



Contents lists available at ScienceDirect

# International Journal of Applied Earth Observation and Geoinformation

journal homepage: [www.elsevier.com/locate/jag](http://www.elsevier.com/locate/jag)



## Leveraging google earth engine cloud computing for large-scale arctic wetland mapping

Michael Merchant <sup>a,b,\*</sup>, Brian Brisco <sup>b</sup>, Masoud Mahdianpari <sup>c,d</sup>, Laura Bourgeau-Chavez <sup>e</sup>, Kevin Murnaghan <sup>b</sup>, Ben DeVries <sup>a</sup>, Aaron Berg <sup>a</sup>

<sup>a</sup> Department of Geography, Environment and Geomatics, University of Guelph, Guelph, ON N1G 2W1, Canada

<sup>b</sup> The Canada Center for Mapping and Earth Observation, Natural Resources Canada, Ottawa, ON K1A 0E4, Canada

<sup>c</sup> C-CORE, 1 Morrissey Rd, St. John's, NL A1B 3X5, Canada

<sup>d</sup> Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada

<sup>e</sup> Michigan Tech Research Institute, Michigan Technological University, Ann Arbor, MI 48105, USA

### ARTICLE INFO

#### Keywords:

Arctic  
Google earth engine  
Machine learning  
Wetlands

### ABSTRACT

Climate-driven permafrost degradation and an intensification of the hydrological cycle are rapidly altering the intricate ecohydrological processes of Arctic wetlands, threatening their long-term carbon sequestration capabilities. Addressing this concern through effective management holds immense potential for climate regulation, mitigation, and adaptation efforts. As such, there is growing need for timely spatial inventory data identifying Arctic wetlands with sufficient accuracy, resolution, and detail. Wetland mapping at large scales necessitates the processing of large volumes of Earth observation (EO) data, a challenge known as “Big Data”. Consequently, in this study, we present a cloud-based methodology exploiting the remarkable collection of EO data and computational power of Google Earth Engine (GEE) to map Arctic wetlands at 10 m spatial resolution. Our workflow evaluated temporally aggregated optical and radar satellite imagery and novel hydro-physiographic layers as inputs into a robust Random Forest (RF) machine learning (ML) algorithm. Both pixel and object-based classification approaches were assessed, whereby ML models were calibrated with a training dataset of sufficient and comprehensive samples. The study was conducted over Canada’s Southern Arctic ecozone ( $830,000 \text{ km}^2$ ). GEE enabled the efficient preprocessing and classification of large volumes of EO data and resulted in excellent yet similar statistical performance for both pixel and object-based approaches, achieving overall accuracies of  $> 89\%$  and mean F1-scores of  $> 0.79$ . Moreover, McNemar tests indicated that these classifications were not statistically different, which has significant implications regarding computing time and processing efficiencies. These results demonstrate the efficacy and scalability of our cloud-based GEE methodology, and as such can support future endeavors around Pan-Arctic wetland mapping and monitoring.

### 1. Introduction

Wetlands are vital ecosystems, providing an abundance of ecological goods and services, including flood mitigation, nutrient cycling, erosion control, greenhouse gas sequestration, and water filtration. Wetlands also help maintain biodiversity, providing habitat for iconic flora and fauna. Across Canada’s Arctic biome, seasonally or permanently waterlogged wetlands frequently dominate the landscape, storing billions of tonnes of carbon and linking numerous ecohydrological and permafrost-related processes (Bring et al., 2016). However, these processes are rapidly transforming (Seifollahi-Aghmuni et al., 2019), a

result of the Arctic warming twice as fast as the global average and an intensification of the hydrological cycle (Cohen et al., 2014). Temperature-driven impacts may alter the biogeochemical cycling of Arctic wetlands and shift them from a carbon sink to source, threatening their role in climate regulation (Kåresdotter et al., 2021). Concerns about the fate of this vast carbon pool, and in the large stocks found in frozen peat deposits, have been expressed for some time due to the permafrost carbon feedback (PCF) – an exacerbation of global warming from thawing permafrost and the release of CO<sub>2</sub> or CH<sub>4</sub> (Schuur et al., 2015). Changes to the Arctic water cycle may disrupt inundation opportunities from enhanced evapotranspiration, lowering of water tables,

\* Corresponding author at: Department of Geography, Environment and Geomatics, University of Guelph, Guelph, ON N1G 2W1, Canada.

E-mail address: [mmerchan@uoguelph.ca](mailto:mmerchan@uoguelph.ca) (M. Merchant).

and wetland desiccation (Woo & Young, 2005). Consequently, the accurate and timely mapping of Arctic wetlands has high ecological relevance in the face of climate change.

Recently, there have been numerous developments in Earth observation (EO) and remote sensing (RS) technologies, resulting in a variety of high-quality datasets, consisting of different spatial, temporal, and radiometric resolutions. This rapid progress, in part due to the continuous launch of state-of-the-art satellites, has helped advance Arctic wetland mapping. For instance, spectral data from passive optical imagers have been used to map a variety of wet Arctic ecosystems (Zou et al., 2021). Parallel studies have demonstrated the value of active Synthetic Aperture Radar (SAR) sensors (Merchant et al., 2017), which are advantageous at high-latitudes due to their all-weather, daylight independent imaging capabilities. SAR microwaves are also particularly sensitive to vegetation canopy structure, moisture, and surface water conditions. While SAR and optical sensors each provide discriminatory information, literature advocates for the integration of both due to their complementarity (Mahdavi et al., 2019). Other studies have found topographic data further enhances fusion analysis (Long et al., 2021).

Despite these scientific accomplishments, there remains a need for consistent, high-resolution, and large-scale data classifying Canada's Arctic wetlands based on their hydrological, biological, and chemical attributes. This information gap is pronounced at high latitudes where in situ reference data is sparse, a result of Arctic environments being remote, and logistically and budgetarily challenging. Providentially, the progress made in achieving open access data policies for numerous EO datasets, together with the advent of cloud-based computing, has effectively addressed the challenge of RS analysis over Arctic regions. At the forefront of such innovative technology is Google Earth Engine (GEE), a cloud-computing platform providing petabytes of satellite data, machine learning (ML) algorithms, and planetary-scale processing capabilities (Tamiminia et al., 2020). The era of big data has brought upon both opportunities and challenges related to volume, velocity, and variety (Liu et al., 2018), and yet GEE, with its ability to store and analyze unprecedented volumes of geospatial data, has provided a practical solution for scientists seeking to map and monitor the Earth (Tamiminia et al., 2020). As a result, GEE has been used in many regional-scale studies, with applications such as forests, agriculture, and floods, to name a few (Del Valle & Jiang, 2022; Islam & Meng, 2022). But up until now, there have been few arctic wetlands-focused studies with GEE.

By leveraging GEE, Mahdianpari et al. (2021) has made remarkable progress in mapping wetlands across Canada's 9.9 million km<sup>2</sup> landscape. These scientific endeavors have cumulated into the Canadian Wetland Inventory Map (CWIM) initiative, which has been updated over three versions. Each CWIM version represents prominent, progressive improvements in terms of methodological complexity and accuracy. The first CWIM (CWIM1; Mahdianpari et al., 2020a, 2020b) used multi-year, multi-source EO imagery and object-oriented analysis, with provincial and territorial boundaries as processing units. The second CWIM (CWIM2; Mahdianpari et al., 2020a, 2020b) incorporated a much larger reference dataset and was completed using ecozones as processing units. The third and most recent CWIM (CWIM3; Mahdianpari et al., 2021) produced improved wetland class discrimination by incorporating L-band SAR and environmental data. Despite this immense progress, the authors concluded that the thematic detail and reliability of CWIM3 in Canada's northern ecozones is considerably lower than in other regions. Confounding this concern in map confidence is the disproportionate availability of reference samples, which is a major determinant of supervised ML performance and approximation. This is because a model's ability to effectively generalize and predict on unseen data requires sufficient and well-labelled training data, a broad finding that has been repeatedly demonstrated (Maxwell et al., 2018). Hence, a refined CWIM classification algorithm for Canada's Arctic region is essential, one in which training data, input variables, and computational efficiency is precisely studied.

The primary aim of this study was to establish a ML, cloud-based

framework for large-scale wetland classification over Canada's Arctic biome. In doing so we, we first created a comprehensive repository of model reference samples using low-cost and efficient techniques, and then tested several methodological approaches within GEE while using the previous findings by Mahdianpari et al. (2021) as a foundation for our data pipeline. Moreover, a thorough evaluation of both pixel and object-based methods was conducted, since other studies have concluded that object-based classification on GEE can be limited, computationally expensive, or susceptible to scaling risks over large areas (Del Valle & Jiang, 2022; Tamiminia et al., 2020). Other specific objectives included the following:

- 1) Enhance the GEE CWIM methodology by calibrating a new ML algorithm with new and innovative EO datasets, such as hydrography layers that capture wetland hydrology, and additional feature extractions such as mathematical indices.
- 2) Establish the accuracy of both a conventional pixel-based approach, and a more modern object-based approach for Arctic wetland classification.
- 3) Evaluate the predictive power of all multi-source EO features used in model development.
- 4) Provide recommendations on scaling the developed methodology across broader regions of the circumpolar region.

Our expectation is this study facilitates future research supporting regular CWIM updates and wetland monitoring at high latitudes, particularly by capitalizing on the extensive coverage and acquisitions of satellites with open data policies. These maps would support the responsible conservation and management of Arctic wetlands.

## 2. Materials and methods

### 2.1. Study area

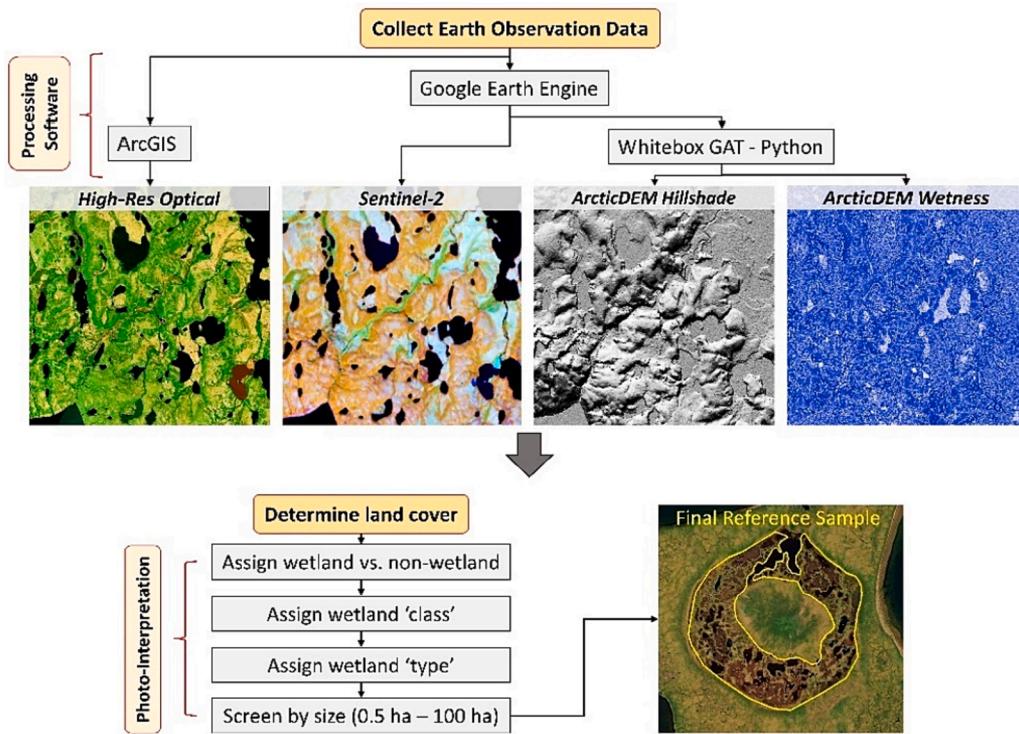
The study area was Canada's Southern Arctic ecozone (~830,000 km<sup>2</sup>), a unique and sparsely inhabited terrestrial ecosystem. Here, permafrost is regularly found, especially above the tree line (~66.5°) which represents the latitudinal limit of trees and the broad southern limit of the ecozone. Sitting on top of this permafrost is the active layer (AL), the surface layer that thaws and refreezes each year. AL thickness is predominately influenced by air temperature and latitude, and thus there is a southward increase in AL thickness across the ecozone. The shallow AL of this ecozone inhibits the establishment of deep-rooted vegetation, resulting in low tree cover and an abundance of low tundra plants; below the tree line exists dense evergreen taiga forests. Common floras include heath-lichen tundra on uplands, shrubby vegetation on slopes and ridges, and sedge, cotton grass, and willows in wetlands.

### 2.2. Wetland classes

This study focused on mapping the wetland class definitions of the Canadian Wetland Classification System (National Wetlands Working Group, 1997): open water, marsh, fen, bog, and swamp. These classes are distinguished by their developmental characteristics, including hydrology, morphology, hydrochemistry, soils, and vegetation, and the environment where they exist. In Canada's Arctic, these wetlands are predominantly open or shrubby systems (i.e., low biomass), since they are beyond the tree line. Other northern countries categorize wetlands similarly (Gunnarsson & Löfroth, 2014), making our methods repeatable over other Arctic regions.

### 2.3. Reference data repository

Considerable effort was committed to creating a large repository of spatially explicit training and testing samples, which is necessary for ML



**Fig. 1.** Reference data collection methodology.

as larger training sets typically generate improved performance (Maxwell et al., 2018). We implemented a low-cost and desktop-based interpretation methodology, guided by expert knowledge, for the collection of well-distributed reference samples. This approach was implemented because conventional field data collection is challenging and costly in Arctic environments. Collection over large areas further magnifies this challenge. This methodology, based on Merchant et al., (2022; Fig. 1), relies on high-resolution (sub-meter) optical, topographic, and multispectral imagery, representing the time period of ~2019–2021, to confidently determine wetland samples. For additional details on each EO source and their processing, see Merchant et al. (2022).

Reference data collection was completed within 30-by-30-kilometer sampling plots. Individual plots were chosen since the procedure required large volumes of EO data, making it impractical to carry out across the entire ecozone. 10 plots were strategically selected (Fig. 2), with each needing to meet a specific suite of criteria. First, plots needed to be located where no prior reference polygons existed from previous CWIM activities, thus filling spatial gaps lacking substantial reference data and where CWIM3 confidence was low. Second, appropriately timed, high-quality EO datasets needed to be available across the entire plot. We also spatially distributed plots across the ecozones entire latitudinal and longitudinal extent as best as possible. This was to capture the full range of environmental conditions and help avoid some spatial autocorrelation, as we recognize there are biases with directed sample collection methods.

About 1000 high confidence sites were digitized within each plot, ranging from 0.5 ha to 100 ha. Very large polygons were excluded because of spectral heterogeneity, while 0.5 ha was chosen to ensure narrow and/or small wetlands, which can be disproportionately abundant in the Arctic, were adequately represented without compromising accuracy. The number (9349) and distribution of sites reflects their presence across the ecozone (Table 1).

#### 2.4. EO data used for classification

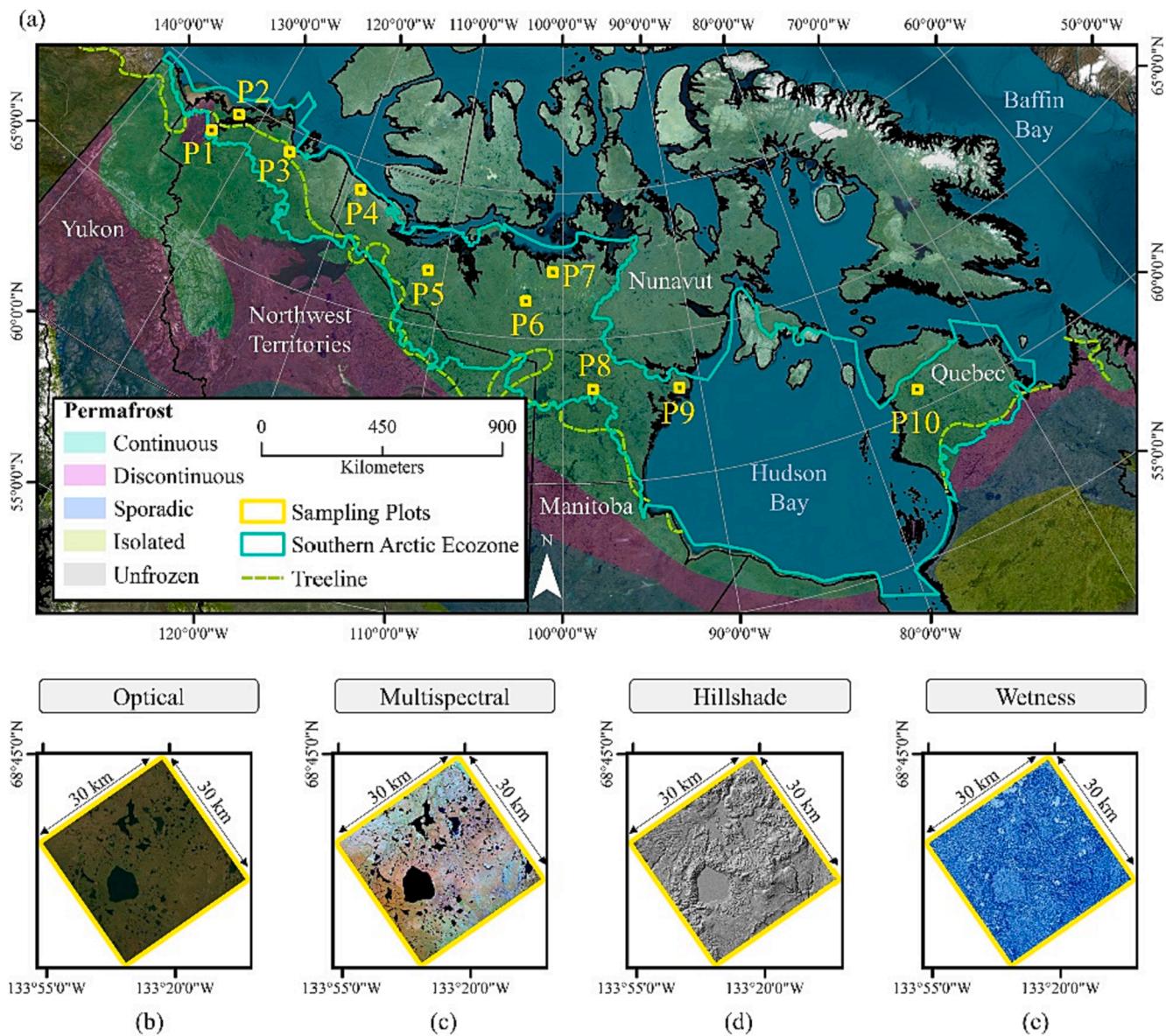
Multi-source EO data was used for classification, thus fusing the strengths of each source, and enhancing wetland discrimination abilities. All EO data was processed within GEE using JavaScript coding (Table 2). The following sections detail each source.

##### 2.4.1. Sentinel-2

Sentinel-2 is a multi-spectral imaging mission, collecting bands at 10 to 60 m with 5-day repeat imaging. Spectral zones include the visible to the shortwave infrared (SWIR) portions of the electromagnetic spectrum (EMS). Both Sentinel-2A/2B Level-2A surface reflectance data, pre-processed with Sen2Cor, was used. Due to the cloudy and/or low light conditions of Canada's Arctic, in addition to the short ice-off season, we queried images from July to August, and for the years 2019–2021. Images were selected with a cloudy pixel score of < 20 % using the QA60 band, and then applied cloud masking. Reflectance composites were created by calculating the index median from the time-series stack, although only for 10 or 20 m resolution bands. Temporal aggregation by median calculation is an advantageous method for gap filling cloud-masked areas (Carrasco et al., 2019).

##### 2.4.2. Sentinel-1

Sentinel-1 collects C-band (5.6 cm wavelength) SAR data from the microwave portion of the EMS. Level-1 Interferometric Wide Swath (IW) Ground Range Detected (GRD) mode was selected, leveraging a dual-polarization sensor with 12-day repeat imaging. Polarization channels included: (1) vertically transmitted and vertically received (VV), and (2) vertically transmitted and horizontally received (VH). Sentinel-1 images made available on GEE have been preprocessed using the Sentinel-1 Toolbox, which included orbital information updating, border noise removal, thermal noise removal, radiometric calibration to backscatter intensity (dB), and terrain correction with ASTER DEM. Collected Sentinel-1 images matched the same time period as the Sentinel-2 data, and were applied a median pixel function.



**Fig. 2.** (a) Southern Arctic ecozone in relation to the 10, 30-by-30-km sampling plots (P1, P2... P10) used for reference data collection. (b-c) EO examples for P1.

**Table 1**  
Reference sample repository.

Class	Samples
Open water	1342
Marsh	206
Swamp	162
Fen	1251
Bog	263
Upland	6125

**Table 2**  
Satellite images processed in GEE.

Source	Number of Processed Images
Sentinel-2	14,379
Sentinel-1	1970
ALOS PALSAR	3

#### 2.4.3. ALOS PALSAR

The global ALOS PALSAR mosaic in GEE was leveraged for its L-band (23.62 cm wavelength) microwave data, which permits improved canopy penetration capabilities over C-band Sentinel-1 (Chapman et al., 2015). JAXA generated this 25 m product by mosaicking strips of PALSAR/PALSAR-2 imagery (Shimada et al., 2009). Processing steps included ortho-rectification, slope correction, speckle filtering, radiometric calibration (dB), and destriping to equalize intensity differences between adjacent strips. Polarization channels included: (1) horizontally transmitted and horizontally received (HH), and (2) horizontally transmitted and vertically received (HV). Yearly mosaics for 2019, 2020, and 2021 were used.

#### 2.4.4. ArcticDEM

Topographic data was acquired from the ArcticDEM, a pan-Arctic digital surface model (DSM) spanning land north of 60°N (Shean et al., 2016). ArcticDEM was produced from high-resolution optical satellite imagery using stereo autocorrelation techniques and high-performance computing, and has an internal accuracy of 0.2 m, with some minor systematic vertical and horizontal offsets (Candela et al.,

**Table 3**  
Predictor variables used for classification.

Variable	Equation	Description
Blue, Green, Red, VRE1, VRE2, VRE3, NIR, NNIR, SWIR1, SWIR2	-	Sentinel-2 bands
NDVI	$\frac{(NIR - Red)}{(NIR + Red)}$	Normalized Difference Vegetation Index
RENDVI	$\frac{(NNIR - VRE2)}{(NNIR + VRE2)}$	Red Edge Normalized Difference Vegetation Index
EVI	$2 \left( \frac{(NIR - Red)}{(NIR + 6^*Red - 7.5^*Blue + 1)} \right)$	Enhanced Vegetation Index
SAVI	$1.5^* \frac{(NIR - Red)}{(NIR + Red + 0.5)}$	Soil Adjusted Vegetation Index
NBR	$\frac{(NIR - SWIR2)}{(NIR + SWIR2)}$	Normalized Burn Ratio
NDWI	$\frac{(Green - NIR)}{(Green + NIR)}$	Normalized Difference Water Index
NDSI	$\frac{(Green - SWIR1)}{(Green + SWIR1)}$	Normalized Difference Snow Index
VV, VH	-	Sentinel-1 linear polarizations
VV/VH	$\frac{VV}{VH}$	Sentinel-1 cross-polarization ratio
Sentinel Span	$ VV ^2 +  VH ^2$	Sentinel-1 total scattered power
Sentinel RVI	$\frac{(VH^*4)}{(VH + VV)}$	Sentinel-1 Radar Vegetation Index
HH, HV	-	ALOS PALSAR linear polarizations
HH/HV	$\frac{HH}{HV}$	ALOS PALSAR cross-polarization ratio
ALOS Span	$ HH ^2 +  HV ^2$	ALOS PALSAR total scattered power
ALOS RVI	$\frac{(HV^*4)}{(HV + HH)}$	ALOS Radar Vegetation Index
Elevation Slope	$\sqrt{D_x^2 + D_y^2}$	Height of elevation Rate of change of elevation, where $D_x$ is horizontal distance and $D_y$ is vertical distance
Aspect	$\arctan \left[ \frac{D_y}{D_x} \right]$	Direction of slope
Flow	$F_{out} = F_{in} + F_{local}$	Flow accumulation, where $F_{in}$ is flow received and $F_{local}$ is flow produced
HAND	$H_{point} - H_{drainage}$	Height Above Nearest Drainage, where $H_{point}$ is the DEM elevation and $H_{drainage}$ is the drainage elevation
PrecipMean	-	Average monthly precipitation
PrecipSD	-	Standard deviation monthly precipitation
TempMean	-	Average monthly temperature
TempSD	-	Standard deviation monthly temperature

2017). We processed three ArcticDEM terrain metrics in GEE: elevation, slope, and aspect.

#### 2.4.5. MERIT Hydro

Multi-Error-Removed-Improved-Terrain (MERIT) Hydro is a global hydrography dataset produced from the 3-arc second MERIT DEM and multiple inland water maps (Yamazaki et al., 2017). We selected flow accumulation (Peucker & Douglas, 1975) and height above nearest drainage (HAND; Nobre et al., 2011) from MERIT. HAND represents the elevation difference between a pixel and the nearest drainage network, whereas flow accumulation is the accumulative weight of all pixels

flowing downslope. Both parameters, which are novel to the CWIM methodology, were included because wetlands are strongly associated with hydrographic data.

#### 2.4.6. ERA5

Auxiliary climate data was acquired from ERA5, the fifth generation European Centre for Medium-range Weather Forecasts (ECMWF) atmospheric reanalysis of the global climate produced by the Copernicus Climate Change Service (Muñoz-Sabater et al., 2021). ERA5 provides global hourly measurements of atmospheric, land, and oceanic climate variables at 30 km resolution, produced by combining model data with global observations. We generated 10 year mean and standard deviation datasets of temperature and precipitation from ERA5, as long-term data have been shown to capture wetland spectral differences (Mahdianpari et al., 2021).

#### 2.5. Feature extraction

The number of features able to be collected by multi-source EO is ever increasing. This has major implications for classification performance, and for improving the translation of EO data into products such as landcover maps. For example, optical or SAR spectral indices combine two or more bands to effectively emphasizing specific targets (Dai et al., 2020). As such, we extracted numerous EO features to support our modelling (Table 3).

#### 2.6. Machine learning classifier

GEE contains several ML algorithms. Among these, RF has been applied extensively for landcover classification due to its flexibility, robustness, and because it frequently outperforms other classifiers (Zhou et al., 2021). For these reasons, we chose the tree-based RF ensemble algorithm.

RF employs bootstrap sampling to build decision trees on subsamples (Breiman, 2001). Samples not used in tree building are then assigned as out-of-bag (OOB). A RF can be expressed as:

$$\{h(X, L_i), i = 1, 2, \dots, M\} \quad (1)$$

where  $X$  are the independent variables,  $\{L_i\}$  are random vectors controlling individual decision tree growth, and  $M$  denotes the number of decision trees. Each individual tree makes a prediction based on  $X$ , with the final prediction the mode of all predictions. This leads to improved accuracies, a high tolerance for outliers, and controls over-fitting (Belgiu & Dragut, 2016).

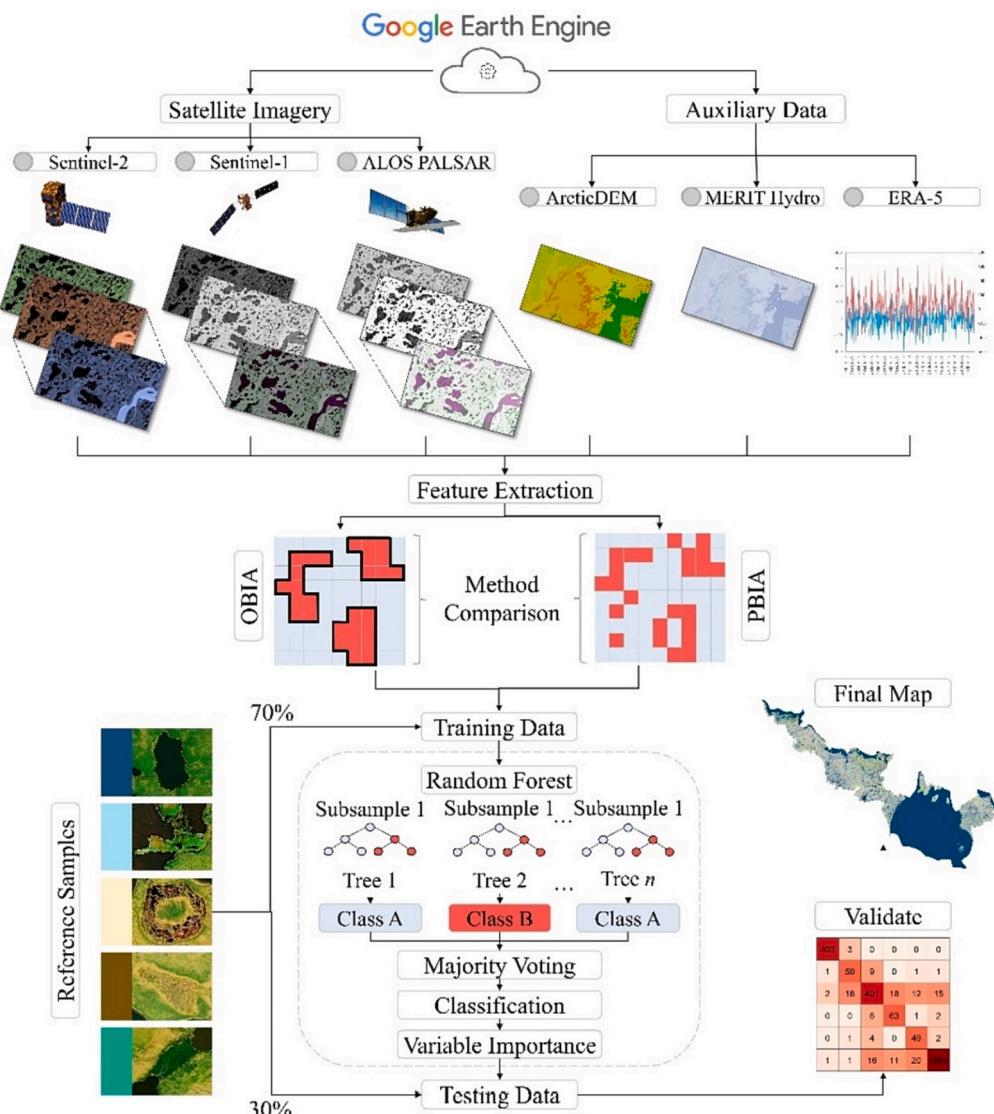
Running RF in GEE requires setting two key hyperparameters: the number of decision trees (nTree) and the minimum number of variables selected at each node (mTry). Based on recommendations within literature (Tamiminia et al., 2020), we assigned nTree to 100, and mTry was the square root of the total number of features.

##### 2.6.1. Variable importance and reduction

The inclusion of numerous high-dimensional datasets in ML can cause processing inefficiencies and information redundancy. To account for this, we quantified feature importance using internal cross-validation from OOB samples. This was later followed by a feature ranking and selection process. The intent here was to potentially improve scalability by removing groups of unimportant variables and reducing processing costs. This may benefit future Pan-Arctic mapping, by ensuring the processing pipeline could remain within GEE, by avoiding memory-intensive issues (Tamiminia et al., 2020).

To quantify feature importance, OOB prediction accuracy,  $A_{OOB}$ , is first calculated by dividing the total number of OOB samples,  $N_{OOB}$ , by the number of correctly classified samples,  $C_{OOB}$ :

$$A_{OOB} = \frac{C_{OOB}}{N_{OOB}} \quad (2)$$



**Fig. 3.** The implemented GEE workflow.

From  $A_{OOB}$ , the Mean Decrease in Accuracy (MDA) can be measured for bootstrap samples  $B_1, B_2, \dots, B_i, \dots, B_n$  with variables  $X_1, X_2, \dots, X_i, \dots, X_n$ . MDA calculates the change in prediction accuracy after features are randomly permuted. This is done by: 1) setting  $i$  to 1, creating a decision tree  $DT_i$  using training samples, and designating OOB samples as  $OOB_i$ , 2) calculating  $A_{OOB_i}$  (i.e., the number of correct predictions) using  $DT_i$  and  $OOB_i$ , 3) adding noise to each variable  $X_i$  to create  $OOB'_i$ , and calculating  $A_{OOB'_i}$  using  $DT_i$ , 4) repeating the previous steps for  $i = 2, 3, \dots, n$ , and lastly, 5) calculating importance for feature  $X_j$  using the following:

$$\overline{MDA}_j = \frac{1}{n} \sum_{i=1}^n A_{OOB_i} - A_{OOB'_i} \quad (3)$$

MDA is considered a reliable measure of importance compared to Gini importance which suffers from a splitting bias (Loecker, 2022).

## 2.7. Classification techniques

RS classification can be allocated into two general techniques: object-based image analysis (OBIA) and pixel-based image analysis (PBIA). OBIA's popularity has rapidly grown, offering an effective alternative to 'traditional' PBIA. With OBIA, pixels are grouped into

objects based on a set of criteria, such as spectral or spatial attributes, and then classified (Dronova, 2015). While studies have demonstrated OBIA advantages over PBIA for wetland classification (Mahdianpari et al., 2019), few have done so over Arctic environments. Thus, we compared both approaches (Fig. 3). The Simple Non-Iterative Clustering (SNIC) algorithm was implemented for OBIA. SNIC uses a regular grid of seeds as input which influences cluster size, and requires a superpixel seed location. This was set to 10 following earlier GEE research (Tassi & Vizzari, 2020). Other SNIC parameters included: 1) compactness, which influences shape, 2) connectivity, which influences adjacent object contiguity, and 3) neighborhood size, which avoids boundary artifacts. Following extensive visual trial and error, these were set to: compactness = 0.1, connectivity = 8, and neighborhood size = 256.

## 2.8. Accuracy assessment and statistical comparison

Models were validated using 2804 independent testing polygons, which were randomly selected from the 10 plots (30 % of samples, and stratified by class, following other wetland studies (Mahdianpari et al., 2019)). Several statistics measured from a confusion matrix, including overall accuracy, precision, recall, F1-scores, and Kappa coefficient, were calculated:

**Table 4**  
McNemar tests contingency matrix.

		Model 1 (OBIA)	Model 2 (PBIA)
		Correct	Incorrect
Model 1 (OBIA)	Correct	$f_{11}$	$f_{12}$
Model 2 (PBIA)	Incorrect	$f_{21}$	$f_{22}$

$$\text{OverallAccuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{Kappa} = \frac{\text{OverallAccuracy} - RA}{1 - RA} \quad (8)$$

where  $TP$  represents true positives,  $TN$  represents true negatives,  $FP$  represents false positives,  $FN$  represents false negatives, and  $RA$  represents random accuracy.

Non-parametric McNemar tests were also used to assess statistical difference between pair-wise classifications (McNemar, 1947). The test is a Chi-square ( $\chi^2$ ) test for goodness of fit whereby  $\chi^2$  follows a chi-squared distribution with one degree of freedom, and compares the

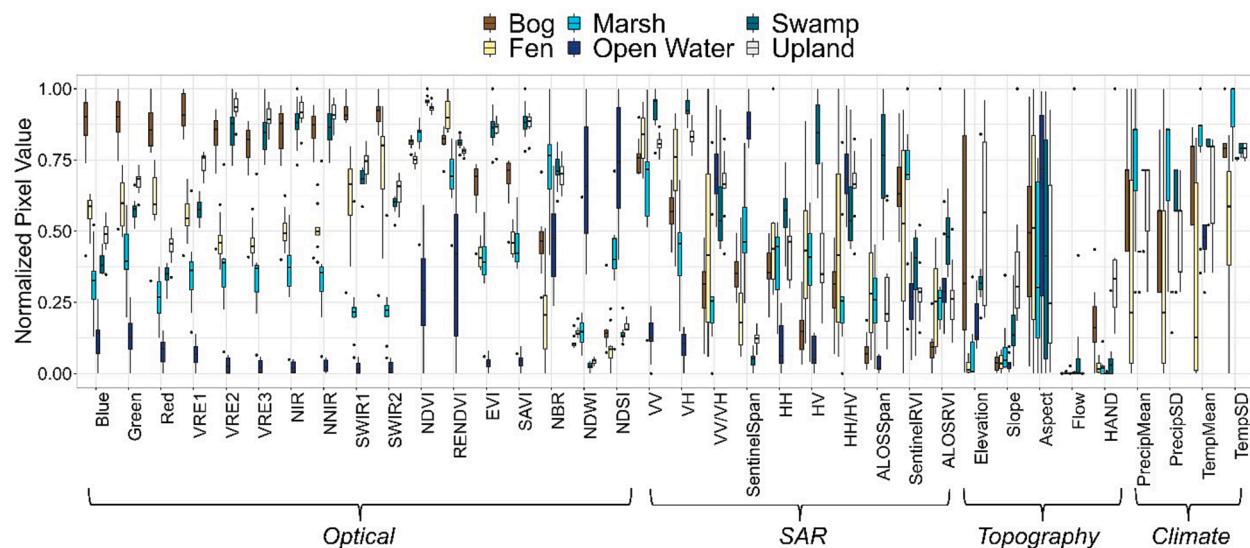


Fig. 4. Spectral response profile for all EO features.

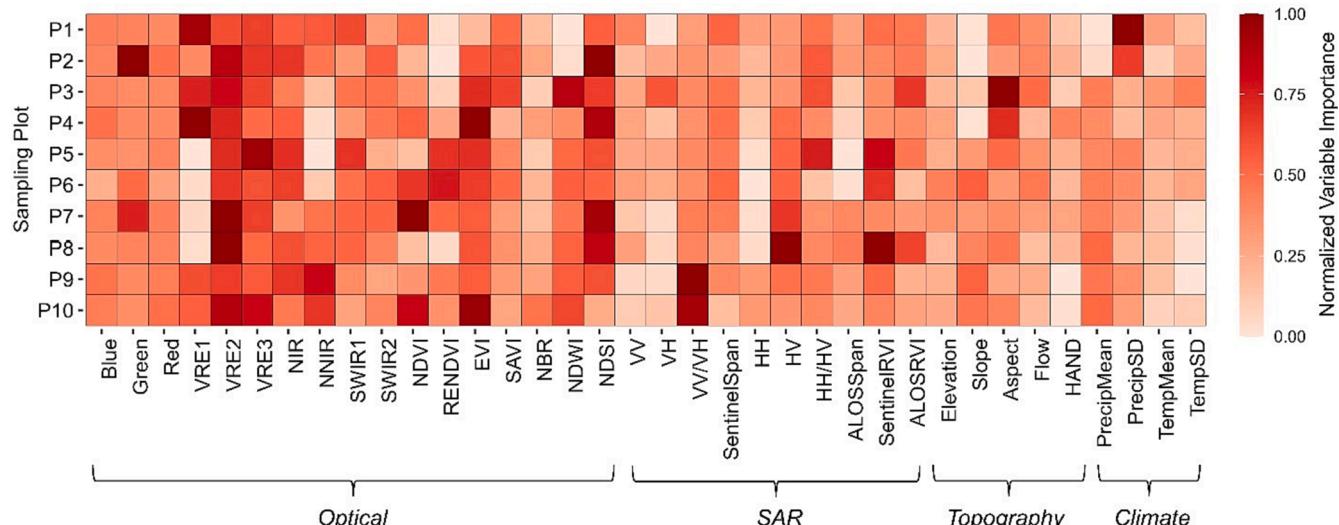
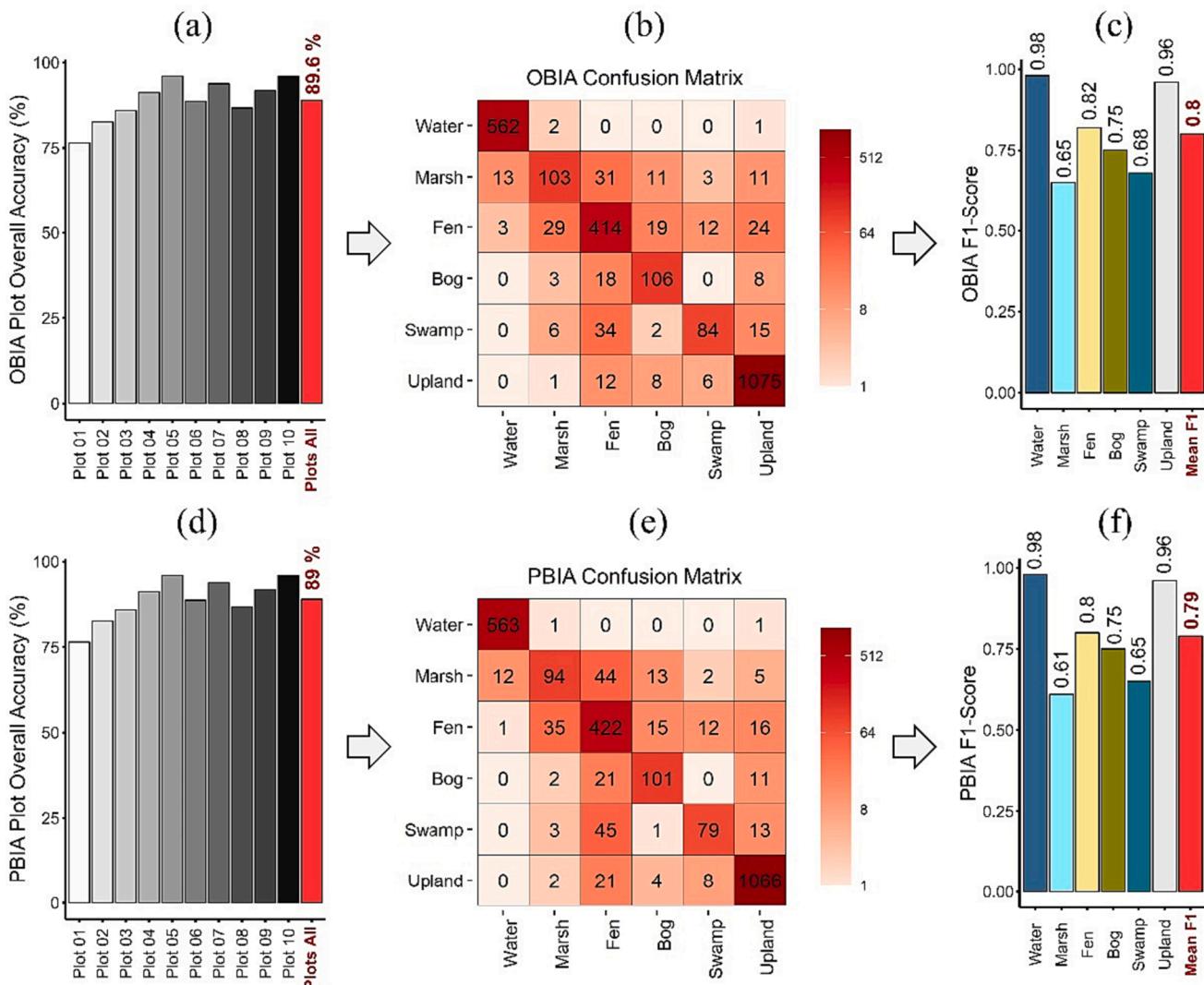


Fig. 5. EO feature variable importance, reported for each plot (P1, P2... P10).



**Fig. 6.** Overall accuracy for individual plots using (a) OBIA and (d) PBIA. Summarized confusion matrix for (b) OBIA and (e) PBIA. F1-scores using (c) OBIA and (f) PBIA.

counts expected under the null hypothesis ( $H_0$ ; that the two classifications perform similarly) against the observed counts. This is calculated using a collapsed  $2 \times 2$ -dimensional error matrix comprised of independent correct and incorrect samples (Table 4). The McNemar test equation is as follows:

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \quad (10)$$

where  $f_{12}$  represents incorrect samples with classification one while classification two is uncorrected, and  $f_{21}$  represents correct samples in classification two while classification one is uncorrected. If the calculated value of  $\chi^2$  is greater than the value of  $\chi^2$ ,  $H_0$  is rejected; if it is lower, than  $H_0$  is accepted.

### 3. Results and discussions

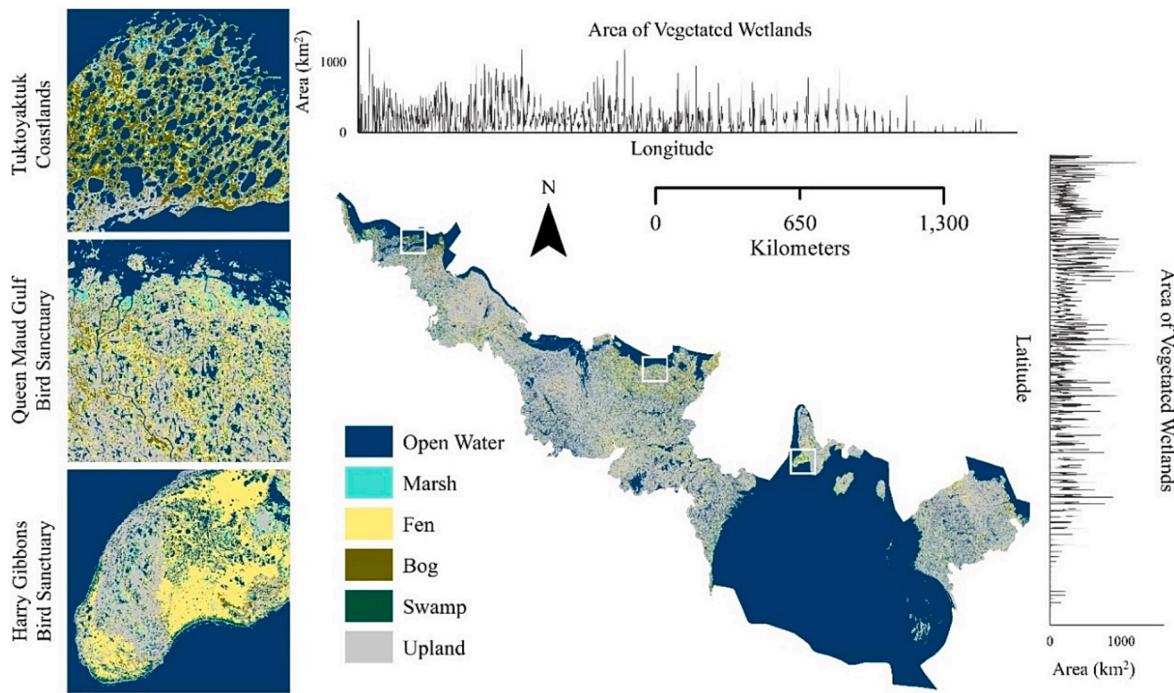
#### 3.1. Signature observations

Different wetlands elicit unique responses across the EMS, resulting from heterogeneous soil, water, and vegetation characteristics (Amani et al., 2018). Thus, all EO features were used to create a comprehensive signature profile of wetlands prior to classification; values were normalized from 0-1 (Fig. 4; note that for aspect, the cosine of degree

values was used). Optical features best discriminated most classes. In particular, several infrared bands displayed noticeable spectral class partitioning, in part because of their sensitivity to vegetation biomass and moisture conditions (Mahdavi et al., 2018). Swamps, however, were more challenging to separate with optical features, due to high canopy closures. In contrast, PALSAR HV and Span features visually exhibited strong separation of swamps due to the L-band wavelength which can penetrate dense biomass and wet soils (Chapman et al., 2015). Several hydrological and topographic variables discriminated wetlands from uplands well, including HAND and Slope. This is because Arctic wetlands develop within well-defined topographical niches, such as in low-lying depressions, riparian areas, or drainage networks (Woo & Young, 2005).

#### 3.2. Variable importance

Fig. 5 presents MDA-derived variable importance. Values were normalized from 0-1. This heatmap was created by quantifying the importance of variables within each sampling plot (i.e., P1 to P10), thus providing an understanding of predictor importance along the ecozone's east-to-west longitudinal extent. Variables discriminating wetlands well in Fig. 4 also contributed strongly to the ML modelling, including several optical features (e.g., EVI). Computing indices is an effective technique,



**Fig. 7.** Final OBIA-based map. The distribution ( $\text{km}^2$ ) of vegetated wetlands (i.e., marsh, fen, bog, and swamp) is plotted against longitude and latitude. Inset maps correspond to white boxes.

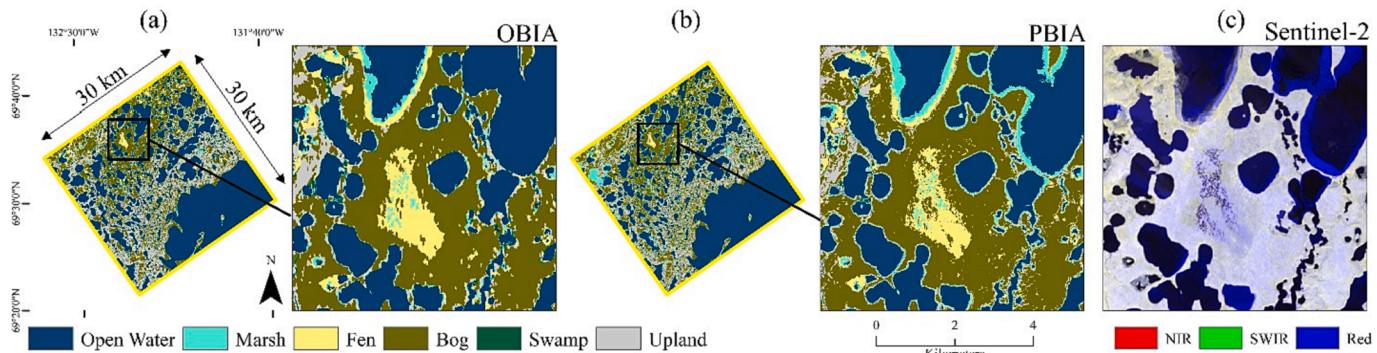
due to their unique ability to amplify spectral differences. For example, EVI has been used to characterize wetland conditions and is advantageous because of its sensitivity to canopy conditions (Dai et al., 2020). RVI indices, which are novel contributions to the CWIM methodology, also indicated strong classification influence. Literature on RVI's use in wetland classification is sparse, although its predictive power can be attributed to its sensitivity to terrestrial vegetation water content (Huang et al., 2016).

Optical and SAR variables had the greatest contributions to the ML modelling, followed closely by ArcticDEM and MERIT Hydro variables. For example, HAND can detect inundated areas because it represents regional hydrological characteristics, such as drainage features, flow paths, and low-lying depressions (Aristizabal et al., 2020). This hydrological information, coupled with ArcticDEM's high-resolution topographic data, evidently helps resolve some of the challenges with Arctic wetland classification, such as variable, within-class spectral signatures (Bartsch et al., 2016). ERA5 climate variables, especially temperature, were less influential, likely because climatic conditions are similar across the ecozone, making it difficult for RF to capitalize on the data at this scale.

### 3.3. OBIA and PBIA comparison

**Fig. 6** displays model accuracy assessments which were first quantified within each plot, and then summarized, for both OBIA and PBIA. PBIA produced an overall accuracy of 89.0 %, while OBIA was 89.6 %, representing an increase of (+) 0.6 % (**Fig. 6**). A McNemar test indicated this was not statistically significant ( $p$ -Value = 0.975) at the 95 % confidence level. Other statistical measures included a kappa of 0.856, mean F1-score of 0.809, and mean F1-score for wetland classes of 0.778 for OBIA, and a kappa of 0.848, mean F1-score of 0.795, and mean F1-score for wetland classes of 0.761 for PBIA. This demonstrates that OBIA is marginally better for this classification problem, but that both methods can produce accurate and statistically similar results. Other studies have reported similar findings in Arctic environments, although at more local scales (Demers et al., 2015).

On a per-class basis, both OBIA and PBIA classified open water accurately with F1-scores of  $\sim 0.98$ . Open water is easily detected due to high absorption of incident radiant flux or SAR specular reflection (Mahdavi et al., 2018). Peatlands, both fen and bog, were mapped more accurately with OBIA with F1-scores of 0.820 and 0.754. Fens were mostly confused with swamps, possibly where biomass was similar. In



**Fig. 8.** OBIA and PBIA comparison over P2, Tuktoyaktuk coastlands, NWT. (a) OBIA. (b) PBIA. (c) Sentinel-2.

**Table 5**

Model validations using various proportions (i.e., percentages) of top-ranked variables.

Model	Overall Accuracy (%)	Kappa	Mean F1-Score	Mean Wetland F1-Score
OBIA 100 % variables	89.6	0.856	0.809	0.778
OBIA 75 % variables	88.7	0.845	0.794	0.761
OBIA 50 % variables	87.8	0.832	0.780	0.746

**Table 6**

McNemar tests comparing different classifications.

Classification Comparisons	$\chi^2$	p-Value
OBIA 100 % variables vs. PBIA 100 % variables	0.001	0.975
OBIA 100 % variables vs. OBIA 75 % variables	2.943	0.086
OBIA 100 % variables vs. OBIA 50 % variables	2.980	0.084

contrast, bogs were mostly mistaken for fens, as both occur along hydrological and nutrient gradients that, at times, can make them difficult to distinguish (Olefeldt et al., 2021). Swamps and marshes were the most challenging to classify, although OBIA represented F1-Score increases of (+) 0.04 and (+) 0.03 over PBIA. We attributed this to two factors: vegetation composition similarities, which inherently lead to spectral confusion (Amani et al., 2018), such as between graminoid dominated marshes and fens, and second being training data distribution. Peatlands are extensive in the Arctic, which develop when organic deposits accumulate from cold temperatures, waterlogged conditions, and slow decomposition (Hugelius et al., 2020). Our reference data repository reflects this distribution, which was done to avoid an imbalance between training data and actual land cover proportions (Millard & Richardson, 2015). Nevertheless, this may have influenced classification performance, particularly for more rare, mineral soil wetlands. The final wetland map, produced using OBIA, is presented in Fig. 7.

Beyond statistical performance, there are several qualitative advantages OBIA holds over PBIA. Most notably is the smoothing of local variation, which reduces “salt-and-pepper” effects, resulting in a cleaner appearing classification (Dronova, 2015). An example the noise suppressing capabilities by OBIA is in Fig. 8.

### 3.4. Variable reduction assessment

A frequent objective in ML is to decrease predictor variables to reduce the inconvenience of data collection and improve computational efficiencies (Maxwell et al., 2018). To achieve stable variable importance values (Behnamian et al., 2017), we averaged the values in Fig. 5, ranked each variable, then processed two additional OBIA classifications representing more parsimonious models. We advanced OBIA for this reduction assessment, rather than PBIA, because it represented the more accurate technique. These models used the top 75 %, and then 50 %, most important variables. Results demonstrated progressively decreasing accuracies compared to using all (i.e., 100 %) variables, with overall accuracy differences of (–) 0.9 % and (–) 1.8 % (Table 5). McNemar tests indicated that neither reduced model produced a significantly different result ( $p$ -Values  $> 0.05$ ) than when using all variables, despite the observed loss in accuracy (Table 6). This finding has significant tradeoff implications, whereby minor losses in accuracy allow for improved computing time, storage, and scalability within GEE, which are common limitations reported by previous studies (Tamiminia et al., 2020). Being able to implement an optimal, low dimensional feature set using feature extraction and/or selection methods provides increased opportunity for repeatable mapping at broader scales.

### 3.5. Comparison to CWIM

Our classification revealed 36.2 % of vegetated wetlands (i.e., marsh, fen, bog, and swamp) and 34.3 % of open water cover the study area. From this, 28.7 % are organic soil wetlands and 7.5 % are mineral soil wetlands. This represents a substantial difference in wetland coverage compared to CWIM3 (Fig. 9). However, this was anticipated as the CWIM3 authors stated that the maps confidence was low across northern Canada, resulting from low reference data abundance and historical disturbance. CWIM3 also did not identify bogs in this ecozone, which are extensive in certain regions of the Arctic. For example, extensive bog coverage was predicted in the northwestern region of the ecozone, particularly around the Tuktoyaktuk coastlands where their occurrence is associated with polygonal ice-wedges (Liljedahl et al., 2016).

Visual assessment showed that wetland spatial patterns were preserved and realistic with our classification, with minimized omission and commission errors. This assessment, supported by high-resolution DEM and satellite images (e.g., WorldView-1 and -2, see Merchant et al., 2022)), was also used to compare against CWIM3. Fig. 10

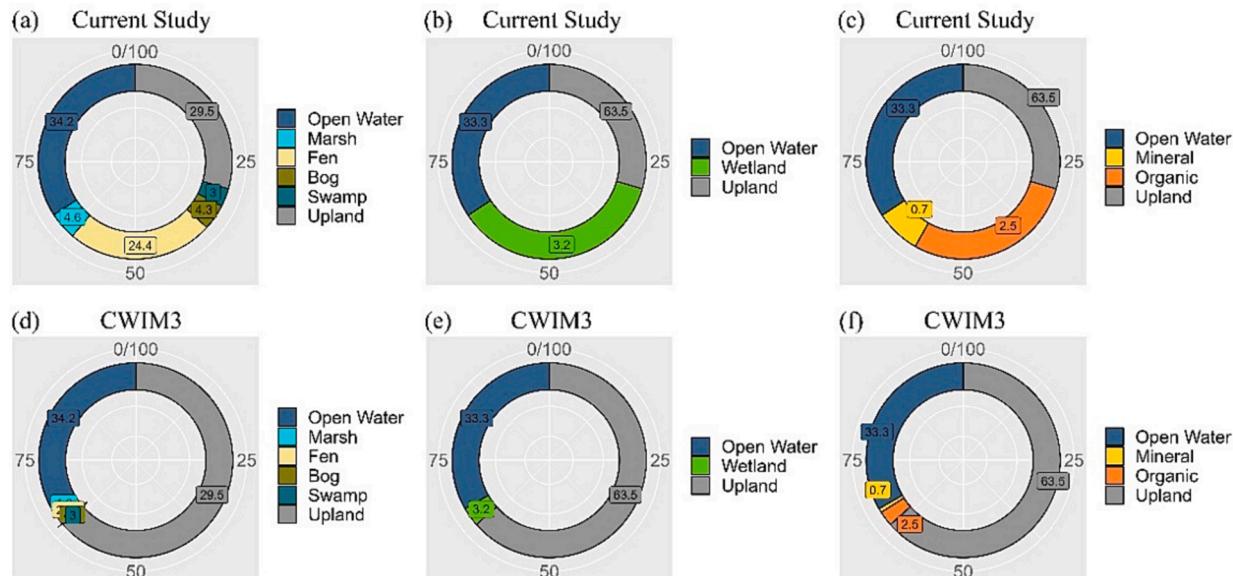
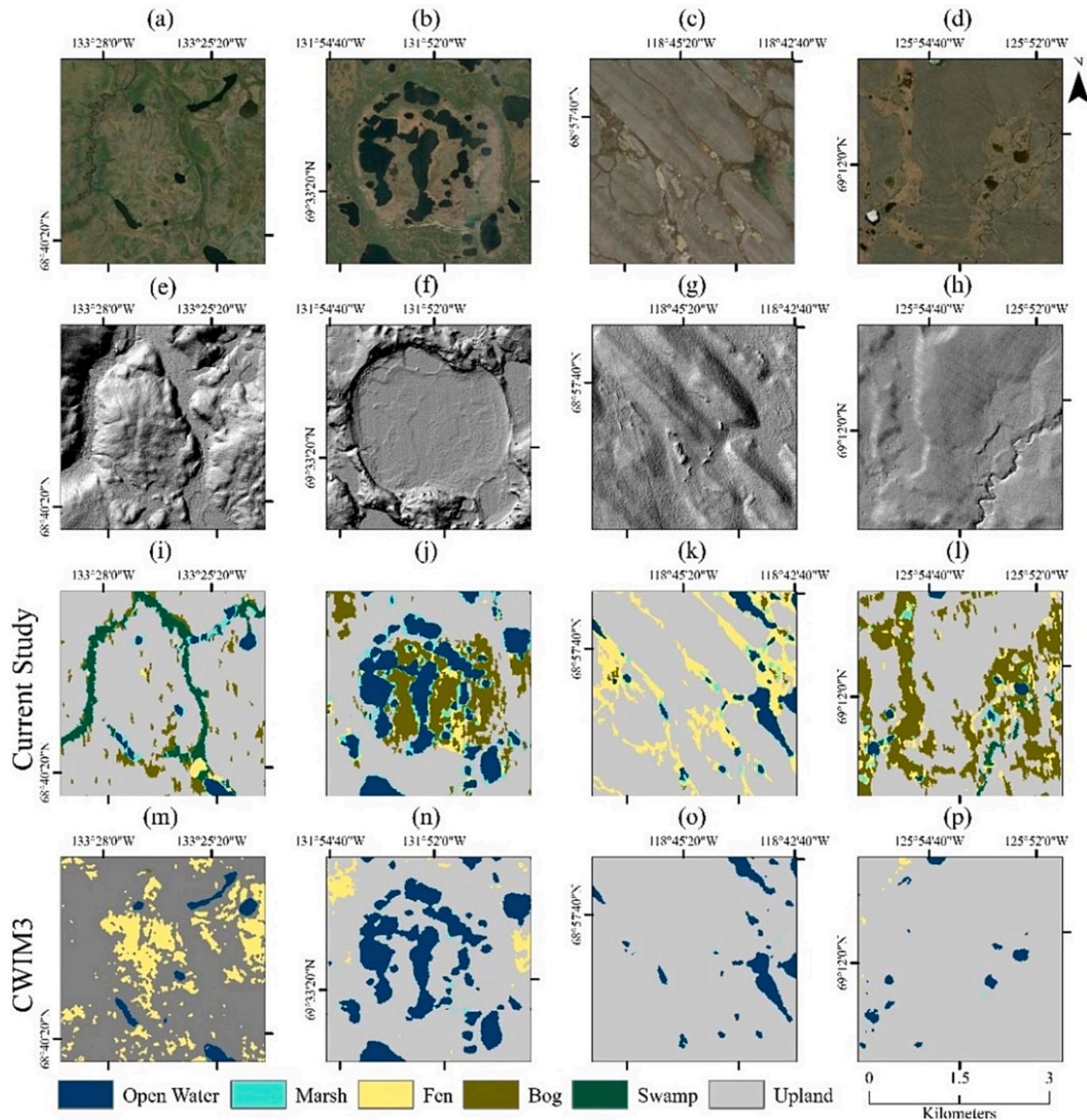


Fig. 9. Wetland distributions from (a-c) current study and (d-f) CWIM3.



**Fig. 10.** (a-d) Optical and (e-h) hillshade imagery for reference. (i-l) Current study classification. (m-p) CWIM3 classification.

demonstrates how our model predicted peatlands in topographical lows, marshes transitioning from water features, and swamps within deeper channels. Hence, a close representation of real land surface formations. Collectively, these observations suggest that CWIM3 underestimates wetlands in Canada's Arctic. To further confirm this, a quantitative accuracy assessment was performed on CWIM3 using our new reference samples. Results indicated an overall accuracy of 79.8 % and kappa of 0.537. Moreover, F1-scores for vegetated wetlands ranged from 0.156 to 0.303, suggesting CWIM3 had difficulty mapping these classes.

### 3.6. Pan-Arctic wetlands mapping

Due to our promising results, we suggest testing this methodology over broader regions of the Arctic as there is a need for improved knowledge of wetland coverages and types at a Pan-Arctic scale, especially for land surface and climate modelling. [Widhalm et al. \(2015\)](#), [Hugelius et al. \(2020\)](#), and [Olefeldt et al. \(2021\)](#) have temporarily helped fill this need, however each source has limitations. For example, [Widhalm et al. \(2015\)](#) used a SAR-only approach and consequently only classified wetland hydrology, whereas [Hugelius et al. \(2020\)](#) and [Olefeldt et al. \(2021\)](#) modelled wetlands at very coarse spatial resolutions. Thus, efforts around high-resolution Pan-Arctic wetland mapping are

still required and should focus on several key elements, including establishing a representative or potentially even unique classification system, optimizing a classification algorithm, and compilation and harmonization of reference data. With the latter, we demonstrated that our desktop-based approach is efficient for training data collection, especially compared to field-based methods. However, this approach requires expert knowledge of Arctic landforms, and the ability to interpret complex spectral signatures from multiple EO sources. Accordingly, the active collaboration between environmental organizations is highly encouraged for sharing of valuable in-situ reference data. Investments in such data could support a classification schema that better categorizes Arctic wetlands, potentially at greater detail such as by dominant vegetation communities (e.g., shrub, lichen, etc.). A limitation of this study is our reference data was not field collected, and our thematic labels omit critical structural information.

Another consideration for optimizing this methodology more broadly is whether to utilize OBIA or PBIA. Some studies have either avoided OBIA and/or reported memory-related issues within GEE (e.g., computation time-outs), limiting the execution of computationally demanding tasks like segmentation with high-resolution imagery over large areas (Yang et al., 2022). Other studies divided their region into multiple processing units to avoid memory limits (Chang et al., 2022). Our results showed minimal and non-statistically significant differences between OBIA and PBIA, which is comparable to some other studies (Duro et al., 2012). Similar results were found when input variables were reduced. Moreover, we observed vast differences in running time, as OBIA processing took 15 days (with several attempts) and PBIA 6 days. Therefore, future Pan-Arctic efforts could consider foregoing OBIA due to processing limitations or reducing large groups of unimportant features. Alternatively, some studies have exported imagery from GEE for further analysis (Jamali et al., 2023), allowing access to functionalities not within GEE (e.g., deep learning). However, this does not take full advantage of GEE's cloud infrastructure, thus limiting the potential for efficiency.

#### 4. Conclusions

Here, we presented one of the first studies with a goal of large-scale Arctic wetland classification using cloud-computing. We demonstrated the value of multi-source EO data, ML, and GEE for mapping these complex ecosystems at high latitudes. We also addressed several sub-objectives; results and significant findings are as follows:

- 1) An efficient yet innovative approach, leveraging several EO sources beyond just optical imagery, was used to create a large repository of wetland samples.
- 2) Using Mahdianpari et al. (2021) as a foundation, we developed an enhanced GEE workflow by incorporating novel EO features.
- 3) Results indicated OBIA to be more accurate than PBIA, although differences were minimal and not statistically significant.
- 4) Signature observations and variable importance measurements indicated optical features had the greatest ML predictive power.
- 5) Recommendations around Pan-Arctic mapping were suggested, including the need for harmonized reference data and further algorithm calibration with a focus on computational efficiencies.

We expect this study to influence future wetland mapping efforts across Canada's Arctic ecozones, and globally over the circumpolar region. Wetland inventories at these scales are necessary for natural resource management and policy development.

#### CRediT authorship contribution statement

**Michael Merchant:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Brian Brisco:** Conceptualization, Supervision, Funding acquisition. **Masoud Mahdianpari:**

Conceptualization, Methodology, Software, Supervision, Funding acquisition. **Laura Bourgeau-Chavez:** Writing – review & editing, Supervision. **Kevin Murnaghan:** Data curation, Writing – review & editing. **Ben DeVries:** Writing – review & editing, Supervision, Funding acquisition. **Aaron Berg:** Writing – review & editing, Supervision, Funding acquisition.

#### 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

GitHub link has been provided for the study code

#### Acknowledgements

We acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (grants RGPIN-2020-05743 and STPGP-521584) and the Global Water Futures through Northern Water Futures, funded by the Canada First Research Excellence Program. Michael Merchant conducted this research through a Research Affiliate Program with the Canada Center for Mapping and Earth Observation.

#### References

- Amani, M., Salehi, B., Mahdavi, S., Brisco, B., 2018. Spectral Analysis of Wetlands Using Multi-Source Optical Satellite Imagery. *ISPRS J. Photogramm. Remote Sens.* 144, 119–136.
- Aristizabal, F., Judge, J., Monsivais-Huertero, A., 2020. High-resolution inundation mapping for heterogeneous land covers with synthetic aperture radar and terrain data. *Rem. Sens.* 12 (900), 1–25.
- Bartsch, A., Höfler, A., Kroisleitner, C., Trofaier, A., 2016. Land Cover Mapping in Northern High Latitude Permafrost Regions with Satellite Data : Achievements and Remaining Challenges. *Rem. Sens.* 8 (979), 1–27.
- Behnamian, A., Millard, K., Banks, S.N., White, L., Richardson, M., Pasher, J., 2017. A Systematic Approach for Variable Selection with Random Forests: Achieving Stable Variable Importance Values. *IEEE Geosci. Remote Sens. Lett.* 14 (11), 1988–1992.
- Belgiu, M., Dragut, L., 2016. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Breiman, L., 2001. Random Forests. *Machine Learn.* 45, 5–32.
- Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mård, J., Mernild, S.H., Prowse, T., Semenova, O., Stuefer, S.L., Woo, M.K., 2016. Arctic terrestrial hydrology: A synthesis of processes, regional effects, and research challenges. *J. Geophys. Res. Biogeosci.* 121, 621–649.
- Candela, S., Howat, I., Noh, M., Porter, C., Morin, P., 2017. ArcticDEM Validation and Accuracy Assessment. *AGU Fall Meeting Abstracts C51A-C.*
- Carrasco, L., O'Neil, A.W., Daniel Morton, R., Rowland, C.S., 2019. Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sens.* 11 (288), 1–21.
- Chang, M., Li, P., Li, Z., Wang, H., 2022. Mapping Tidal Flats of the Bohai and Yellow Seas Using Time Series Sentinel-2 Images and Google Earth Engine. *Remote Sens.* 14 (1789), 1–20.
- Chapman, B., McDonald, K., Shimada, M., Rosenqvist, A., Schroeder, R., Hess, L., 2015. Mapping regional inundation with spaceborne L-Band SAR. *Remote Sens.* 7 (5), 5440–5470.
- Cohen, J., Screen, J.A., Furtado, J.C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., Jones, J., 2014. Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.* 7, 627–637.
- Dai, X., Yang, G., Liu, D., Wan, R., 2020. Vegetation carbon sequestration mapping in herbaceous wetlands by using a MODIS EVI time-series data set: A case in Poyang Lake Wetland, China. *Remote Sens.* 12 (3000), 1–19.
- Del Valle, T., Jiang, P., 2022. Comparison of common classification strategies for large-scale vegetation mapping over the Google Earth Engine platform. *Int. J. Appl. Earth Obs. Geoinf.* 115, 103092.
- Demers, A.M., Banks, S.N., Pasher, J., Duffe, J., Laforest, S., 2015. A comparative analysis of object-based and pixel-based classification of RADARSAT-2 C-band and optical satellite data for mapping shoreline types in the Canadian arctic. *Can. J. Remote Sens.* 41, 1–19.
- Dronova, I., 2015. Object-Based Image Analysis in Wetland Research: A Review. *Remote Sens.* 7, 6380–6413.
- Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification

- of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* 118, 259–272.
- Gunnarsson, U., & Löfroth, M. 2014. The Swedish Wetland Survey – Compiled Excerpts From The National Final Report.
- Huang, Y., Walker, J.P., Gao, Y., Wu, X., Monerris, A., 2016. Estimation of Vegetation Water Content From the Radar Vegetation Index at L-Band. *IEEE Trans. Geosci. Remote Sens.* 54 (2), 981–989.
- Hugelius, G., Loisel, J., Chadburn, S., Jackson, R.B., Jones, M., MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M.B., Treat, C., Turetsky, M., Voigt, C., Yu, Z., 2020. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proc. Natl. Acad. Sci. U.S.A.* 117 (34), 20438–20446.
- Islam, M., Meng, Q., 2022. An exploratory study of Sentinel-1 SAR for rapid urban flood mapping on Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* 113, 103002.
- Jamali, A., Roy, S., Ghamisi, P., 2023. WetMapFormer: A unified deep CNN and vision transformer for complex wetland mapping. *Int. J. Appl. Earth Obs. Geoinf.* 120, 103333.
- Kåresdotter, E., Destouni, G., Ghajarnia, N., Hugelius, G., Kalantari, Z., 2021. Mapping the Vulnerability of Arctic Wetlands to Global Warming. *Earth's Future* 9 (e2020EF001858), 1–14.
- Liljedahl, A.K., Boike, J., Daanen, R.P., Fedorov, A.N., Frost, G.V., Grosse, G., Hinman, L.D., Iijima, Y., Jorgenson, J.C., Matveyeva, N., Necsoiu, M., Raynolds, M.K., Romanovsky, V.E., Schulla, J., Tape, K.D., Walker, D.A., Wilson, C.J., Yabuki, H., Zona, D., 2016. Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology. *Nat. Geosci.* 9, 312–319.
- Liu, P., Di, L., Du, Q., Wang, L., 2018. Remote sensing big data: Theory, methods and applications. *Remote Sens.* 10 (711), 1–4.
- Loecker, M., 2022. Unbiased variable importance for random forests. *Commun. Stat. - Theory Methods.* 51 (5), 1413–1425.
- Long, X., Li, X., Lin, H., Zhang, M., 2021. Mapping the vegetation distribution and dynamics of a wetland using adaptive-stacking and Google Earth Engine based on multi-source remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102453.
- Mahdavi, S., Salehi, B., Granger, J., Amani, M., Brisco, B., Huang, W., 2018. Remote sensing for wetland classification: a comprehensive review. *GIScience remote sens.* 55 (5), 623–658.
- Mahdavi, S., Salehi, B., Amani, M., Granger, J., Brisco, B., Huang, W., 2019. A dynamic classification scheme for mapping spectrally similar classes: Application to wetland classification. *Int. J. Appl. Earth Obs. Geoinf.* 83, 101914.
- Mahdianpari, M., Salehi, B., Mohammadimansh, F., Homayouni, S., 2019. The First Wetland Inventory Map of Newfoundland Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform. *Remote Sens.* 11 (43), 1–27.
- Mahdianpari, M., Brisco, B., Granger, J.E., Mohammadimansh, F., Salehi, B., Banks, S., Homayouni, S., Bourgeau-Chavez, L., Weng, Q., 2020a. The Second Generation Canadian Wetland Inventory Map at 10 Meters Resolution Using Google Earth Engine. *Can. J. Remote Sens.* 46 (3), 360–375.
- Mahdianpari, M., Salehi, B., Mohammadimansh, F., Brisco, B., Homayouni, S., Gill, E., DeLancey, E., Bourgeau-Chavez, L., 2020b. Big Data for a Big Country: The First Generation of Canadian Wetland Inventory Map at a Spatial Resolution of 10-m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform. *Can. J. Remote Sens.* 46 (1), 15–33.
- Mahdianpari, M., Brisco, B., Granger, J., Mohammadimansh, F., Salehi, B., Homayouni, S., Bourgeau-Chavez, L., 2021. The third generation of pan-canadian wetland map at 10 m resolution using multisource earth observation data on cloud computing platform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 8789–8803.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* 39 (9), 2784–2817.
- McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12 (2), 153–157.
- Merchant, M., Adams, J., Berg, A., Baltzer, J., Quinton, W., Chasmer, L., 2017. Contributions of C-Band SAR Data and Polarimetric Decompositions to Subarctic Boreal Peatland Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10 (4), 1467–1482.
- Merchant, M., Brisco, B., Mahdianpari, M., Granger, J., Devries, B., Berg, A., 2022. A desktop-based methodology for collecting wetland reference data over inaccessible arctic landscapes. *IGARSS* 5893–5896.
- Millard, K., Richardson, M., 2015. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sens.* 7, 8489–8515.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data.* 13, 4349–4383.
- National Wetlands Working Group. 1997. The Canadian Wetland Classification System. In National Wetlands Working Group (Second Edi). University of Waterloo.
- Nobre, A.D., Cuartas, L.A., Hodnett, M., Rennó, C.D., Rodrigues, G., Silveira, A., Waterloo, M., Saleska, S., 2011. Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *J. Hydrol.* 404, 13–29.
- Olefeldt, D., Hovemyr, M., Kuhn, M.A., Bastviken, D., Bohn, T.J., Connolly, J., Crill, P., Euskirchen, E.S., Finkelstein, S.A., Genet, H., Grosse, G., Harris, L., Heffernan, L., Helbig, M., Hugelius, G., Hutchins, R., Juutinen, S., Lara, M., Malhotra, A., Watts, J., 2021. The Boreal – Arctic Wetland and Lake Dataset (BAWLD). *Earth Syst. Sci. Data.* 13, 5127–5149.
- Peucker, T.K., Douglas, D.H., 1975. Detection of Surface-Specific Points by Local Parallel Processing of Discrete Terrain Elevation Data. *Comput. Graph. Image Process.* 4, 375–387.
- Schuur, E., McGuire, A., Schadel, C., Grosse, G., Harden, J., Hayes, D., Hugelius, G., Koven, C., Puhry, P., Lawrence, D., Natali, S., Olefeldt, D., Romanovsky, V., Schaefer, K., Turetsky, M., Treat, C., Vonk, J., 2015. Climate change and the permafrost carbon feedback. *Nature* 520, 171–179.
- Seifollahi-Aghmuni, S., Kalantari, Z., Land, M., Destouni, G., 2019. Change drivers and impacts in Arctic wetland landscapes-literature review and gap analysis. *Water (Switz.)* 11 (722), 1–11.
- Shean, D., Alexandrov, O., Moratto, Z., Smith, B., Joughin, I., Porter, C., Morin, P., 2016. An Automated, Open-Source Pipeline for Mass Production of Digital Elevation Models (DEMs) from Very-High-Resolution Commercial Stereo Satellite Imagery. *ISPRS J. Photogramm. Remote Sens.* 116, 101–117.
- Shimada, M., Isoguchi, O., Tadono, T., Isono, K., 2009. PALSAR radiometric and geometric calibration. *IEEE Trans. Geosci. Remote Sens.* 47 (12), 3915–3932. <https://doi.org/10.1109/TGRS.2009.2023909>.
- Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J. Photogramm. Remote Sens.* 164, 152–170.
- Tassi, A., Vizzari, M., 2020. Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms. *Remote Sens.* 12 (3776), 1–17.
- Widhalm, B., Bartsch, A., Heim, B., 2015. A novel approach for the characterization of tundra wetland regions with C-band SAR satellite data. *Int. J. Remote Sens.* 36 (22), 5537–5556.
- Woo, M.K., Young, K.L., 2005. High Arctic wetlands: Their occurrence, hydrological characteristics and sustainability. *J. Hydrol.* 320, 432–450.
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* 44, 5844–5853.
- Yang, L., Driscoll, J., Sarigai, S., Wu, Q., Chen, H., Lippitt, C.D., 2022. Google Earth Engine and Artificial Intelligence (AI): A Comprehensive Review. *Remote Sens.* 14 (3253), 1–110.
- Zhou, L., Luo, T., Du, M., Chen, Q., Liu, Y., Zhu, Y., He, C., Wang, S., Yang, K., 2021. Machine learning comparison and parameter setting methods for the detection of dump sites for construction and demolition waste using the google earth engine. *Remote Sens.* 13 (787), 1–17.
- Zou, Z., Devries, B., Huang, C., Lang, M.W., Thielke, S., McCarty, G.W., Robertson, A.G., Knopf, J., Wells, A.F., Macander, M.J., Du, L., 2021. Characterizing wetland inundation and vegetation dynamics in the arctic coastal plain using recent satellite data and field photos. *Remote Sens.* 13 (1492), 1–22.