. Phys. 21 (2021) 5117–5136, https://doi.org/10.5194/acp-21-5117-2021.
[4] S. Bhatnagar, L. Gill, S. Regan, O. Naughton, P. Johnston, S. Waldren, B. Ghosh, Mapping vegetation communities inside wetlands using Sentinel-2 imagery in

Ireland, Int. J. Appl. Earth Obs. Geoinf. 88 (2020) 102083, https://doi.org/10.1016/j.jag.2020.102083.
[5] S. Bhatnagar, L. Gill, S. Regan, S. Waldren, B. Ghosh, A nested drone-satellite approach to monitoring the ecological conditions of wetlands, ISPRS J. Photogramm. Remote Sens. 174 (2021) 151–165, https://doi.org/10.1016/j.isprsjprs.2021.01.012.
[6] T.J. Bohn, J.R. Melton, A. Ito, T. Kleinen, R. Spahni, B.D. Stocker, B. Zhang, X. Zhu, R. Schroeder, M.V. Glagolev, S. Maksyutov, WETCHIMP-WSL: intercomparison of wetland methane emissions models over West Siberia, Biogeosciences 12 (2015) 3321–3349, https://doi.org/10.5194/bg-12-3321-2015.
[7] L. Collins, G. McCarthy, A. Mellor, G. Newell, L. Smith, Training data requirements for fire severity mapping using Landsat imagery and random forest, Remote Sens. Environ. 245 (2020) 111839, https://doi.org/10.1016/j.rse.2020.111839.
[8] N.T. Comer, P.M. Lafleur, N.T. Roulet, M.G. Letts, M. Skarupa, D. Verseghy, A test of the Canadian land surface scheme (class) for a variety of wetland types, Atmosphere-Ocean 38 (1) (2000) 161–179, https://doi.org/10.1080/07055900.2000.9649644.
[9] A.R. Desai, K. Xu, H. Tian, P. Weishampel, J. Thom, D. Baumann, A.E. Andrews, B. D. Cook, J.Y. King, R. Kolka, Landscape-level terrestrial methane flux observed from a very tall tower, Agric. For. Meteorol. 201 (2015) 61–75, https://doi.org/10.1016/j.agrformet.2014.10.017.
[10] Didan, K. (2015). MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020–9-1 from https://doi.org/10.5067/MODIS/MOD13A3.006.
[11] Environment and Climate Change Canada (2016) Canadian Environmental Sustainability Indicators: Extent of Canada's Wetlands. Consulted on 04 October, 2020. Available at ⟨www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=69E2D25B-1⟩.
[12] J. Hachmeister, O. Schneising, M. Buchwitz, A. Lorente, T. Borsdorff, J.P. Burrows, J. Notholt, M. Buschmann, On the influence of underlying elevation data on Sentinel-5 Precursor TROPOMI satellite methane retrievals over Greenland, Atmos. Meas. Tech. 15 (2022) 4063–4074, https://doi.org/10.5194/amt-15-4063-2022.
[13] Hall, D.K., G.A. Riggs & V.V. Salomonson. 2006. MODIS/Terra Snow Cover 5-Min L2 Swath 500m, Version 5. [MOD10A1]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. Accessed 2020–9-1 from https://doi.org/10.5067/ACYTYZB9BEOS.
[14] L.I. Harris, K. Richardson, K.A. Bona, S.J. Davidson, S.A. Finkelstein, M. Garneau, J. McLaughlin, F. Nwaishi, D. Olefeldt, M. Packalen, N.T. Roulet, F.M. Southee, M. Strack, K.L. Webster, S.L. Wilkinson, J.C. Ray, The essential carbon service provided by northern peatlands, Front. Ecol. Environ. (2021), https://doi.org/10.1002/fee.2437.
[15] H. Hu, O. Hasekamp, A. Butz, A. Galli, J. Landgraf, J. Aan de Brugh, T. Borsdorff, R. Scheepmaker, I. Aben, The operational methane retrieval algorithm for TROPOMI, Atmos. Meas. Tech. 9 (2016) 5423–5440, https://doi.org/10.5194/amt-9-5423-2016.
[16] H. Hu, J. Landgraf, R. Detmers, T. Borsdorff, J. Aan de Brugh, I. Aben, A. Butz, O. Hasekamp, Toward global mapping of methane with TROPOMI: First results and intersatellite comparison to GOSAT, Geophys. Res. Lett. 45 (8) (2018) 3682–3689, https://doi.org/10.1002/2018GL077259.
[17] L.O. Jimenez, A. Morales-Morell, A. Creus, Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks, IEEE Trans. Geosci. Remote Sens 37 (1999) 1360–1366. https://ieeexplore.ieee.org/document/763300.
[18] Lambert, J.-C., A. Keppens, S. Compernolle, K.-U. Eichmann, M. de Graaf, D. Hubert, B. Langerock, A. Ludewig, M.K. Sha, T. Verhoelst, T. Wagner, C. Ahn, A. Argyrouli, D. Balis, K.L. Chan, M. Coldewey-Egbers, I. De Smedt, H. Eskes, A.M. Fjæraa, K. Garane, J.F. Gleason, F. Goutail, J. Granville, P. Hedelt, C. Ahn, K.-P. Heue, G. Jaross, Q. Kleipool, M.L. Koukouli, R. Lutz, M.C. Martinez Velarte, K. Michailidis, S. Nanda, S. Niemeijer, A. Pazmiño, G. Pinardi, A. Richter, N. Rozemeijer, M. Sneep, D. Stein Zweers, N. Theys, G. Tilstra, O. Torres, P. Valks, J. van Geffen, C. Vigouroux, P. Wang, and M. Weber. S5P MPC Routine Operations Consolidated Validation Report series, Issue #19, Version 19.01.00, 196 pp., 3 July 2023.
[19] Langerock, B. & Sha, M.K.. MPC + TCCON4S5P Validation Presentation, Sentinel-5P Third products Release Workshop. Available at ⟨https://earth.esa.int/documents/247904/3753563/S5PCH4-VAL-MPC-TCCON4S5P⟩, 2019.
[20] A. Lorente, T. Borsdorff, M.C. Martinez-Velarte, J. Landgraf, Accounting for surface reflectance spectral features in TROPOMI methane retrievals, Atmos. Meas. Tech. 16 (2023) 1597–1608, https://doi.org/10.5194/amt-16-1597-2023.
[21] J.R. Melton, R. Wania, E.L. Hodson, B. Poulter, B. Ringeval, R. Spahni, T. Bohn, C. A. Avis, D.J. Beerling, G. Chen, A.V. Eliseev, Present state of global wetland extent and wetland methane modelling: conclusions from a model intercomparison project (WETCHIMP), Biogeosciences 10 (2013) 753–788, https://doi.org/10.5194/bg-10-753-2013.
[22] K. Millard, M. Richardson, On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping, Remote Sens. 7 (7) (2015) 8489–8515, https://doi.org/10.3390/rs70708489.
[23] National Wetlands Working Group (1997) The Canadian Wetland Classification System. 2nd edition. Edited by BG Warner and CDA Rubec. Wetlands Research Centre, University of Waterloo, Waterloo, Ontario,1997.
[24] B. Poulter, P. Bousquet, J.G. Canadell, P. Ciais, A. Peregon, M. Saunois, V.K. Arora, D.J. Beerling, V. Brovkin, C.D. Jones, F. Joos, Global wetland contribution to 2000–2012 atmospheric methane growth rate dynamics, Environ. Res. Lett. 12 (2017) 094013, https://doi.org/10.1088/1748-9326/aa8391.
[25] M.J. Prather, C.D. Holmes, J. Hsu, Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, L09803, Geophys. Res. Lett. 39 (2012), https://doi.org/10.1029/2012GL051440.
[26] R. Real, J.M. Vargas, The Probabilistic Basis of Jaccard's Index of Similarity, Syst. Biol. 45 (1996) 380–385, https://doi.org/10.1093/sysbio/45.3.380.
[27] M. Saunois, A.R. Stavert, B. Poulter, P. Bousquet, J.G. Canadell, R.B. Jackson, P. A. Raymond, E.J. Dlugokencky, S. Houweling, P.K. Patra, P. Ciais, The Global Methane Budget 2000–2017, Earth Syst. Sci. Data 12 (2020) 1561–1623, https://doi.org/10.5194/essd-12-1561-2020.
[28] O. Schneising, M. Buchwitz, J. Hachmeister, S. Vanselow, M. Reuter, M. Buschmann, H. Bovensmann, J.P. Burrows, Advances in retrieving XCH4 and XCO from Sentinel-5 Precursor: improvements in the scientific TROPOMI/WFMD algorithm, Atmos. Meas. Tech. 16 (2023) 669–694, https://doi.org/10.5194/amt-16-669-2023.
[29] M.K. Sha, B. Langerock, J.-F.L. Blavier, T. Blumenstock, T. Borsdorff, M. Buschmann, A. Dehn, M. De Mazière, N.M. Deutscher, D.G. Feist, O.E. García, D. W.T. Griffith, M. Grutter, J.W. Hannigan, F. Hase, P. Heikkinen, C. Hermans, L. T. Iraci, P. Jeseck, N. Jones, R. Kivi, N. Kumps, J. Landgraf, A. Lorente, E. Mahieu, M.V. Makarova, J. Mellqvist, J.-M. Metzger, I. Morino, T. Nagahama, J. Notholt, H. Ohyama, I. Ortega, M. Palm, C. Petri, D.F. Pollard, M. Rettinger, J. Robinson, S. Roche, C.M. Roehl, A.N. Röhling, C. Rousogenous, M. Schneider, K. Shiomi, D. Smale, W. Stremme, K. Strong, R. Sussmann, Y. Té, O. Uchino, V.A. Velazco, M. Vrekoussis, P. Wang, T. Warneke, T. Wizenberg, D. Wunch, S. Yamanouchi, Y. Yang, M. Zhou, Validation of methane and carbon monoxide from Sentinel-5 Precursor using TCCON and NDACC-IRWG stations, Atmos. Meas. Tech. 14 (9) (2021), https://doi.org/10.5194/amt-14-6249-2021.
[30] M. Story, R.G. Congalton, Accuracy assessment: a user's perspective, Photogramm. Eng. Remote Sens. 52 (1986) 397–399.
[31] A. Stumpf, N. Lachiche, J.P. Malet, N. Kerle, A. Puissant, Active learning in the spatial domain for remote sensing image classification, IEEE Trans. Geosci. Remote Sens. 52 (5) (2013) 2492–2507, https://doi.org/10.1109/TGRS.2013.2262052.
[32] M.M. Swenson, S. Regan, D.T. Bremmers, J. Lawless, M. Saunders, L.W. Gill, Carbon balance of a restored and cutover raised bog: implications for restoration and comparison to global trends, Biogeosciences 16 (2019) 713–731, https://doi.org/10.5194/bg-16-713-2019.
[33] D.W. Valentine, E.A. Holland, D.S. Schimel, Ecosystem and physiological controls over methane production in northern wetlands, J. Geophys. Res. 99 (1994) 1563–1571, https://doi.org/10.1029/93JD00391.
[34] J.P. Veefkind, I. Aben, K. McMullan, H. Förster, J. De Vries, G. Otter, J. Claas, H. J. Eskes, J.F. De Haan, Q. Kleipool, M. Van Weele, 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 Sens. Environ. 120 (2012) 70–83, https://doi.org/10.1016/j.rse.2011.09.027.
[35] R. Wania, I. Ross, I.C. Prentice, Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1, Geosci. Model Dev. 3 (2010) 565–584, https://doi.org/10.5194/gmd-3-565-2010.
[36] A.J. Webb, H. Bösch, R.J. Parker, L.V. Gatti, E. Gloor, P.I. Palmer, L.S. Basso, M. P. Chipperfield, C.S. Correia, L.G. Domingues, L. Feng, CH4 concentrations over the Amazon from GOSAT consistent with in situ vertical profile data, 11-006, J. Geophys. Res.: Atmospheres 121 (18) (2016), https://doi.org/10.1002/2016JD025263.
[37] S.C. Whalen, Biogeochemistry of Methane Exchange between Natural Wetlands and the Atmosphere, Environ. Eng. Sci. 22 (2005) 73–94, https://doi.org/10.1089/ees.2005.22.73.
[38] Z. Yu, D.W. Beilman, S. Frolking, G.M. MacDonald, N.T. Roulet, P. Camill, D. J. Charman, Peatlands and their role in the global carbon cycle, Eos, Trans. Am. Geophys. Union 92 (2011) 97–108, https://doi.org/10.1029/2011EO120001.