The Importance of Lake Littoral Zones for Estimating Arctic-Boreal Methane Emissions

Ethan Kyzivat[1,1,1], Laurence Smith[1,1,1], Fenix Garcia Tigeros[2,2,2], Chang Huang[3,3,3], Chao Wang[4,4,4], Theodore Langhorst[4,4,4], Jessica V Fayne[5,5,5], Merritt Harlan[6,6,6], Yuta Ishitsuka[7,7,7], Dongmei Feng[6,6,6], Wayana Dolan[4,4,4], Lincoln H Pitcher[8,8,8], Tamlin M Pavelsky[4,4,4], David Butman[2,2,2], Kimberly P Wickland[9,9,9], Mark M Dornblaser[9,9,9], Robert Striegl[9,9,9], and Colin Joseph Gleason[6,6,6]

[1]Brown University
[2]University of Washington
[3]Northwest Unversity
[4]University of North Carolina at Chapel Hill
[5]University of California
[6]University of Massachusetts Amherst
[7]Department of Civil and Environmental Engineering, University of Massachusetts Amherst
[8]University of Colorado Boulder
[9]United States Geological Survey

November 30, 2022

Abstract

Areas of lakes that support emergent aquatic vegetation emit disproportionately more methane than open water but are under-represented in upscaled estimates of lake greenhouse gas emissions. These shallow areas are typically less than ~1.5 m deep and can be estimated through synthetic aperture radar (SAR) mapping. To assess the importance of lake emergent vegetation (LEV) zones to landscape-scale methane emissions, we combine airborne SAR mapping with field measurements of vegetated and open-water methane flux. First, we use Uninhabited Aerial Vehicle SAR (UAVSAR) data from the NASA Arctic-Boreal Vulnerability Experiment (ABoVE) to map LEV in 4,572 lakes across four Arctic-boreal study areas and find it comprises ~16% of lake area, exceeding previous estimates, and exhibiting strong regional differences (averaging 59 [50–68]%, 22 [20-25]%, 1.0 [0.8-1.2]%, and 7.0 [5.0-12]% of lake areas in the Peace-Athabasca Delta, Yukon Flats, and northern and southern Canadian Shield, respectively). Next, we account for these vegetated areas through a simple upscaling exercise using paired methane fluxes from regions of open water and LEV. After excluding vegetated areas that could be accounted for as wetlands, we find that inclusion of LEV increases overall lake emissions by 21 [18-25]% relative to estimates that do not differentiate lake zones. While LEV zones are proportionately greater in small lakes, this relationship is weak and varies regionally, underscoring the need for methane-relevant remote sensing measurements of lake zones and a consistent criterion for distinguishing wetlands. Finally, Arctic-boreal lake methane upscaling estimates can be improved with more measurements from all lake zones.

The Importance of Lake Emergent Aquatic Vegetation for Estimating Arctic-Boreal Methane Emissions

Ethan D. Kyzivat[1], Laurence C. Smith[1], Fenix Garcia-Tigreros[2], Chang Huang[1, 3], Chao Wang[4], Theodore Langhorst[4], Jessica V. Fayne[5], Merritt E. Harlan[6], Yuta Ishitsuka[6], Dongmei Feng[10], Wayana Dolan[4], Lincoln H Pitcher[5, 8], Kimberly P. Wickland[7], Mark M. Dornblaser[7], Robert G. Striegl[7], Tamlin M. Pavelsky[4], David E. Butman[2,9], and Colin J. Gleason[6]

[1]Department of Earth, Environmental & Planetary Sciences and Institute at Brown for Environment & Society, Brown University, Providence, RI, 02912 USA

[2]School of Environmental and Forest Sciences, University of Washington, Seattle, WA, 98195 USA

[3]School of Urban and Environmental Sciences, Northwest University, Xi'an, Shaanxi, 710127 China

[4]Department of Earth, Marine and Environmental Sciences, University of North Carolina, Chapel Hill, NC, 27599 USA

[5]Department of Geography, University of California-Los Angeles, Los Angeles, CA, 90095 USA

[6]Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, MA, 01003 USA

[7] U.S. Geological Survey, Water Resources Mission Area, Boulder, CO, 80303 USA

[8]Cooperative Institute for Research in Environmental Sciences (CIRES). University of Colorado, Boulder. Boulder, CO, 80309, USA.

[9]School of Engineering and Environmental Sciences, University of Washington, Seattle, WA, 98195 US

[10]Department of Chemical and Environmental Engineering, University of Cincinnati, OH, 45221 USA

Corresponding author: Ethan D. Kyzivat (ethan.kyzivat@aya.yale.edu)

Key Points:

- We provide a first quantification of emergent vegetation area across 4,572 lakes in four Arctic-boreal study areas using airborne mapping.
- Lake emergent vegetation coverage varies regionally from 1 to 59 percent of lake area and seasonally to a lesser degree.
- Accounting for this coverage could increase Arctic-boreal lake methane upscaling estimates by 21 percent.

Abstract

Areas of lakes that support emergent aquatic vegetation emit disproportionately more methane than open water but are under-represented in upscaled estimates of lake greenhouse gas emissions. These shallow areas are typically less than ~1.5 m deep and can be estimated through synthetic aperture radar (SAR) mapping. To assess the importance of lake emergent vegetation (LEV) zones to landscape-scale methane emissions, we combine airborne SAR mapping with field measurements of vegetated and open-water methane flux. First, we use Uninhabited Aerial Vehicle SAR (UAVSAR) data from the NASA Arctic-Boreal Vulnerability Experiment (ABoVE) to map LEV in 4,572 lakes across four Arctic-boreal study areas and find it comprises ~16% of lake area, exceeding previous estimates, and exhibiting strong regional differences (averaging 59 [50–68]%, 22 [20-25]%, 1.0 [0.8-1.2]%, and 7.0 [5.0-12]% of lake areas in the Peace-Athabasca Delta, Yukon Flats, and northern and southern Canadian Shield, respectively). Next, we account for these vegetated areas through a simple upscaling exercise using paired methane fluxes from regions of open water and LEV. After excluding vegetated areas that could be accounted for as wetlands, we find that inclusion of LEV increases overall lake emissions by 21 [18-25]% relative to estimates that do not differentiate lake zones. While LEV zones are proportionately greater in small lakes, this relationship is weak and varies regionally, underscoring the need for methane-relevant remote sensing measurements of lake zones and a consistent criterion for distinguishing wetlands. Finally, Arctic-boreal lake methane upscaling estimates can be improved with more measurements from all lake zones.

Plain Language Summary

Lakes are one of the largest natural sources of the greenhouse gas methane and are especially common in high latitudes. Shallow, near-shore areas of lakes having emergent aquatic vegetation emit disproportionately more methane than open water areas but are under-represented in broad-scale estimates of lake greenhouse gas emissions. While lake depths are difficult to map from remote sensing, emergent vegetation, which typically grows in water less than ~ 1.5 m deep, can be detected via radar remote sensing. To assess the importance of these areas to landscape-scale methane emissions, we combine airborne radar mapping with field measurements of vegetated and open-water methane emissions. Zones of emergent vegetation vary regionally and comprise ~16% of lake area on average. A simple estimate that accounts for both open water and emergent vegetation methane emissions results in 21% increased overall lake methane emissions estimates. Emergent aquatic vegetation coverage has only a weak relationship with lake size, making it hard to predict. Therefore, to better estimate broad-scale methane emissions, we suggest using remote sensing to create lake vegetation distribution maps and measuring methane emissions from both vegetated and open water zones within lakes.

1 Introduction

Inland waters (lakes, reservoirs, rivers, and wetlands) are the single largest natural source of the greenhouse gas methane (CH_4) (Saunois et al., 2020). Lakes are estimated to be responsible for ~24% of all inland water emissions, second only to wetlands (Bastviken et al., 2011; Saunois et al., 2020). They emit methane via diverse pathways of diffusion, ebullition,

transport through aquatic plant tissue, and through a storage flux during turnover and/or ice melt in stratified lakes. Emissions are strongly dependent on temperature, sediment carbon content, redox environment, and gas transfer velocity (Bastviken, Cole, Pace, & Tranvik, 2004; Wik et al., 2016). Uncertainties in upscaling lake emissions therefore have vast spatial and temporal heterogeneities (Loken et al., 2019; Natchimuthu et al., 2016; Stephanie et al., 2020; Saunois et al., 2020).

Unlike for wetlands, there are few process-based models for lake fluxes, so estimates have relied on data-driven extrapolations (Saunois et al., 2020). Lake emission upscaling efforts have only recently begun to account for lake surface area (DelSontro et al., 2016; Hastie et al., 2018; Holgerson & Raymond, 2016), but it is still rare to consider other aspects of morphometry, such as slope, vegetation, and littoral area (Casas-Ruiz et al., 2021). "Bottom-up," or process-based, methane estimates tend to over-predict aquatic methane fluxes compared to "top-down," or inversion-based, models (Saunois et al., 2020), and double-counting of small lakes as wetlands caused by mismatch in scale and methods among datasets has been suggested as a possible cause (Thornton et al., 2016). Small (< 0.001 km^2) lakes and wetlands are poorly mapped, especially in Arctic-boreal regions containing the world's greatest abundance of lakes (Verpoorter et al., 2014). Indeed, uncertainty in wetland extent is frequently cited as the leading cause of uncertainty in bottom-up methane estimates (Zhang et al. 2017), and errors arising from large-scale extrapolations of heterogeneous wetlands have also been noted (Bridgham et al., 2013).

One key challenge to upscaling is the high within-lake spatial variability of methane emissions. Total fluxes measured from vegetated (Villa et al., 2021) and shallow (Natchimuthu et al., 2016) zones can be statistically greater than those from open water and have been attributed to the majority of whole-lake emissions (Saunois et al., 2020). Estimates derived from deep lake centers have been shown to underestimate total flux by 5-78% in select lakes (Natchimuthu et al., 2016). Plant-mediated fluxes can be significant at the landscape scale, for example exceeding peatland emissions in southern Finland by 30%, despite covering only 40% as much area (Bergström et al., 2007). Another study of three Finnish lakes found that the vegetated littoral zone produced 66-77% of whole-lake emissions (Juutinen et al., 2003). Combined globally, emergent macrophytes are estimated to emit 11% of the equivalent from all open water lakes, rivers, and reservoirs (Bastviken et al., 2011). As the most "wetland-like" zone within lakes, littoral zones are important sources of carbon and known methane emission hot spots (Bergström et al., 2007; Burger et al., 2016; Huttunen et al., 2003; Juutinen et al., 2003; Larmola et al., 2004), with exceptions (Jansen et al., 2020a). However, littoral zone area is difficult to quantify accurately because its extent is classified by light penetration into the water column (Wetzel, 2001) and not by characteristics that are easily delineated by remote sensing. In contrast, the extent of emergent macrophytes growing in water < ~1.5 m deep in the upper littoral zone are more easily detected. These plants can act as conduits to the atmosphere for methane produced in lake sediments (Dacey and Klug, 1979; Colmer, 2003). They also produce carbon compounds that are preferentially consumed by methanogens (methane-producing archea), and their decomposing biomass and root exudates are a large contributor to sediment organic carbon (Christensen et al., 2003; Joabsson, Christensen, & Wallén, 1999; Ström et al., 2005). Previous studies have noted the tendency for small (Michmerhuizen, Striegl, & McDonald, 1996; Bastviken et al., 2004; Holgerson & Raymond, 2016; Engram et al. 2020) and shallow (West et al., 2015; Wik et al., 2016a; Li et al., 2020) lakes to emit more methane than larger and deeper ones. Delsontro et al. (2018b) successfully modeled lake methane

concentration as a function of distance from the littoral zone, horizontal transport and oxidation, and oxic epilimnetic production, which highlights the outsized importance of littoral methane production. Notably, not all properties of littoral zones come from their vegetation. Their relative shallowness is also a factor, as depth often prohibits methane ebullition due to water overburden pressure (Bastviken et al., 2004, Langenegger et al., 2019), although there are exceptions (Huttunen et al., 2003). Shallow waters may also contain distinct sediment organic matter composition and less opportunity for microbe-mediated oxidation of dissolved methane (DelSontro et al., 2016). Finally, diffusive fluxes measured in the littoral zone may be driven by terrestrial inflows (Paytan et al., 2015, Natchimuthu et al., 2016), and offshore fluxes are diminished by oxidation during transport (DelSontro et al., 2018b). Thus, methane emissions in lakes are spatially variable, with highest emissions coming from littoral zones, particularly with vegetation.

This challenge of accounting for spatial heterogeneity is exacerbated by lack of data in the littoral or vegetated zones (DelSontro et al. 2018b; Desrosiers et al., 2022). The Boreal–Arctic Wetland and Lake Methane Dataset (BAWLD-CH_4; Kuhn et al., 2021a; Kuhn et al., 2021b) is the first synthesis study we are aware of that notes which part of the lake ebullition fluxes were measured (center, edge, or whole lake). However, only 143 of the 553 records actually contain within-lake location, and of these, only one was measured from an edge, with 19 from centers and 123 from whole-lakes. Among lake methane studies, plant-mediated emissions are measured least frequently of all lake pathways (Bastviken et al., 2011; Wik et al., 2016a), along with open-water emissions near plants, so methane upscaling estimates in lakes (DelSontro et al., 2018a; Tranvik et al., 2009) usually rely solely on pelagic diffusion and ebullition (DelSontro et al. 2018; Desrosiers et al., 2021), with biases introduced by insufficient within-lake sampling sites (Wik et al., 2016b). For these reasons, lake methane measurements are under-represented in vegetated and littoral zones, even among the few studies that report sampling location.

Another key challenge to upscaling is that littoral and vegetation coverage in lakes are poorly constrained. Duarte et al. (1986) suggested that emergent macrophytes colonize on average 7% of a lake regardless of its area, while submerged macrophyte coverage generally declines with area. They list light availability, sediment characteristics, and trophic status as key characteristics for macrophyte growth, with slope as the greatest predictor of emergent macrophyte coverage. Others have theorized that the percent of a lake's surface area covered with macrophytes scales with nitrogen concentration and the inverse of mean depth (Smith and Wallsten 1986), or scales inversely with lake area (Michmerhuizen et al., 1996) or perimeter (Bergström et al., 2007). Mäkelä et al. (2004) similarly found that an average of 6% (range: 1-100%) of total lake area was covered by macrophytes in a sample of 50 lakes and that total fractional macrophyte coverage per lake steeply declined with lake area. Zhang et al. (2017) compiled a synthesis database of aquatic macrophytes in 155 global lakes and observed an average coverage of 26% (range: 0.000-100%) with an accelerating decline since 1900.

Remote sensing studies have used both optical and synthetic aperture radar (SAR) sensors to map macrophytes in lakes. Optical satellites are better suited to detecting vegetation type, while SAR can detect water even through vegetation canopies (Hess et al., 1990). Ghirardi et al. (2019) used optical Sentinel-2 satellite data to map submerged aquatic macrophytes in an Italian lake and noted both inter- and intra-annual variations in aerial coverage. Nelson et al. (2006) used Landsat Thematic Mapper imagery to map various types of macrophytes in 13 lakes

in Michigan, USA and found total macrophyte coverage ranging from 5-42%. Ganju et al. (2017) used air imagery and elevation data to derive the unvegetated/vegetated marsh ratio (UVVR) for tidal marshes, which scales with sediment budget and has typical values < 0.4. Zhang et al. (2018) used TerraSAR-X SAR imagery to map macrophytes in nine Brazilian reservoirs and similarly found large spatial and temporal variation in coverage. Thus, many remote sensing studies have demonstrated spatial and/or temporal differences in aquatic macrophyte cover, yet few have measured total coverage across large geographical areas and numerous lakes. Lake macrophyte area statistics, therefore, remain confined to a handful of studies of small numbers of lakes.

Here, we aim to quantify the fractional coverage of emergent vegetation for 4,572 lakes across four Arctic-boreal regions in order to assess their potential importance in scaling methane emissions. To estimate coverage, we use the canopy-penetrating properties of L-band synthetic aperture radar (SAR) flown during the NASA Arctic-Boreal Vulnerability Experiment (ABoVE) airborne campaign (2017-2019). Although floating-leafed macrophytes are relevant to the methane budget, they cannot be reliably detected with this technique due to similar surface roughness with water waves and thus are omitted here. Next, we compile paired measurements of methane flux (new data and literature) via all pathways from open water and emergent macrophyte regions of lakes. Finally, we use these flux measurements and our remote sensing-derived ranges in emergent vegetation coverage to estimate its impact on lake methane emissions. We conclude with discussion of the causes of regional differences, some broader recommendations for landscape-scale methane upscaling, study limitations, and recommendations for future research.

2 Study areas, data sources, and methods

2.1 Study areas

The NASA Arctic-Boreal Vulnerability Experiment (ABoVE) campaign is a decade-long effort to measure environmental change in the Arctic and boreal regions of western North America via coordinated ground measurements and airborne remote sensing (Miller et al., 2019). Here, we focus on four study areas within the ABoVE domain, each corresponding to one or more flight lines from its airborne campaigns:

1) Peace-Athabasca Delta, Alberta, Canada (PAD);
2) Southern Canadian Shield near Baker Creek (CSB), Northwest Territories, Canada;
3) Interior Canadian Shield near Daring Lake (CSD), Northwest Territories, Canada; and
4) Yukon Flats National Wildlife Refuge, Alaska, USA (YF).

These four study areas were chosen because of their high lake density and contrasting geological, hydrological, and ecological conditions. The PAD is one of the world's largest inland deltas and is located on the western edge of Lake Athabasca (**Figure 1**). The overall relief of its lowland regions is 11 m, causing numerous marsh-type wetlands, mudflats, and lakes, many of which are recharged by the Athabasca River (Pavelsky & Smith, 2008), and more rarely, by ice-jam floods in the Peace River (Timoney, 2013). These floods can inundate up to 80% of the 5,600 km^2 delta (Töyrä & Pietroniro, 2005; Wolfe et al., 2006), while in typical years, 26% is covered by intermittently-inundated wetlands (Ward & Gorelick 2018). It is a Ramsar Wetland, UNESCO World Heritage site, and home to numerous endemic species of birds, fish, and mammals

including the endangered whooping crane and the largest remaining herd of wood bison (Parks Canada, 2019). The two Northwest Territories study areas (CSD, CSB) are located on the Canadian Shield, the world's largest deposit of Precambrian-age bedrock and source of the oldest known terrestrial rocks (Slaymaker, 2016). Deglaciated only nine thousand years ago and with a rocky, sparse surface drainage pattern, the Shield is also the world's most lake-rich region and contains many peatlands (Slaymaker, 2016; Spence & Woo, 2006). CSB is underlain by discontinuous permafrost, while CSD crosses the tree line and contains a transition to continuous permafrost and the tundra/taiga ecotone (**Figure 1**). The YF is underlain by discontinuous permafrost in alluvial soils and contains lakes of various hydrologic connectivity to the Yukon River and its tributaries (Anderson et al. 2013, Johnston et al., 2020). Like the PAD, the YF has flat topography, permitting seasonal flooding during the early summer to cover large areas, and it is a source of both lateral riverine and water-air carbon fluxes (Striegl, et al., 2012). All four study areas are home to multiple indigenous and First Nation communities, as well as the city of Yellowknife (CSB) and numerous smaller settlements.

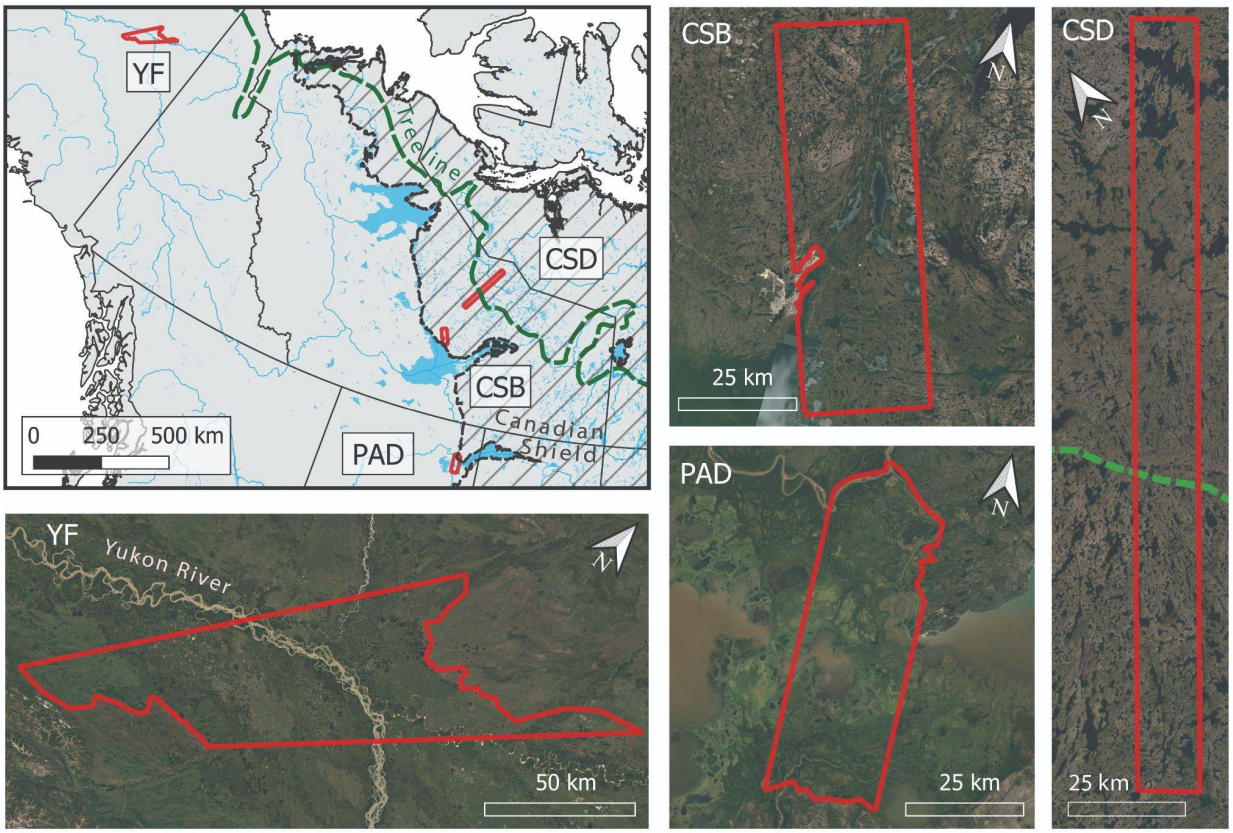

Figure 1. Location map of study areas (YF = Yukon Flats; CSD = Canadian Shield, Daring Lake; CSB = Canadian Shield, Baker Creek; PAD = Peace-Athabasca Delta). Study area boundaries (red polygons) are derived from intersecting UAVSAR airborne flight coverage with physiographic boundaries. Major water bodies are shown in blue; Canadian Shield with stippling, and the northern tree line limit (Brown et al., 2002) in green.

2.2 Data sources

2.2.1 Airborne polarimetric SAR

L-band synthetic aperture radar (SAR) data from the Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) were obtained in multi-look ground-projected format (GRD) and reprojected to ~5.5 m spatial resolution (NASA/JPL 2017-2019) on the ABoVE Science Cloud computing environment. With a wavelength of 23.8 cm, UAVSAR has been used extensively for vegetation mapping and inundation detection, including in lowlands or deltas with flooded vegetation (Ayoub et al., 2018; Jensen et al., 2021; Z. Zhang et al., 2017). All available ABoVE UAVSAR flight dates from non-contiguous days during summers 2017-2019 were used. Both early (June) and late (August-September) summer images were acquired by UAVSAR in 2017, and only late summer/early autumn dates were imaged in 2018 and 2019.

2.2.2 Water and land cover maps

Several ABoVE land cover data sets were referenced to help build a land cover training dataset for UAVSAR (see **Section 2.3.1**). High-resolution imagery and derivative water masks were obtained from the AirSWOT color-infrared camera (Kyzivat et al. 2018; Kyzivat et al. 2019; Kyzivat, et al. 2020), supplemented by high-resolution satellite imagery from Maxar (https://evwhs.digitalglobe.com/myDigitalGlobe/). Two satellite-based land cover maps available for the ABoVE domain were also referenced (Bourgeau-Chavez et al., 2017, 2019; Wang et al., 2019; Wang et al., 2019). Although these maps use a different classification scheme than our derived UAVSAR classification, they are particularly useful for partitioning between trees, shrubs, and graminoid vegetation.

2.3 Methods

2.3.1 Land cover classification training dataset

To estimate lake emergent macrophyte coverage (A_{EV}), a land cover training dataset was created using inundation status from field measurements in 2015 and 2017-2019 and vegetation categories from ABoVE land cover maps (Bourgeau-Chavez et al., 2017, 2019; Wang et al., 2019; Wang et al., 2019). As part of the field measurements, lake and wetland shorelines and vegetation zones were mapped by field teams carrying handheld GPS receivers, as described in Kyzivat et al. (2019). In YF, airborne GPS tracks from a low-hovering helicopter were used, as no suitable ground GPS tracks were available. Contextual photos were also taken by camera, both from the ground and from aircraft windows, and by uninhabited airborne vehicles (UAVs). UAV photos were processed into orthomosaics using DroneDeploy web software. All of these measurements were digitized into polygon shapefiles in ArcGIS 10.6 denoting 13 land cover classes falling into five broad categories of open water, dry land and three types of emergent vegetation (**Table 1)**. The resulting vector data set was used to train and validate a supervised classification from the radar data (Kyzivat et al., 2021a).

Broad Grouping	**UAVSAR land cover class**

Open surface water	Open Water (OW), Rough Water (RW), Sedimentary Bar (SB), Wet Herbaceous (WH)
Wet Graminoid	Wet Graminoid (WG)
Wet Shrub	Wet Shrub (WS)
Wet Forest	Wet Forest (WF)
Dry land	Dry Graminoid (DG), Dry Shrub (DS), Dry Forest (DF), Bank Scarp Double-Bounce (BS), Dry Woodland (DW), Bare Ground (BG)

Table 1. Classification Schema: RW refers to wind roughening at the time of acquisition. WG refers to cattails (*Typha latifolia*), bulrushes (*Scirpus* spp.), and sedges (*Carex* spp.), as well as aquatic horsetails (*Equisetum fluviatile*). WS typically refers to willows (*Salix* spp.). DW refers to a mix of trees and shrubs as defined by Wang (2019). WH refers to water lilies (*Nuphar variegatum*), and both WH and SB were not separable from the other open water classes. Further details are in the accompanying data publication (Kyzivat et al., 2021a).

2.3.2 Synthetic aperture radar data pre-processing

UAVSAR GRD data for the PAD, YF and CSB flight lines were transformed to the C3 complex covariance matrix using PolSAR Pro 6.0 software. Images were corrected for incidence angle-dependent backscatter using a fitted exponential function multiplied by the cosine of incidence angle as per Ulander (1996) and Zhang et al. (2017). Due to its more rugged topography, CSD was corrected for both incidence angle and terrain slope as per the look-up table method of Simard et al. (2016). For all flight lines, a Freeman-Durden polarimetric decomposition was performed. The decomposition comprises a physical scattering model and is commonly used to identify scattering mechanism contributions to each pixel (single bounce, modeled as Bragg scattering; double bounce, modeled as from a pair of orthogonal surfaces; and volume scattering, modeled as from a cloud of randomly-oriented dipoles) (Freeman & Durden, 1998). Although it is known to overestimate the double bounce component (Chen et al., 2014), it is sufficient as an input feature to an empirical, machine-learning based classification.

2.3.3 Land cover classification

Each of the three scattering mechanism output bands was used for feature extraction via three moving-window filters designed to introduce spatial contextual information for the classifier. The chosen filters were standard deviations, offsets oriented along the radar look direction, and an edge-preserving guided filter to reduce speckle (**Table S.2**). Additional input bands of incidence angle and elevation-derived indexes were tested, but ultimately omitted, due to their high spatial autocorrelation, which led to model over-fitting. The training class BS was developed specifically to identify bright double bounce scattering between water surfaces and steep bank scarps, which would otherwise have appeared as inundated vegetation. SB and WH (defined as protruding <20 cm from the water surface, as determined from field measurements) were found to be inseparable from OW, so they were treated as open surface water in the analysis. The radar dataset was further prepared for classifier training by randomly under-sampling the majority training classes and cropping out pixels taken at low incidence angles.

Incidence angle limits as well as filter parameters (**Table S.2**) were chosen by trial and error. Finally, pixel values within training polygons in all input bands from the appropriate date were extracted, and the results split using stratified sampling into training (85%) and validation (15%) datasets with 15 bands each. A description of this workflow, parameter settings, and other technical details is provided in **Table S.2**.

Finally, a random forests classifier was trained using the TreeBagger function in MATLAB R2017b and evaluated using the validation dataset via the confusion matrix and Cohen's kappa coefficient. One model was used for the areas with incidence angle correction and another for the CSD area with the look-up table correction. The models were then applied over the extent of their corresponding study areas for all available dates. The original 13 classes were aggregated into the five generalized classes for analysis (**Table 1**).

2.3.4 Quality control and conversion to emergent vegetation coverage

The derived five-class land cover maps were used to identify emergent macrophyte and open water areas and quantify their total landscape coverage. First, maps were clipped to the intersection of all flight lines per study area excluding any roads or urban areas, if present. Raster mosaics were created for the PAD and YF, since they were acquired in multiple flight lines on most dates (**Table S.1**). Next, candidate lakes were identified as connected pixel groups of at least five pixels with at least one open water pixel and any number of inundated vegetation pixels (or none at all). This criterion permitted inclusion of open water wetlands, because there is no reliable way to differentiate them from lakes and ponds. Rivers were removed by applying a manually-created river mask, modified from Kyzivat et al. (2019). Lake emergent vegetation (A_{LEV}) were operationally defined as emergent vegetation classes 8-connected to lakes, with the remaining emergent vegetation pixels considered wetlands (A_{WEV}). Although dependent on pixel size, this definition permitted a consistent definition across all study areas. At this stage, the total landscape coverage of A_{LEV} (wet graminoid, shrub, and forest classes) and open water were calculated so they could be compared between dates.

Although there is scarce data for methane emission from trees and shrubs along lake shores, we included them in the sensitivity analysis because: 1) 69% of A_{LEV} is comprised of graminoid vegetation and this value increases to >97% after correcting for double counting (see **3.1.1**); 2) There is no mixed coverage class, meaning there is likely still graminoid vegetation present, but hard to detect; 3) Data scarcity makes it hard to account for them separately; and 4) Many of the factors that make vegetated water surfaces high emitters are shared between vegetation types, such as shallowness, proximity to terrestrial inputs, variable inundation, and presence of root systems. In fact, these dynamically-inundated water surfaces with woody vegetation, which could also be called littoral swamps, have been shown to emit methane four orders of magnitude greater than temperate forest soil uptake (Hondula et al., 2021). This observation underscores the importance of accounting for regions of emergent lake vegetation separately from open water, while being sure to exclude any regions otherwise accounted for as wetlands (see **2.3.7**).

To calculate A_{LEV} coverage on a per-lake basis, water bodies smaller than 250 m^2 (0.00025 km^2 or 7-8 px) were discarded, since they were too small to consistently resolve and likely included false detections. Although hardly affecting total lake area, false detections of

lakes would be disproportionately small and thus impact the distribution of A_{LEV}. Partially observed lakes intersecting the flight line boundary were discarded as well, since A_{LEV} could not be reliably measured. A third category of lakes were discarded if they did not overlap with any water pixels in the 2017 AirSWOT color-infrared camera open water masks, which had a slightly narrower ground footprint in all study areas. By comparing our UAVSAR retrievals to an independent, optical data set, this step removed many falsely-identified lakes caused by classification error. Finally, we calculated the areas of the remaining lakes and the fractional area of their emergent vegetation (A_{LEV}) coverages, defined as the proportion of pixels in a lake classified as any of the three inundated vegetation classes. For visualization and analysis, these data were divided into 24 logarithmically-spaced lake area bins across the four study areas, and the mean, lake area-weighted mean, and median A_{LEV} computed for each study area. For each study area, confidence intervals were calculated for each of the 24 bins and for the area-weighted means using the 95th percentile of 10,000 bootstrapped simulated datasets.

2.3.5 Adjusting estimate to avoid double-counting wetlands

Our method for detecting emergent vegetation excludes wetlands based on lack of pixel connectivity to open water. Although this method conserves total area and thus does not double-count any pixel to more than one land cover class, this partitioning includes open-water and littoral wetlands as parts of lakes. As a result, our estimate of A_{LEV} would be too high because it treats areas typically considered to be wetlands (e.g. in methane models) as parts of lakes, which is precisely the double-counting between datasets described by Thornton et al. (2016). To correct for this over-estimate of total lake area, we obtained two leading global lake datasets, GLOWABO (Verpoorter et al., 2014) and HydroLAKES (Messager et al., 2016) and compared total lake extent between the datasets and our own. First, since the global datasets were made at a coarser geographic scale, USAVSAR lakes below the appropriate minimum size threshold were excluded (0.002 km^2 for GLOWABO and 0.1 km^2 for HydroLAKES). Even so, there were still many more lakes detected by UAVSAR (and some only detected by one of the other datasets), so spatial selection in the python package geopandas 0.10.2 (Jordahl et al., 2021) was used to exclude any lakes in either dataset that did not overlap at least partly with a lake in the dataset to which it was being compared. This exclusion ensured that we were only comparing areas within commonly-detected lakes and not simply assessing lake mapping accuracy between the datasets, which have vastly different scales and time domains. Next, both datasets were rasterized to the UAVSAR pixel grid for the corresponding scene, typically 5.5 by 5.5 m pixels. Then, for each study area, a confusion matrix was computed between the UAVSAR dataset and each of the others for all pixels not denoted as land in both candidate datasets. These matrices were used to compute the scalar c, which is used in **Equation [1]** and denotes how much of UAVSAR A_{LEV} falls within global dataset lakes, with the remainder assumed to already be mapped as wetlands with adequate accounting of methane emissions.

The calculation ignores the effects of changing inundation during the 10-20 years between data acquisitions, as well as errors arising from the global datasets having less-precise georeferencing. It is also limited to only the large lakes that could be compared between datasets. Since these biases would also exist in any modeling study using GLOWABO or HydroLAKES, we have made no attempt to correct for them, which would also be beyond the scope of this work.

2.3.6 Methane flux chamber measurements

24 methane fluxes were measured at 15 lakes in the PAD during July and August 2019 (Kyzivat et al. 2021, **Figure S.6**). The sampling schedule permitted no more than one or two visiting days per lake, so the measurements represent a broad, geographic sampling within the PAD at the expense of frequent measurements in any one lake. This sampling approach allowed for better, but still limited extrapolation to the 470 UAVSAR-observed lakes in the PAD. In all 15 lakes, single 15-minute fluxes were taken from an open water region near the lake center via inflatable raft, anchored canoe, or motorboat. In five lakes, one to three additional flux measurements were made amidst emergent macrophytes of different species (corresponding to the wet graminoid land cover class) short enough to fit into the flux chamber without excessive disturbance. The chamber comprised an inverted 25.4 cm tall, opaque white bucket with a 34.2 cm diameter opening wrapped with a buoyant skirt made of foam tubing. An infrared greenhouse gas analyzer (EGM-4, PP Systems) was used to measure chamber air carbon dioxide (CO_2) concentration and circulate chamber air via an inlet on the side of the chamber and an outlet in the center of its ceiling. A metal handle was used to steady the bucket for a 15-minute measurement period. At 0, 5, 10, and 15 minutes, gas samples were drawn from the chamber's headspace through the gas analyzer inlet tubing and injected into evacuated exetainers using a 30 mL polypropylene syringe fitted with a 3-way stopcock for subsequent analyses of methane concentration.

The samples were analyzed on a Shimadzu GC-2014 gas chromatograph for methane partial pressure within two months of collection. Gas flux across the water-air interface was calculated from the rate of change in the chamber methane concentration over the deployment time and chamber area ($mol \cdot min^{-1} \cdot m^{-2}$). The rates of change of methane concentrations in the chamber were generally linear with r^2 values greater than 0.90. Given this linear response, ebullition was deemed negligible during the measurement periods. Thus, the closed, static chamber measurements included both diffusive fluxes from the water surface as well as any plant-mediated fluxes. For the three lakes where multiple emergent macrophyte fluxes were taken at one location, measurements from each water zone were averaged by lake. Finally, for sites where paired open water vs. littoral zone measurements were collected, we calculated the vegetated: open water flux ratio (hereafter: flux ratio) as the ratio between the average emergent macrophyte and open water measurements for each lake, where open water could include submerged macrophytes not detectable with UAVSAR.

During sampling, care was taken not to disturb the sediment, and if any bubbles were observed before or during the period, the measurement was aborted. Even so, three measurements were extremely high, implying sediment disturbance. To avoid potential bias, these measurements, which were greater than 2.2 standard deviations from the median, were discarded (the next-highest value was 0.17 standard deviations from the median). These three measurements all came from vegetated sites, so this data omission lessened the impact of emergent vegetation in our subsequent analyses.

2.3.7 Published flux chamber measurements

In addition to our own field measurements, we compiled a synthesis dataset of 58 paired flux measurements, with the aim of determining the flux ratio for each lake. Six of these

measurements corresponded to shallow (typically with a 2-4 m cutoff) versus deep regions of the lake, with no mention of adjacent macrophytes, and were only included for reference, while the remaining 52 were taken from vegetated versus open water fluxes, and were used for subsequent calculations. Each lake pair corresponded to one of 41 distinct lakes or lake regions during a single or multi-year-averaged sampling season, published in 21 papers (Kankaala et al. 2005; 2013; Smith and Lewis 1992; Larmola et al. 2004; Huttunen et al. 2003; Juutinen et al. 2003; Villa et al. 2021; Burger et al. 2016; DelSontro et al. 2016; Bergström et al. 2007; Striegl and Michmerhuizen 1998; Ribaudo et al. 2012; Casper et al. 2000; Dove et al. 1999; Elder et al., 2022; Rey-Sanchez et al., 2018; Desrosiers et al., 2021; Engram et al. 2020; Natchimuthu et al., 2016; Wik et al., 2013, Jansen et al., 2020a; **Table S.3**). Lakes included boreal, tropical and temperate regions and were located in Finland, Quebec, Ontario, Alaska, Colorado, Ohio, Minnesota, Italy, the UK, and the Amazon and Orinoco river basins. For each paper, the average—whether seasonal or annual—vegetated and open water measurements were recorded and converted, if necessary, to units of mg CH_4/m^2/day. Four papers (Burger et al., 2016; Casper et al., 2000; Dove et al., 1999; Desrosiers et al., 2021) separately measured each of the three methane emission pathways, and most of the others focused on diffusion and/or plant-mediated fluxes. An additional six (Huttunen et al., 2003; Juutinen et al., 2003; Larmola et al., 2004; Striegl and Michmerhuizen, 1998; Jansen et al., 2020a; Villa et al., 2021) measured diffusion and ebullition in both lake zones, but did not place the flux chamber over plants, thus not accounting for that pathway. One study (Bergström et al., 2007) did not provide open water values, which we estimated based on lake area via the relationship of Holgerson and Raymond (2016). The dataset includes 55 diffusion, 40 plant-mediated, and 17 ebullition pairs, with some measurements counting towards multiple pathways.

The vegetated: open water flux ratio *R* was calculated for each applicable lake (including our field lakes) and divided by a correction factor of 1.33 to account for most measurements being made either during ice-covered or ice-free seasons, but not during ice melt, when open-water emissions can temporarily spike. The correction factor, averaged from Wik et al. (2016a) and Denfeld et al. (2018), comes from statements that 23% and 27% of emissions of ice-covered lakes, respectively, are attributed to ice-melt fluxes. Although the lake upscaling calculation by Rosentreter et al. (2021) also uses a spatiotemporal ice-cover correction with the opposite effect of the ice-melt pulse correction, we have omitted it here, assuming it affects both vegetated and unvegetated areas equally. The adjusted flux ratio *R'* therefore comes from measurements of three methane flux pathways, collected from both littoral vegetation and shallow open water in all seasons, and reflects adjustments to account for unmeasured ice-melt pulses.

Many papers stated the area covered by emergent macrophytes, but if not, Google Earth Pro and QGIS 3.10.11 were used to digitize, map project, and measure the approximate coverage area, with attention paid to the papers' description of the vegetation for context. Coverage areas were assigned an uncertainty value (typically 2–5%) based on interpretation of the methods used or confidence in our digitizing result. Although challenging to compare across methodologies, geographic regions, and plant types, this dataset served as a best estimate of flux ratios from a diverse global sample of lakes.

2.3.8 Sensitivity analysis

Likely ranges in whole-lake methane emissions were calculated using the following equation, mapped lake areas, and the compiled flux dataset:

$$F_{total} = c * A_{LEV} * \Omega * R' * f_{OW} + (1 - c * A_{LEV} * \Omega) * f_{OW} \quad [1]$$

where F_{total} is the total lake flux (mg CH_4/day), calculated as a weighted average of vegetated and open water zones; c is a scalar ≤ 1, described in section 2.3.7, that corrects for potential double-counting of UAVSAR-observed emergent vegetation as wetlands contained in modeling datasets (unitless); A_{LEV} is the emergent vegetation area as a fraction of total lake area (unitless); Ω is the total lake area (m^2), f_{OW} is the flux per area of open water (mg CH_4/m^2/day); and R' is the corrected ratio between emergent macrophyte and open water fluxes per area (unitless). All areas and fluxes are expressed relative to the total lake area Ω, and the flux per unit area of open water (f_{OW}), both of which cancel out when applying equations [1] and [2].

The impact of vegetation on whole-lake flux was calculated as a percent difference via:

$$I = \frac{F_{total} - f_{OW} * \Omega}{f_{OW} * \Omega} \quad [2]$$

where I represents the percent increase from differentiating between open water and emergent vegetation within lakes. I is sensitive only to the measured parameters R', A_{LEV}, and c, and independent of the absolute magnitudes of the fluxes or areas attributed to each lake zone, which cancel out.

Equations **[1]** – **[3]** were applied using the median values of R' and f_{OW} and the lake area-weighted mean A_{LEV}. Median values were used due to the skewed distributions of R' and f_{OW}. The equations were also applied to the bootstrapped confidence intervals of A_{LEV} in order to estimate uncertainty.

3 Results

3.1 Inundation patterns at the landscape scale

3.1.1 Regional and seasonal inundation characteristics

Significant open water, emergent vegetation, and wetland fractional areas are found in all study areas, vary seasonally as well as regionally, and are particularly extensive in the PAD and YF. The total area of the landscape covered by lake emergent vegetation (LEV) varies from 0.5 – 0.6 % (CSD), 2.2 – 3.4 % (CSB), 7.6 – 15.5 % (PAD), and 1.7 – 2.8 % (YF) over the 2017-2019 observational period (**Figure 2, Table 2**). In comparison, wetland emergent vegetation (A_{WEV}) covers ≤ 2.7% of the area in all sites (mean of 1.4%, **Table 2**). Most of the emergent vegetation is classified as either wet graminoid (WG, weighted mean of 69%) or shrub vegetation (WS, 29%), with wet forest comprising <1% of this area for all areas except YF, for which it covers a mean of 5.9%. When only considering LEV that falls within a global dataset lake (the double-counting correction), the graminoid fraction increases to 99.1% (GLOWABO) or 98.7% (HydroLakes), which provides further confidence that the remaining LEV is indeed littoral vegetation and not an adjacent, forested wetland, at least for large lakes in the global datasets. Virtually all detected emergent vegetation lies adjacent to shorelines, with < 0.2% of their area

occurring completely within a lake with no connectivity to non-island land. These patterns show that the dominant littoral vegetation type in the study areas is graminoids, which almost always occur at the interface between land and water.

In all applicable study areas, total inundation (open water plus emergent vegetation) is greater or equal in the early summer (June) than in late summer (August/September), likely due to snowmelt. In the PAD, this change is caused by decreased LEV, with emergent wetland vegetation remaining constant, implying that seasonal inundation changes occurred in flood-tolerant eulittoral vegetation (**Figure 2, Table 2**). Thus, regional variations in emergent vegetation, as well as open water, are greater than seasonal/interannual variations within study areas.

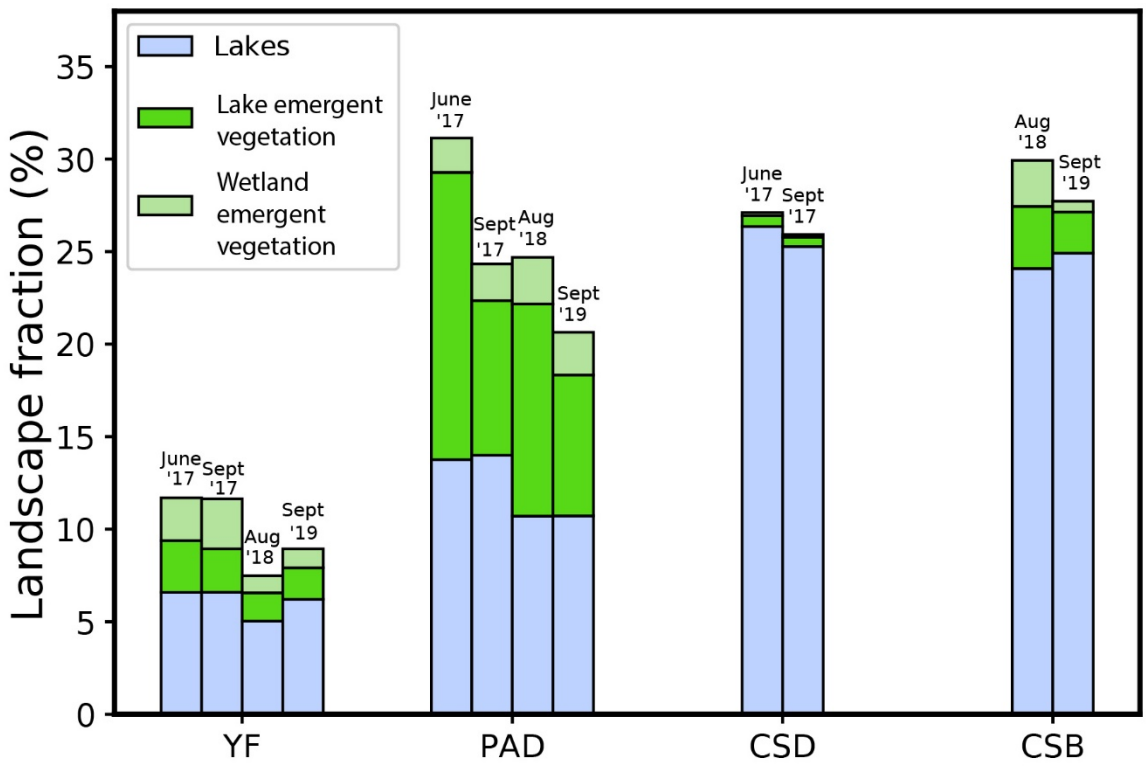

Figure 2. Significant lake emergent vegetation (LEV) is found in all study areas, varies seasonally as well as regionally, and is particularly extensive in the lowland PAD and YF. This chart shows landscape fractional areas of open water and LEV classes for the Yukon Flats (YF), Peace-Athabasca Delta (PAD), Canadian Shield – Daring Lake (CSD), and Canadian Shield – Baker Creek (CSB), derived from airborne UAVSAR. LEV is defined as emergent vegetation

adjacent to open water, with remaining areas assigned to wetlands (WEV). Month and year of UAVSAR flight acquisitions appear in text above each column.

				Lake fraction (%)						Landscape area (km^2, %)					
	Study area	**Extent (km^2)**	**Lake count**	**A_{LEV}**	**A_{WF}**	**A_{WS}**	**A_{WG}**	**A_{LEV} (median)**	**A_{LEV} (unweighted)**	**Lake open water**	**LEV**	**WF**	**WS**	**WG**	**WEV**
CSD June 2017	CSD	3037	1918	1.1 [0.9, 1.4]	0.0 [0.0, 0.0]	0.0 [0.0, 0.1]	1.1 [0.9, 1.3]	0.0%	2.0%	800 (26.4%)	18 (0.6%)	0 (0.0%)	1 (0.0%)	17 (0.6%)	3 (0.1%)
CSD Sept 2017	CSD	3037	1975	0.9 [0.6, 1.1]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.8 [0.6, 1.1]	0.0%	3.8%	767 (25.3%)	16 (0.5%)	0 (0.0%)	0 (0.0%)	15 (0.5%)	2 (0.1%)
CSD		3037	1947	1.0 [0.8, 1.2]	0.0 [0.0, 0.0]	0.0 [0.0, 0.1]	1.0 [0.7, 1.2]	0.0%	2.9%	784 (25.9%)	17 (0.5%)	0 (0.0%)	1 (0.0%)	16 (0.5%)	3 (0.1%)
CSB Aug 2018	CSB	1155	376	8.6 [5.8, 14.1]	0.0 [0.0, 0.1]	2.3 [1.7, 3.6]	6.2 [4.1, 10.5]	20.4%	26.6%	278 (24.1%)	39 (3.4%)	0 (0.0%)	11 (1.0%)	28 (2.4%)	29 (2.5%)
CSB Sept 2019	CSB	1160	378	5.5 [3.6, 9.0]	0.0 [0.0, 0.1]	0.7 [0.5, 1.1]	4.7 [3.1, 7.9]	11.3%	17.5%	289 (24.9%)	26 (2.2%)	0 (0.0%)	4 (0.3%)	22 (1.9%)	7 (0.6%)
CSB		1158	377	7.0 [4.7, 11.5]	0.0 [0.0, 0.1]	1.5 [1.1, 2.3]	5.5 [3.6, 9.2]	15.9%	22.1%	284 (24.5%)	32 (2.8%)	0 (0.0%)	7 (0.6%)	25 (2.1%)	18 (1.5%)
PAD June 2017	PAD	1339	347	65.5 [56.5, 75.3]	0.7 [0.2, 1.3]	35.3 [28.2, 42.6]	29.5 [21.4, 38.8]	63.5%	58.3%	184 (13.8%)	208 (15.5%)	2 (0.1%)	73 (5.4%)	133 (10.0%	25 (1.8%)
PAD Sept 2017	PAD	1338	729	52.1 [42.8, 61.6]	0.1 [0.0, 0.3]	13.8 [9.0, 19.5]	38.2 [31.4, 45.3]	60.5%	56.2%	187 (14.0%)	112 (8.4%)	0 (0.0%)	18 (1.3%)	94 (7.0%)	26 (1.9%)
PAD Aug 2018	PAD	1338	366	61.4 [51.8, 70.8]	1.1 [0.3, 1.9]	39.3 [31.1, 47.6]	21.1 [15.8, 27.7]	68.4%	62.0%	143 (10.7%)	153 (11.4%)	1 (0.1%)	64 (4.8%)	88 (6.6%)	34 (2.5%)
PAD Sept 2019	PAD	1336	437	56.6 [49.2, 65.2]	0.3 [0.0, 0.6]	33.3 [26.9, 40.2]	22.9 [16.8, 31.1]	57.1%	57.7%	143 (10.7%)	102 (7.6%)	0 (0.0%)	42 (3.1%)	60 (4.5%)	31 (2.3%)
PAD		1338	470	58.9 [50.1, 68.2]	0.6 [0.1, 1.0]	30.4 [23.8, 37.5]	27.9 [21.3, 35.7]	62.4%	58.6%	164 (12.3%)	144 (10.7%)	1 (0.1%)	49 (3.7%)	94 (7.0%)	29 (2.1%)
YF June 2017	YF	2739	2687	24.9 [22.8, 27.2]	1.2 [0.2, 2.5]	4.0 [3.4, 4.8]	19.7 [18.0, 21.6]	31.8%	36.8%	180 (6.6%)	77 (2.8%)	4 (0.1%)	14 (0.5%)	58 (2.1%)	63 (2.3%)
YF Sept 2017	YF	2739	2857	22.6 [20.7, 24.7]	1.3 [0.3, 2.6]	5.5 [4.3, 6.8]	15.8 [14.6, 17.3]	27.0%	33.5%	180 (6.6%)	64 (2.3%)	4 (0.1%)	15 (0.6%)	45 (1.6%)	74 (2.7%)
YF Aug 2018	YF	2739	1784	22.4 [19.7, 25.3]	1.8 [0.3, 3.8]	4.6 [3.6, 6.0]	16.0 [14.3, 17.9]	17.0%	28.2%	138 (5.0%)	42 (1.5%)	3 (0.1%)	10 (0.4%)	30 (1.1%)	25 (0.9%)
YF Sept 2019	YF	2739	1533	18.5 [16.1, 21.2]	1.9 [0.4, 4.0]	2.3 [1.8, 3.0]	14.3 [12.7, 16.1]	15.6%	25.5%	170 (6.2%)	47 (1.7%)	3 (0.1%)	9 (0.3%)	35 (1.3%)	28 (1.0%)
YF		2739	2215	22.1 [19.8, 24.6]	1.5 [0.3, 3.2]	4.1 [3.3, 5.2]	16.5 [14.9, 18.2]	22.8%	31.0%	167 (6.1%)	57 (2.1%)	3 (0.1%)	12 (0.4%)	42 (1.5%)	47 (1.7%)
Mean				22.3 [18.9, 26.4]	0.5 [0.1, 1.1]	9.0 [7.1, 11.3]	12.7 [10.1, 16.1]	25.3%	28.6%	350 (17.2%)	63 (4.0%)	1 (0.1%)	17 (1.2%)	44 (2.8%)	24 (1.4%)
Weighted mean				16.2 [13.9, 19.1]	0.5 [0.1, 1.1]	5.8 [4.5, 7.2]	10.0 [8.2, 12.2]	17.9%	21.8%	409 (16.9%)	53 (3.0%)	1 (0.1%)	13 (0.8%)	38 (2.1%)	24 (1.2%)
Mean (late summer)				21.4 [18.0, 25.6]	0.5 [0.1, 1.1]	8.6 [6.7, 10.9]	12.3 [9.8, 15.5]	24.4%	28.4%	343 (16.9%)	55 (3.6%)	1 (0.0%)	15 (1.0%)	39 (2.5%)	23 (1.3%)
Weighted mean (lt. s.)				15.0 [12.7, 17.8]	0.5 [0.1, 1.1]	5.3 [4.1, 6.7]	9.2 [7.5, 11.3]	16.4%	21.1%	401 (16.6%)	47 (2.7%)	1 (0.0%)	12 (0.7%)	34 (1.9%)	22 (1.1%)

Table 2. Within-lake emergent vegetation coverages (A_{LEV}) by vegetation type (A_{WF} = area of wet forest, A_{WS} = area of wet shrub, A_{WG} = area of wet graminoid, A_{WEV} = area of wetland emergent vegetation, as opposed to lake vegetation) and by study area, along with landscape coverage in km^2 and as percent coverages. Numbers in brackets give the bootstrapped 95% confidence intervals. Weighted mean columns are weighted by individual lake area, and summary weighted mean rows are weighted by the total lake area of each study area for all dates and late summer only (August and September, abbreviated as lt. s. when necessary).

3.1.2 Validation of UAVSAR classifier

The land cover classifier successfully retrieves the three broad classes of emergent vegetation. Based on visual inspection of the land cover maps, the most significant misclassification is evidenced by false detections of water in areas actually covered by dry graminoid vegetation (**Figure 3e**, top middle) and false detections of inundated vegetation in areas of forest. The most frequent misclassification occurs between Wet Shrub and Rough Water, although errors of omission and commission are roughly equal, implying a near-zero net effect on the landscape totals (**Figure S.1**). Any misclassification among the dry land classes does not affect our lake analysis, and misclassification between the flooded and dry classes is rare, as expected, given the sensitivity of SAR to water presence (**Figure S.1**). Prior to the quality control measures (**Section 2.3.4**), Cohen's kappa coefficients are 0.862 for the model used on the simpler CSD landscape and 0.824 for the model used for the remaining sites, implying good agreement with the validation data. Since the analysis only uses flooded classes connected to open water that could be validated by optical imagery, errors of commission (**Figure S.1**) represent an upper bound.

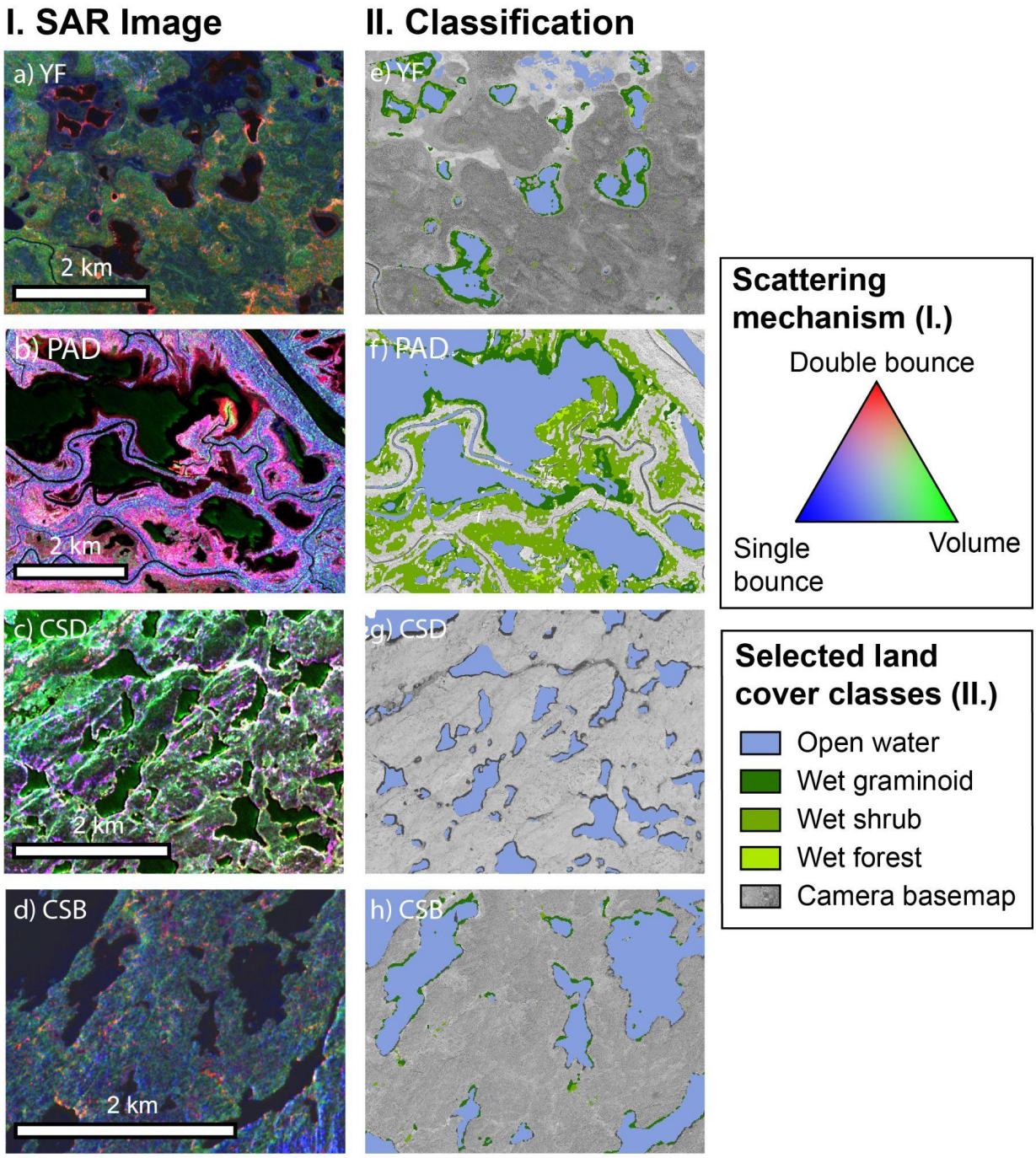

Figure 3. Example L-band SAR images of subsets within the four study areas (**Column I. a-d**, YF 6/2017, PAD 9/2019, CSD 9/2017, CSB 8/2018, respectively) and corresponding classification (**Column II. e-h**). SAR images are colorized by Freeman-Durden scattering mechanism (double bounce in red, primarily indicating emergent vegetation; volume scattering in green, primarily indicating leafy vegetation; and single bounce scattering in blue, primarily indicating bare ground, bedrock, and some types of trees) and are stretched identically, with visual adjustments for brightness and color saturation. In column **II**., only inundated classes are

shown and are superimposed over a grayscale version of the color-infrared camera base map from Kyzivat et al. (2018), in which forests appear darker than grasslands or bedrock.

3.2 Emergent vegetation extent

3.2.1 Regional and morphological trends

Although useful for integrating all flux components, landscape-scale descriptors obscure the nuance of individual lake characteristics. Consequently, we also present results normalized by each lake's area and aggregated via weighted averaging (**Table 2, Figure 4**). With this normalization, it is more apparent that emergent vegetation (A_{LEV}) is quite prevalent in lakes, averaging 16.2 [13.9 – 19.1]% across the four study areas, weighted by lake area. Again, coverage is especially extensive in the lowland PAD and YF (**Figure 2**), averaging 59 [50 – 68]% and 22 [20 – 25]%, respectively. A_{LEV} in the more topographically constrained, colder, sparsely vegetated CSB and CSD areas averages 7.0 [4.7 – 11.5]% and 1.0 [0.8 – 1.2]%, respectively. The lowland sites, therefore, have the most A_{LEV}, both as a percentage of total lake area as well as landscape area.

While emergent vegetation is observed in every size bin in every area, we find only a weak relationship between A_{LEV} and lake area that holds for all study areas. The area bins comprising small to medium-sized lakes between 0.002 to 0.02 km^2 always contain the primary histogram peak, with the exception of the PAD, for which these bins contain the secondary peak (**Figure 4b**). In all regions except the PAD, the smallest observable lakes ($\geq$ 250 m^2) have similar coverage to the largest (> 10 km^2), resulting in unimodal area-binned histograms, even within the confidence intervals (**Figure 4**). The drop in A_{LEV} for small lakes is likely caused by mixed pixels in narrow littoral zones being detected as water. Even so, Pearson correlation is weak between log-transformed A_{LEV} and lake area ($r^2 = 0.124$, $p < 0.001$, **Figure 5**), implying that the inverse relationship between the two variables is not consistent across sites. On an individual basis, the two Canadian Shield study areas have significant regression relationships ($p < 0.001$, **Figure 5**), with $r^2 = 0.25$ (CSB) and 0.48 (CSD), likely explained by their simpler, bedrock-dominated landscapes.

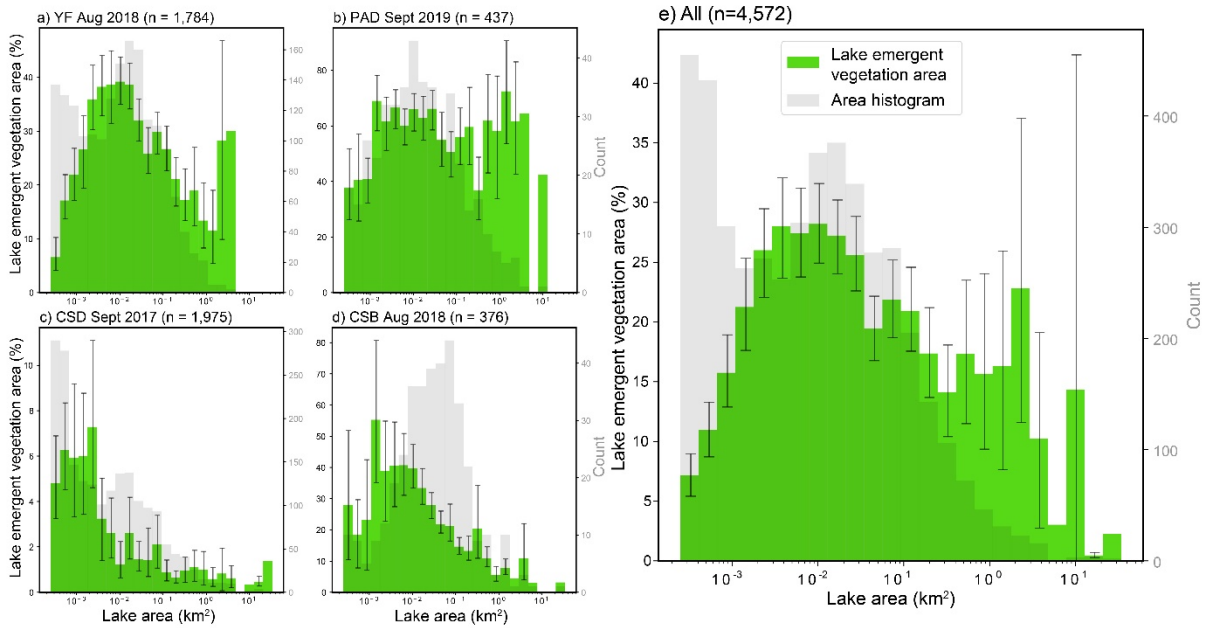

Figure 4. Emergent vegetation (A_{LEV}) is most prevalent in small to medium-sized lakes. Here, mean A_{LEV}, in green, is calculated for logarithmic lake area bins for each region (a) and for all regions combined (b). Error bars give the 95% confidence interval for A_{LEV} for all bins with > 2 observations. The lake count in each bin is plotted in grey and shows that most observed lakes are much smaller than 1 km^2. Accordingly, bins with fewer lakes generally have greater uncertainty in A_{LEV}, and the rightmost bins, which contain < 10 lakes, have considerable uncertainty. For a version of this figure showing bin sums, rather than means, see **Figure S.2**.

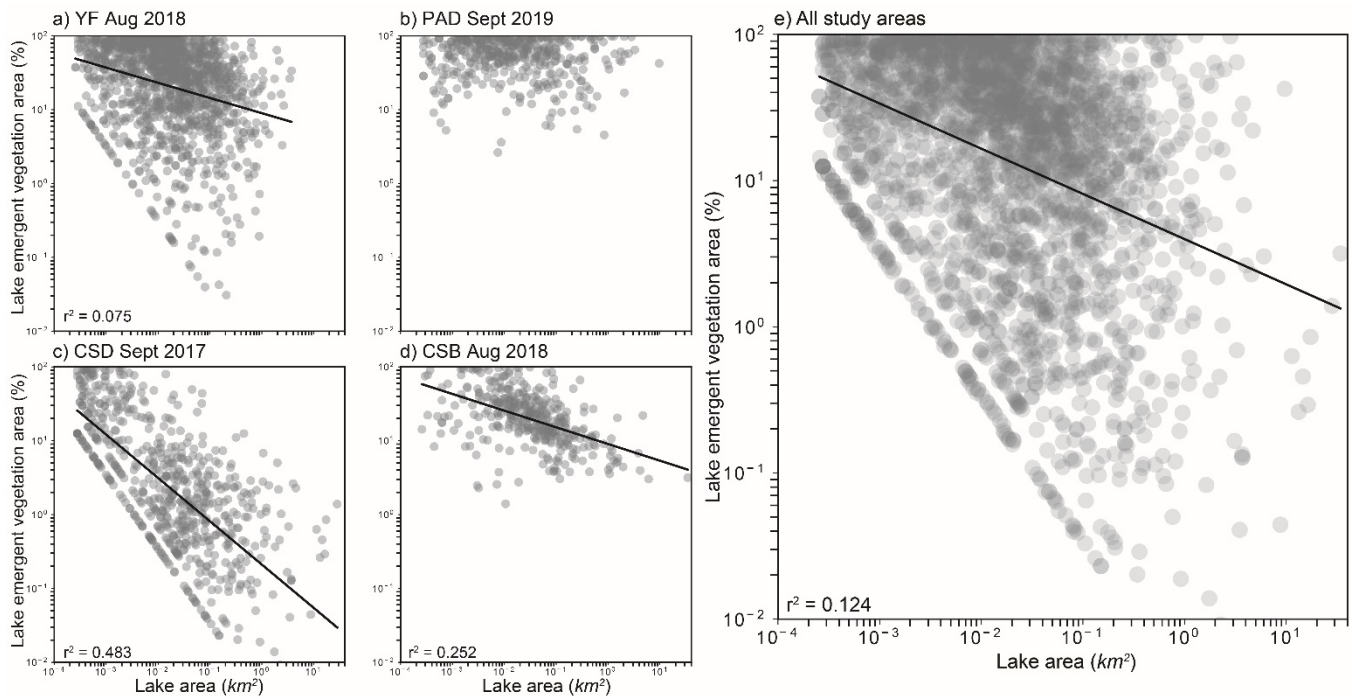

Figure 5. Scatter plot of lake area and emergent vegetation coverage (A_{LEV}) for all 4,572 lakes by study area (**a-d**) and aggregated (**e**). There is only a weak relationship between the two log-transformed variables. The diagonal bottom-left boundary in most plots is caused by area quantization by pixilation; since A_{LEV} is a fraction, the minimum possible A_{LEV} corresponding to a one-pixel vegetated zone decreases as the denominator increases. Lakes with $A_{LEV} = 0$ are not shown nor included in the regression and regression lines are only included for $p < 0.001$.

3.2.2 Seasonal trends

Despite fluctuating water levels, the distribution of A_{LEV} across lakes of varying areas remains largely similar across seasons and years (**Figure S.3**). In all study areas, there is a histogram peak at lakes with little or no emergent vegetation (**Figure S.3 a-d**, leftmost bin), as many areas lack the necessary conditions to support emergent macrophytes. The histogram drops sharply with increasing A_{LEV} coverage: extremely quickly in the sparsely-vegetated CSD, somewhat quickly in the more southern CSB, and gradually in YF. The negative-skewed PAD distribution (tail on left) is an anomaly with high-coverage lakes common. Accordingly, the area-weighted mean (58.9 %) is barely greater than the arithmetic mean coverage (58.6 %) in the PAD, unlike the rest of the study areas and the aggregated total, for which these values can differ by a factor of two (**Table 2**). There are also more lakes overall detected in the PAD during early summer (**Figure S.3**), likely because temporarily submerged macrophytes would be detected as open water and thus constitute lakes in our analysis. These effects are likely due to prevalence of shallow open water wetlands, which are ubiquitous in the delta and are included in our lake dataset as long as some area of open water (> one pixel, or ~30 m^2) is detected. Although there is little seasonal variance to the A_{LEV} distribution, the corresponding methane fluxes may depend greatly on plant activity, which varies between seasons. To avoid including seasonal wetlands as lakes, we used only the late summer (low water season) land cover maps to calculate mean A_{LEV} and have broken down available flux data by season. The temporal invariance of the A_{LEV}

histograms provides further validation of the consistency of the classifier, and it shows how changes in A_{LEV} are not relegated to the same small subset of lakes.

3.3 Methane fluxes from emergent macrophytes vs. open water

Field measurements confirm that methane fluxes per unit area from emergent macrophytes are consistently higher than open water, even within the same lake (**Figure 6**). Although macrophyte fluxes were only collected at five of the 15 visited PAD lakes, four have higher mean macrophyte values than open water, leading to a mean macrophyte: open water flux ratio of 2.3 (Kyzivat et al., 2021b). Given the small sample size, differences are not significant ($u = 2.0$, $p = 0.19$, $n = 5$) based on the non-parametric Mann-Whitney test. Strong variability in the measurement may also contribute, since these short-term measurements exclude ebullition and the other key episodic open water fluxes (ice-out flux, water column turnover fluxes) are accounted for afterwards via a correction factor. Similarly, plants, as well as open water, can have pronounced diel and seasonal variability in their fluxes, and these measurements were all made during the day.

The fluxes obtained by literature synthesis (**Table S.3**) have an even more extreme median ratio of 8.8 (**Figure 6**; **Figure 7**, top histogram), with a significant difference between open water and vegetation ($u = 1{,}800$, $p < 0.001$, $n = 47$). Of the 56 paired vegetation versus open water measurements, all but eight have flux ratios > 1, implying greater emissions from vegetated regions. The PAD and literature measurements combined have a median flux ratio of 6.1, or 15.9 if only Arctic-boreal lakes are included. We use the former, smaller value, since it comes from a larger sample size, and multiply it by the ice-melt flux correction factor to obtain 4.6, which is used for the subsequent sensitivity calculation (**Table 3**). Due to limited data, studies from all seasons and measurement periods were used, and some only measured one or two of the emission pathways (see **2.3.6**). The four studies that defined lake zones based on depth rather than vegetation yielded a median flux ratio of 15.8. Despite a limited and spatiotemporally uneven global sampling, lakes in our study areas and worldwide unequivocally trend towards higher emissions from emergent macrophyte environments than from open water.

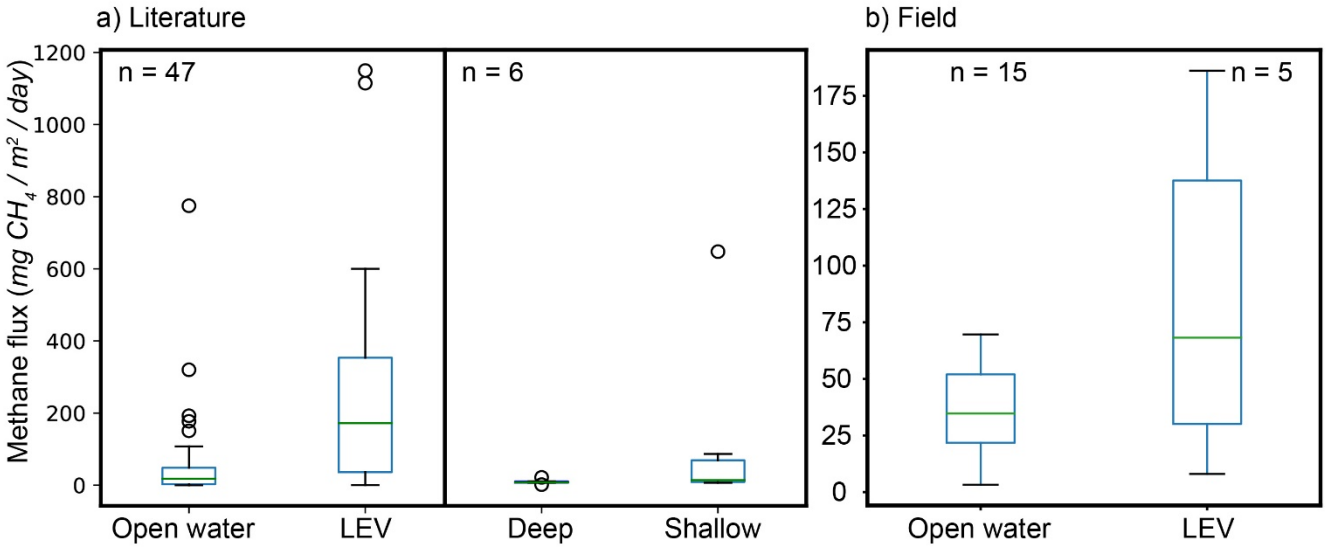

Figure 6. Lake emergent vegetation (LEV) and shallow regions produce greater methane fluxes than open water zones and deep regions, respectively, based on the literature (**a**) and from field measurements in the Peace-Athabasca Delta in July and August 2019 (**b**). Green lines show the median, hinges are drawn at the lower and upper quartiles, and flyer bars give the extent of data not considered outliers, which are plotted as points. Note the different scales demonstrating much greater flux values (mg of CH_4 /day) from the literature (**a**) than in the PAD (**b**).

3.4 Sensitivity of whole-lake methane emissions to inclusion of vegetated areas

By applying the median corrected macrophyte: open water ratio of 4.6 (**Section 3.3**) to our remotely sensed UAVSAR LEV maps (**Figure 3**), we estimate the relative importance of accounting for emergent vegetation in whole-lake methane flux estimates (**Table 3**). Assuming a lake area weighted average A_{LEV} of 16.2 [13.9 - 19.1]% increases the overall methane emissions from the four study areas by 21 [18 - 25]% (**Figure 7**). Although the flux ratio *R'* has variability, we have not included it within the bounds of the estimate, relying instead on the more robust measurement of variance of A_{LEV}. Spatiotemporally, the impact ratio *I* varies from 4% to 321%, with the lower bound coming from CSD in September 2017 (where only ~0.9% of lake areas contains emergent vegetation) and the upper bound from the PAD in June 2017 (~66% coverage, **Table 2**). Although these are the most extreme values observed, these scenarios show that accounting for small, but numerous LEV zone areas significantly raises whole-lake emissions estimates.

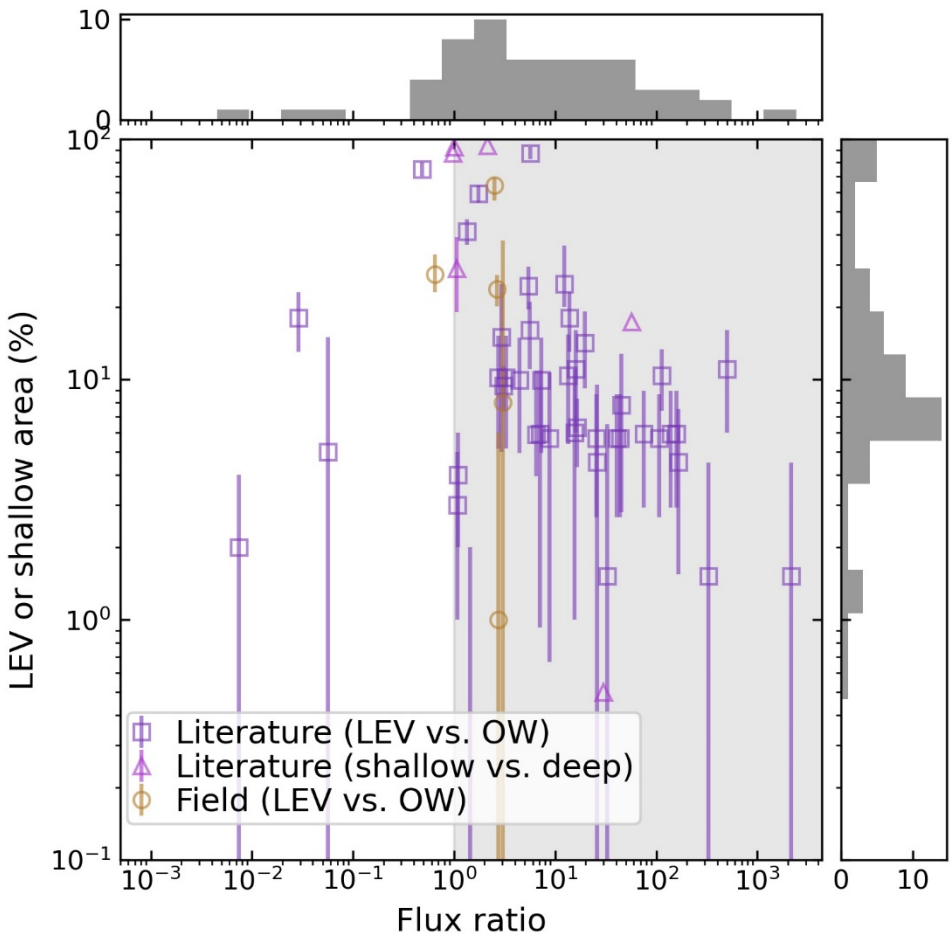

Figure 7. Plotting study lakes in a flux ratio-emergent vegetation fraction feature space shows that most emit more methane from lake emergent vegetation (LEV) than from open water on a per-area basis (shaded region), leading to an overall median flux ratio of 6.1. Studies that partitioned fluxes into shallow versus deep, rather than vegetated vs. open water zones (triangular markers) are shown for reference but are not used for further analysis. The distributions of both variables are shown as histograms along the relevant axes. Vertical error bars show the temporal range in coverage for the field data (orange circles) and the estimated mapping uncertainty for the literature data (purple squares) and can extend to zero (beyond axis limits). For scale, the uppermost square data point in the figure (peat pond, Ontario, Canada, A_{LEV} = 88%, R=5.7) corresponds to a 113% increase in emissions compared to the no LEV zone

case. Note the logarithmically-scaled x and y axes. For a version of this figure with contour lines for the impact *I*, see **Figure S.4.4**.

A_{LEV}	*c*	*R*	R'	*I*
16.2 [13.9 – 19.1]%	0.36	6.1	4.6	21 [18-25]%

Table 3. Parameters and results of sensitivity calculation (**Equations 2** and **3**). A_{LEV} is area, with 95% confidence intervals, of lake emergent vegetation and is corrected for double-counting with wetlands by the scalar *c*. *R* is the median global vegetated: open water flux ratio obtained from the literature and is adjusted to *R'* correct for unmeasured ice-melt fluxes. The summary statistic *I* represent the impacts of accounting for LEV in whole-lake methane flux estimates.

4 Discussion and Conclusion

4.1 Emergent vegetation coverage in lakes

Littoral zones are often theorized to cover greater portions of small lakes than of large lakes (Bergström et al., 2007; Wetzel, 1990, 2001). It is logical that smaller lakes with larger perimeter: area ratios would be dominated by near-shore areas, which are overwhelmingly shallow. However, while our results generally show greater fractional emergent vegetation area (A_{LEV}) in small and medium-sized lakes (**Figure 4**), there is weak correlation at best (Pearson r^2 = 0.124, $p < 0.001$; **Figure 5**). This discrepancy can likely be explained by lake emergent vegetation (LEV) comprising only a portion of the littoral zone, as well as mixed pixels obscuring narrow littoral margins in small lakes. Bergström et al. (2007) similarly observed that medium-sized lakes (0.1 to 1 km^2) had the greatest A_{LEV} of ~11% on average for 50 Fennoscandian Shield lakes in Finland, which, plotted as an area-binned histogram, also resembles an inverted V-shaped curve. Mäkelä et al (2004), using the same dataset, pointed out that large, lowland lakes had the largest total macrophyte coverage, also noting that area and pH only account for 15% variation in A_{LEV}.

In comparison, the Canadian Shield areas we sampled contained the greatest A_{LEV} in small-to-medium lakes (0.0001 - 0.002 km^2 in area), with values ranging from 7.3 [4.5 – 10.7] % (CSD) to 55 [35 – 81] % (CSB). We also observe a large contribution to total A_{LEV} from the large lakes (**Figure S.2**), underscoring the need not to discount them. Incidentally, these lakes are under-represented in lake methane datasets (Deemer & Holgerson, 2021). The largest 100 lakes (area ≥ 0.9 km^2) comprise 62.7% of total lake area and 39.2% of total LEV area across all four study areas, and this trend holds across all study areas (**Fig S.2**). The observed region-specific dependence on lake area further highlights the need for remote sensing to estimate littoral or vegetated zone coverage as well as to identify the interface between wetlands and open waters in the context of aggregated methane emission estimates.

The ~16% mean A_{LEV} coverage we observe is greater than the globally-inclusive estimate of 7% (Duarte et al.,1986) and Southern Finland estimate of 5.2% (Bergström et al., 2007). Since the number is an intermediate average derived from much lower values on the Canadian Shield (1.0%, and 7.0% for CSD and CSB, respectively, **Table 2**) and much higher values for the PAD

(59%) and YF (22%), it is highly sensitive to the choice of study areas and their relative sizes. Even though the relationship between coverage and lake area does not appear as simple as suggested by Duarte et al. (1986), their conclusion that lake area is not a strong predictor of emergent macrophyte coverage is still supported. Although the Boreal–Arctic Wetland and Lake Dataset (BAWLD; Olefeldt et al., 2021a; Olefeldt et al., 2021b) does not explicitly map littoral vegetation, the authors defined all open-water ecosystems as lakes, which includes shallow open-water wetlands. As a result, their lake class is defined nearly identically to ours, and they cite similar reasons regarding the importance emergent macrophytes as controls on net emissions. Indeed, comparison between datasets shows similar (ranging from 3-46% difference) lake coverage in each study area and an identical area-weighted mean over all study areas (16.6%, **Table S.6**). The roughly equivalent emergent vegetation and/or wetland classes are 24% greater in BAWLD (3.8% areal coverage from UAVSAR, 4.7% from BAWLD), which indicates that some or all LEV is included within BAWLD wetlands. BAWLD therefore represents best practices not only in ensuring a consistent lake-wetland distinction, but also presumably in including lake emergent vegetation within a wetland class, where it can be assigned a more appropriate methane flux.

4.2 Importance of emergent vegetation for methane upscaling

4.2.1 Toward improved upscaling of lake methane emissions

This broad-domain study supports previous studies demonstrating the importance of accounting for vegetated and/or littoral areas in upscaling lake methane flux estimates (Bergström et al., 2007; Casas-Ruiz et al., 2021; DelSontro et al., 2018a; Juutinen et al., 2003; Kankaala et al., 2013; Natchimuthu et al., 2016; Smith & Lewis, 1992; Striegl & Michmerhuizen, 1998). However, in addition to the challenges of measuring wetland extent more generally (Melton et al., 2013), a knowledge gap remains about the distribution and area of lake littoral zones (Huttunen et al., 2003). Our airborne UAVSAR approach for detecting LEV has limited spatial coverage and is unsuitable for broader-scale studies. Satellite approaches, however, have good utility for pan-Arctic or global wetland mapping (Hess et al. 1990, Nelson et al. 2006, Ghirardi et al. 2019, Zhang et al. 2021) and are well suited for study of large lakes, which contribute most to total LEV area (**Fig S.2**). These lakes are otherwise considered low methane emitters on a per-area basis (Holgerson & Raymond, 2016) and have little risk of being double-counted in wetland datasets, so they would be a good starting point for future studies. Incidentally, DelSontro et al. [2018] define an underestimation ratio between pelagic and littoral methane concentrations (roughly the inverse of I) and show that it approaches unity for larger lakes, although they do not calculate the impact of these lakes to total lake emissions. The upcoming NISAR satellite mission is likely to provide high-resolution, freely available global coverage of L-band SAR, which may facilitate similar analysis for A_{LEV} over larger scales.

Unfortunately, our results do not reconcile the gap between modeled methane fluxes from bottom-up and top-down models (Thornton et al. 2016; Saunois et al., 2020). In fact, they suggest bottom-up fluxes are slightly greater than previously thought, which further widens the discrepancy. The most recent aquatic upscaling studies (Saunois et al., 2020; Rosentreter et al., 2021) and a recent wetland synthesis dataset for modeling (Zhang et al., 2021) used a consistent lake mask when defining lake and wetland areas, and this careful lake masking has not significantly improved the discrepancy (Saunois et al., 2020). These masks either come from global lake datasets (HydroLakes, GLWD, GLOWABO), or the more recent global surface water

explorer (GSW). Both GLOWABO and GSW were derived from 30 m resolution, optical Landsat satellite data, which is quite effective at detecting open water. It is unclear whether these methodologies include vegetation as part of lakes, although GLOWABO and HydroLakes show good agreement with our open water class (**Table S.4**). Wetland detection is more challenging and hampered by scale disparities between the relevant satellite sensors and inconsistent wetland definitions between disciplines ([Poulter et al., 2017]; Zhang et al., 2021). Thus, the practice of using consistent open-water lake masks to differentiate between lakes and wetlands is a good one.

Our results show that even after correcting for double-counted wetlands, UAVSAR detects emergent vegetation in 5.8% of lakes contained in global datasets. Whether through temporal change or dissimilar mapping methods, this discrepancy is large enough to have an impact on estimates of the lake contribution to the global methane budget. Equally important, but not demonstrated here, is accounting for the uniquely high emissions from non-vegetated lake littoral zones, which are less likely to be confused with wetlands, and are probably at least as extensive as LEV (Seekell et al., 2021). Non-vegetated littoral zones can also be high emitters, especially when within the reach of carbon-exuding roots and rhizomes (Bansal et al. 2020). Since mapped LEV falls within littoral zones by definition, it shares some of their properties, but our analysis does not attempt to separate these drivers. Even so, Jansen et al. (2020a) found no clear depth difference in the diffusive fluxes from two lakes in Stordalen Mire, Sweden, despite maximum depths of 5 and 7 m and a robust sampling strategy. However, ebullitive emissions from these same lakes showed a clear depth gradient (Wik et al. 2013). Our compiled synthesis data on depth, while limited, also shows no significant difference between deep and shallow emissions (Mann Whitney test, $u = 10$, $p = 0.24$, $n = 6$), highlighting the need for more reporting of fluxes from different pathways and depth zones.

Given that our LEV flux data includes all emission pathways in a variety of lake types, the derived flux ratios represent a combination of many correlated drivers, including shallowness, methane oxidation, variable inundation, proximity to terrestrial inputs, and microbial community. In the context of deriving spatially explicit representations of methane emissions, it could be preferable to move away from using discrete land cover classes, and develop continuous representations of the processes that control methane production and rates of flux. These representations could better describe gradually-varying conditions, such as water table depth, the resulting concentration of oxygen in the subsurface, and the inclusion of new estimates of soil moisture, and they could improve estimates of methane emissions along hydrologic gradients.

4.2.2 Limitations and future directions

Our 21% estimate for I (**Equation 2**), the percent increase due to including emergent vegetation in lake methane flux accounting, uses assumptions chosen to lead to the smallest possible value. Regardless, it is highly sensitive to the data-limited input parameters c and the flux ratio, which has a large variability that we have not accounted for. The double-counting correction factor c may suffer from lack of generality, since it was calculated only within the boundaries of our study regions using global datasets collected ~ 20 years prior. It also assumes that LEV zones have similar areal emission to wetlands, which may not be valid. Clearly, more methane flux measurements in shallow or vegetated zones and estimates of total macrophyte

coverage are needed (Bergström et al., 2007; Schmiedeskamp et al., 2021). While our approach for correcting for double counting is only based on lakes large enough to be included in global datasets, the small magnitude of *c* shows how easy it is to count wetlands as LEV. Without this correction factor, *I* would be more than doubled to 58%. Future work should look more generally at the cause and magnitude of lake/wetland double counting (Thornton et al., 2016) and develop continuous metrics for methane emission habitats that don't rely on discrete land cover classes.

Our estimate for *I* may still be too high because our A_{LEV} includes up to 2.6% emergent shrubs and trees, even after the double-counting correction (**Table S.5**). This woody vegetation lacks the aerenchyma tissue that allows most wetland plants to transport methane from the sediments. Recent work has shown the potential for microbes living inside trees to produce methane (Covey & Megonigal, 2019), although this effect is likely less than soil microbe production. Secondly, the relatively narrow swath width of UAVSAR causes large (and likely less-vegetated) lakes to be under-represented in the calculation of weighted mean A_{LEV}. Adding to this effect is the use of the same vegetated: open water flux ratio for lakes of all sizes, when smaller lakes and ponds are known to be higher open-water methane emitters than large (Michmerhuizen, Striegl, & McDonald, 1996; Bastviken et al., 2004; Holgerson & Raymond, 2016; Engram et al. 2020), probably because littoral zones (vegetated and unvegetated) cover most of their areas. Indeed, Kankaala et al. (2013) showed that the flux ratio increases with lake size. It follows that our concept of a vegetated: open water flux ratio is less useful for small lakes and would likely be even larger for the largest lakes, which were under-represented in our literature synthesis. Future studies could better quantify how this ratio varies based on lake area. Nevertheless, since the contribution to total A_{LEV} from the small lakes is so slight (**Fig S.2**), they don't have a large negative impact on our estimate. Finally, the estimate compares to a hypothetical upscaling using solely open water fluxes, while in reality, some studies include open-water measurements from entire littoral zones. While many of the studies cited here used area-weighted approaches with regard to lake depth zones (Natchimuthu et al., 2016; DelSontro et al, 2016; Jansen et al. 2020a), they appear to be a minority and are not available on the global scale (Kuhn et al. 2021b; Wik et al. 2016b).

Comparison of our sensitivity study with previous Arctic-boreal and global lake studies suggests that our finding of a 21% increase in whole-lake methane flux is conservative. Using flux chamber measurements from two Swedish lakes, Natchimuthu et al. (2016) found that methane emissions from lake centers are 2.1 times less than whole-lake fluxes, although fluxes were not explicitly measured near lake macrophytes. Similarly, Kankaala et al. (2013) found that 74-82% of diffusive and plant-mediated emissions in 12 Finnish lakes derived from littoral macrophyte stands comprising only 5% of their total area. These amounts correspond to a flux ratio of 54-86, leading to an impact, *I*, on whole-lake fluxes between 270 and 430% greater than a case where open water fluxes were assumed throughout. Most recently, Desrosiers et al. (2022) found that the 26% of a boreal lake covered in macrophytes was responsible for 81% of its carbon emissions. The impact of considering the *Typha latifolia* stands alone can be calculated at 102%. Although focused only on extremely high-emitting lake and wetland thermokarst hot spots, Elder et al. (2021) conducted a study of remotely-sensed methane "hot spot" emissions across a 70,000 km^2 Arctic-boreal domain and found an even greater disproportionality, where 0.005% of the domain was estimated to emit 0.3-16.2 % of the total. The higher reported flux ratios from lake studies can be partly attributed to area-weighted analyses including much larger, and thus lower-emitting per unit area, lakes than our airborne-based study. Yet, they also

underscore the pitfalls of assigning higher areal fluxes to vegetated lake zones without ensuring these zones are not otherwise counted as wetlands.

Even when using best practices to avoid double-counting lakes with wetlands, the coarse resolution of global lake maps can still cause uncertainty in the precise location of shorelines. At the medium resolution of Landsat (30 m), the entire littoral zone could be "hidden" inside of mixed pixels at lake boundaries, even for large lakes, if they have steep margins. If only unmixed pixels are classified as lakes, it is unclear how near-shore land pixels would be treated, especially given that global wetland maps are typically made from coarser-resolution sensors (Zhang et al., 2021). Unfortunately, this hard-to-resolve small strip of land/water interface is precisely the area with the greatest impact on full-lake (DelSontro et al. 2018, Thornton et al. 2016) and landscape (Elder et al. 2021) fluxes, so it cannot be rounded off. Furthermore, due to changing inundation and vegetation coverage, lakes can contain LEV even if attempts are made to exclude it, such as from static lake maps. Littoral zones often have fluctuating inundation, and there are valid reasons to count them as either lakes or wetlands, even though current upscaling efforts require making this distinction. Just as with wetlands, lakes can be defined differently across disciplines. Although plant-mediated emissions are often reported in studies focused on lakes, upscaling studies frequently exclude vegetated areas from their lake estimates (Bastviken et al., 2011; Wik et al., 2013; Olefeldt et al., 2021a; Rosentreter et al., 2021), a best practice to avoid double-counting. This exclusion requires careful treatment of the fluxes from which "lake" estimates should be derived. Future work should develop techniques that can more accurately measure littoral zone area (Seekell et al., 2021), produce consistent and methane-relevant lake versus wetland criteria from remote sensing (Olefeldt et al., 2021a), and make use of temporally-dynamic inundation maps (Pekel et al., 2016; Zhang et al., 2021) for both wetlands and lakes.

Finally, since ebullition is under-represented in the synthesis dataset and not present in the field dataset, there may be biases present due to its episodic temporal pattern. We would expect a positive bias to R, since there is evidence that both diffusion and porewater CH_4 concentrations are reduced when there is an available plant pathway (Bansal et al. 2020). If this trend holds for ebullition as well, then ebullition would be greater in non-vegetated zones. Even so, of the 10 flux ratios that include ebullition among the measured pathways and use zones based on vegetation presence/absence, the median ratio is 6.5 (**Table S.3**), which hardly differs from the full dataset median of 6.1 (unpaired Mann-Whitney $u = 1{,}600$; $p = 0.052$, $n = 13$ and 39). Similarly, the use of a correction factor to compensate for missing ice-out flux measurements may too presumptive. Jammet et al. (2015) measured large spring fluxes in a very shallow peatland lake (<2 m), which suggests that methane accumulates in the sediments as well as in the water column over winter, and both shallow and deep areas would contribute to the spring efflux of CH_4. Further research is necessary to investigate how the flux ratio might change based on seasonality and pathway. In the absence of robust flux ratio data collected separately for each pathway, we do not attempt to correct for under-reported ebullition measurements.

Estimating Arctic-boreal lake methane emissions is constrained by limited data and reliance on assumptions such as discrete land cover classes. As noted by Saunois et al. (2020), methane upscaling can be improved by considering spatiotemporal variability and increasing sampling efforts in lakes with diverse morphologies and environmental conditions. Previous estimates have calculated a high bias caused by most measurements being made during waking hours (Sieczko et al., 2020) or summertime sampling (Wik et al., 2016a; Denfeld et al., 2018; Jansen et al., 2020b); or from static inundation maps (Hondula et al., 2021). Others have shown

low biases from insufficient seasonal (Wik et al., 2016b), or spatial (Wik et al., 2016b; Natchimuthu et al., 2016; Desrosiers et al., 2021) sampling. This study also suggests a low bias from not separately accounting for LEV, on par with the contribution of under-sampled ice melt flux, which ranges from 23 to 27%. Even so, inadequate and geographically-uneven sampling of the world's > 117 million lakes (Verpoorter et al., 2014) is likely the greatest source of uncertainty in lake upscaling. In the absence of sufficient data, upscaling estimates should make use of available quantitative corrections and continue to find and remediate sampling biases.

4.3 Conclusion

Lake emergent vegetation (LEV) is ubiquitous in Northern lakes but limited data prohibit its inclusion in upscaling lake methane emissions. We provide a first assessment of its prevalence across 4,572 lakes in four Arctic-boreal regions using airborne UAVSAR mapping and find that they cover 16.2 [13.9 – 19.1]% of Arctic-boreal lakes on average, a higher amount than other estimates, but with strong differences between study areas. LEV is greatest in lowland riverine areas, where changing water levels cause seasonal variability. Consistent with previous studies, we find that it is more common in small than large lakes, but this relationship is weak and varies regionally. Accounting for LEV, using a synthesis of paired open water and LEV field measurements of methane flux, leads to an upscaling estimate 21 [18 - 25]% greater than an estimate that assigns the same open water flux to the entire lake. We conclude that multi-temporal remote sensing of littoral zones, based on vegetation or otherwise, and collection of flux data from all parts of a lake are necessary for accurate upscaling of lake methane emissions. Future studies should continue using consistent definitions to separate lakes and wetlands, incorporate temporal wetland and lake change into analyses, remain vigilant against double counting with wetlands, and use multiple lake zones or continuous metrics for upscaling.

Acknowledgments

We thank Robert and Barbara Grandjambe for guiding services; Queenie Gray and Wood Buffalo National Park for enabling our field visits; Alex Shapiro of Alaska Land Exploration LLC for helicopter piloting; the U.S. Fish and Wildlife Service, the Gwichyaa Zhee Gwich'in Tribal Government and Doyon Limited for access to Yukon Flats National Wildlife Refuge; Michael W. Denbina for assistance with UAVSAR data processing; and Kevin Timoney for assistance with botanical identifications. Field work was conducted on or near land belonging to the Athabasca-Chipewyan, Mikisew Cree, Métis, and Gwich'in indigenous peoples. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center via the ABoVE Science Cloud. UAV Orthomosaics were created using DroneDeploy software. Data processing was made possible in part by the free and open-source software, including python, numpy, pandas, as well as with PolSAR Pro v5.1, which is provided by the European Space Agency (ESA). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Thanks to Lawrence Vulis and two anonymous reviewers who have greatly improved the manuscript.

Funding

This work was funded by the following NASA programs; Future Investigators in NASA Earth and Space Science and Technology (FINESST), grant number 80NSSC19K1361, managed by Dr. Allison Leidner; Terrestrial Ecology Program Arctic-Boreal Vulnerability Experiment (ABoVE), grant number 80NSSC19M0104, managed by Dr. Hank Margolis; and the NASA Surface Water and Ocean Topography mission, grant number NNX16AH83G, managed by Dr. Nadya Vinogradova Shiffer. US Geological Survey (USGS) LandCarbon Program provided funding for KPW, RGS, MMD, and FGT.

Data and software availability

UAVSAR data used for this study can be downloaded at https://uavsar.jpl.nasa.gov/cgi-bin/data.pl. The derivative land cover maps and lake emergent vegetation shapefiles can be found at the accompanying data publication: https://doi.org/10.3334/ORNLDAAC/1883. Methane flux data from the PAD can be found at https://doi.org/10.6073/pasta/1e0cadadd8024c8fabc692ee21dc1f57. All MATLAB, Python and shell scripts used in data processing can be found at https://doi.org/10.5281/zenodo.5974901 and https://doi.org/10.5281/zenodo.5974915.

References

Ayoub, F., Jones, C. E., Lamb, M. P., Holt, B., Shaw, J. B., Mohrig, D., & Wagner, W. (2018). Inferring surface currents within submerged, vegetated deltaic islands and wetlands from multi-pass airborne SAR. *Remote Sensing of Environment*, *212*, 148–160. https://doi.org/10.1016/j.rse.2018.04.035

Bansal, S., Johnson, O. F., Meier, J., & Zhu, X. (2020). Vegetation Affects Timing and Location of Wetland Methane Emissions. *Journal of Geophysical Research: Biogeosciences*, *125*(9), e2020JG005777. https://doi.org/10.1029/2020JG005777

Bastviken, D., Cole, J., Pace, M., & Tranvik, L. (2004). Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Global Biogeochemical Cycles*, *18*(4), 1–12. https://doi.org/10.1029/2004GB002238

Bastviken, D., Tranvik, L. J., Downing, J. A., Crill, P. M., & Enrich-Prast, A. (2011). Freshwater Methane Emissions Offset the Continental Carbon Sink. *Science*, *331*(6013), 50–50. https://doi.org/10.1126/SCIENCE.1196808

Bergström, I., Mäkelä, S., Kankaala, P., & Kortelainen, P. (2007). Methane efflux from littoral vegetation stands of southern boreal lakes: An upscaled regional estimate. *Atmospheric Environment*, *41*, 339–351. https://doi.org/10.1016/J.ATMOSENV.2006.08.014

Bourgeau-Chavez, L. L., Endres, S., Powell, R., Battaglia, M. J., Benscoter, B., Turetsky, M., … Banda, E. (2017). Mapping boreal peatland ecosystem types from multitemporal radar and optical satellite imagery. *Canadian Journal of Forest Research*, *47*(4), 545–559. https://doi.org/10.1139/cjfr-2016-0192

Bourgeau-Chavez, L. L., Graham, J. A., Endres, S., French, N. H. F., Battaglia, M., Hansen, D.,

& Tanzer, D. (2019). ABoVE: Ecosystem Map, Great Slave Lake Area, Northwest Territories, Canada, 1997-2011 (Version 1). ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1695

Bridgham, S. D., Cadillo-Quiroz, H., Keller, J. K., & Zhuang, Q. (2013). Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. *Global Change Biology*, *19*(5), 1325–1346. https://doi.org/10.1111/gcb.12131

Brown, J., Ferrians, O., Heginbottom, J. A., and Melnikov E. (2002). Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2 [Data set]. NSIDC. https://doi.org/10.7265/SKBG-KF16

Burger, M., Berger, S., Spangenberg, I., & Blodau, C. (2016). Summer fluxes of methane and carbon dioxide from a pond and floating mat in a continental Canadian peatland. *Biogeosciences*, *13*(12), 3777–3791. https://doi.org/10.5194/bg-13-3777-2016

Casas-Ruiz, J. P., Jakobsson, J., & Giorgio, P. A. del. (2021). The role of lake morphometry in modulating surface water carbon concentrations in boreal lakes. *Environmental Research Letters*, *16*(7), 074037. https://doi.org/10.1088/1748-9326/AC0BE3

Casper, P., Maberly, S. C., Hall, G. H., & Finlay, B. J. (2000). Fluxes of methane and carbon dioxide from a small productive lake to the atmosphere. *Biogeochemistry*, *49*(1), 1–19. https://doi.org/10.1023/A:1006269900174

Cheng, X., Huang, W., & Gong, J. (2014). Improved van Zyl polarimetric decomposition lessening the overestimation of volume scattering power. Remote Sensing, 6(7), 6365–6385. https://doi.org/10.3390/rs6076365

Christensen, T. R., Panikov, N., Mastepanov, M., Joabsson, A., Stewart, A., Öquist, M., … Svensson, B. (2003). Biotic controls on CO2 and CH4 exchange in wetlands – a closed environment study. *Biogeochemistry*, *64*(3), 337–354. https://doi.org/10.1023/A:1024913730848

Colmer, T. D. (2003). Long-distance transport of gases in plants: a perspective on internal aeration and radial oxygen loss from roots. Plant, Cell and Environment, 26(1), 17–36. https://doi.org/10.1046/j.1365-3040.2003.00846.x

Covey, K. R., & Megonigal, J. P. (2019). Methane production and emissions in trees and forests. *New Phytologist*, *222*(1), 35–51. https://doi.org/10.1111/NPH.15624

Dacey, J. W. H., & Klug, M. J. (1979). Methane efflux from lake sediments through water lilies. Science, 203(4386), 1253–1255. https://doi.org/10.1126/science.203.4386.1253

Deemer, B. R., & Holgerson, M. A. (2021). Drivers of Methane Flux Differ Between Lakes and Reservoirs, Complicating Global Upscaling Efforts. *Journal of Geophysical Research: Biogeosciences*, *126*(4), e2019JG005600. https://doi.org/10.1029/2019JG005600

DelSontro, T., Boutet, L., St-Pierre, A., del Giorgio, P. A., & Prairie, Y. T. (2016). Methane

ebullition and diffusion from northern ponds and lakes regulated by the interaction between temperature and system productivity. *Limnology and Oceanography*, *61*(S1), S62–S77. https://doi.org/10.1002/lno.10335

DelSontro, T., Beaulieu, J. J., & Downing, J. A. (2018a). Greenhouse gas emissions from lakes and impoundments: Upscaling in the face of global change. *Limnology and Oceanography Letters*, *3*(3), 64–75. https://doi.org/10.1002/lol2.10073

DelSontro, T., del Giorgio, P. A., & Prairie, Y. T. (2018b). No Longer a Paradox: The Interaction Between Physical Transport and Biological Processes Explains the Spatial Distribution of Surface Water Methane Within and Across Lakes. *Ecosystems*, *21*(6), 1073–1087. https://doi.org/10.1007/s10021-017-0205-1

Denfeld, B. A., Baulch, H. M., del Giorgio, P. A., Hampton, S. E., & Karlsson, J. (2018). A synthesis of carbon dioxide and methane dynamics during the ice-covered period of northern lakes. *Limnology and Oceanography Letters*, *3*(3), 117–131. https://doi.org/10.1002/LOL2.10079

Desrosiers, K., DelSontro, T., & del Giorgio, P. A. (2022). Disproportionate Contribution of Vegetated Habitats to the CH4 and CO2 Budgets of a Boreal Lake. *Ecosystems 2021*, 1–20. https://doi.org/10.1007/S10021-021-00730-9

Dove, A., Roulet, N., Crill, P., Chanton, J., & Bourbonniere, R. (1999). Methane dynamics of a northern boreal beaver pond. *Ecoscience*, *6*(4), 577–586.

Duarte, C. M., Kalff, J., Peters, R. H., & Peters, R. H. (1986). Patterns Biomass and Cover of Aquatic Macrophytes in Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *43*, 1900–1908.

Elder, C. D., Thompson, D. R., Thorpe, A. K., Chandanpurkar, H. A., Hanke, P. J., Hasson, N., James, S. R., Minsley, B. J., Pastick, N. J., Olefeldt, D., Anthony, K. M. W., & Miller, C. E. (2021). Characterizing Methane Emission Hotspots From Thawing Permafrost. *Global Biogeochemical Cycles*, *35*(12), e2020GB006922. https://doi.org/10.1029/2020GB006922

Engram, M., Walter Anthony, K. M., Sachs, T., Kohnert, K., Serafimovich, A., Grosse, G., & Meyer, F. J. (2020). Remote sensing northern lake methane ebullition. *Nature Climate Change, 10(6)*, 511–517. https://doi.org/10.1038/s41558-020-0762-8

Freeman, A., & Durden, S. L. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/36.673687

Ganju, N. K., Defne, Z., Kirwan, M. L., Fagherazzi, S., D'Alpaos, A., & Carniello, L. (2017). Spatially integrative metrics reveal hidden vulnerability of microtidal salt marshes. *Nature Communications 2017 8:1*, *8*(1), 1–7. https://doi.org/10.1038/ncomms14156

Ghirardi, N., Bolpagni, R., Bresciani, M., Valerio, G., Pilotti, M., & Giardino, C. (2019). Spatiotemporal dynamics of submerged aquatic vegetation in a deep lake from sentinel-2

data. *Water*, *11*(3), 1–14. https://doi.org/10.3390/w11030563

Hastie, A., Lauerwald, R., Weyhenmeyer, G., Sobek, S., Verpoorter, C., & Regnier, P. (2018). CO2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections. *Global Change Biology*, *24*(2), 711–728. https://doi.org/10.1111/GCB.13902

Hess, L. L., Melack, J. M., & Simonett, D. S. (1990). Radar detection of flooding beneath the canopy a review. *International Journal of Remote Sensing*, *11*(7), 1313–1325.

Holgerson, M. A., & Raymond, P. A. (2016). Large contribution to inland water CO2 and CH4 emissions from very small ponds. *Nature Geoscience*, *9*(3), 222–226. https://doi.org/10.1038/ngeo2654

Hondula, K. L., Jones, C. N., & Palmer, M. A. (2021). Effects of seasonal inundation on methane fluxes from forested freshwater wetlands. *Environmental Research Letters*, *16*(8), 084016. https://doi.org/10.1088/1748-9326/AC1193

Huttunen, J. T., Alm, J., Liikanen, A., Juutinen, S., Larmola, T., Hammar, T., … Martikainen, P. J. (2003). Fluxes of methane, carbon dioxide and nitrous oxide in boreal lakes and potential anthropogenic effects on the aquatic greenhouse gas emissions. *Chemosphere*, *52*, 609–621. https://doi.org/10.1016/S0045-6535(03)00243-1

Jammet, M., Crill, P., Dengel, S., & Friborg, T. (2015). Large methane emissions from a subarctic lake during spring thaw: Mechanisms and landscape significance. *Journal of Geophysical Research: Biogeosciences*, *120*(11), 2289–2305. https://doi.org/10.1002/2015JG003137

Jansen, J., Thornton, B. F., Cortés, A., Snöälv, J., Wik, M., MacIntyre, S., & Crill, P. M. (2020a). Drivers of diffusive CH4 emissions from shallow subarctic lakes on daily to multi-year timescales. *Biogeosciences*, *17*(7), 1911–1932. https://doi.org/10.5194/bg-17-1911-2020

Jansen, J., Thornton, B. F., Wik, M., MacIntyre, S., & Crill, P. M. (2020b). Temperature Proxies as a Solution to Biased Sampling of Lake Methane Emissions. *Geophysical Research Letters*, *47*(14). https://doi.org/10.1029/2020GL088647

Jensen, D., Cavanaugh, K. C., Simard, M., Christensen, A., Rovai, A., & Twilley, R. (2021). Aboveground biomass distributions and vegetation composition changes in Louisiana's Wax Lake Delta. *Estuarine, Coastal and Shelf Science*, *250*, 107139. https://doi.org/10.1016/j.ecss.2020.107139

Joabsson, A., Christensen, T. R., & Wallén, B. (1999). Vascular plant controls on methane emissions from northern peatforming wetlands. *Trends in Ecology & Evolution*, *14*(10), 385–388. https://doi.org/10.1016/S0169-5347(99)01649-3

Johnston, S. E., Striegl, R. G., Bogard, M. J., Dornblaser, M. M., Butman, D. E., Kellerman, A. M., … Spencer, R. G. M. (2020). Hydrologic connectivity determines dissolved organic matter biogeochemistry in northern high-latitude lakes. *Limnology and Oceanography Oceanography*. https://doi.org/10.1002/lno.11417

Jordahl, Kelsey, Joris Van den Bossche, Martin Fleischmann, James McBride, Jacob Wasserman, Adrian Garcia Badaracco, Jeffrey Gerard, Alan D. Snow, Jeff Tratner, Matthew Perry, Carson Farmer, Geir Arne Hjelle, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Brendan Ward, Giacomo Caria, Mike Taves, … Leah Wasser. (2021). geopandas/geopandas: v0.10.2 (v0.10.2). Zenodo. https://doi.org/10.5281/zenodo.5573592

Juutinen, S., Alm, J., Larmola, T., Huttunen, J. T., Morero, M., Martikainen, P. J., & Silvola, J. (2003). Major implication of the littoral zone for methane release from boreal lakes. *Global Biogeochemical Cycles*, *17*(4). https://doi.org/10.1029/2003gb002105

Kankaala, P., Huotari, J., Tulonen, T., & Ojala, A. (2013). Lake-size dependent physical forcing drives carbon dioxide and methane effluxes from lakes in a boreal landscape. *Limnology and Oceanography*, *58*(6), 1915–1930. https://doi.org/10.4319/lo.2013.58.6.1915

Kankaala, P., Kaki, T., Makela, S., Ojala, A., Pajunen, H., & Arvola, L. (2005). Methane efflux in relation to plant biomass and sediment characteristics in stands of three common emergent macrophytes in boreal mesoeutrophic lakes. *Global Change Biology*, *11*(1), 145–153. https://doi.org/10.1111/j.1365-2486.2004.00888.x

McKenzie Kuhn, Ruth Varner, David Bastviken, Patrick Crill, Sally MacIntyre, et al. 2021a. BAWLD-CH4: Methane Fluxes from Boreal and Arctic Ecosystems. Arctic Data Center. doi:10.18739/A2DN3ZX1R.

Kuhn, M. A., Varner, R. K., Bastviken, D., Crill, P., Macintyre, S., Turetsky, M., Anthony, K. W., Mcguire, A. D., & Olefeldt, D. (2021b). BAWLD-CH 4 : a comprehensive dataset of methane fluxes from boreal and arctic ecosystems. *Earth Syst. Sci. Data*, *13*, 5151–5189. https://doi.org/10.5194/essd-13-5151-2021

Kyzivat, E. D., Smith, L. C., Pitcher, L. H., Arvesen, J., Pavelsky, T. M., Cooley, S. W., & Topp, S. (2018). ABoVE: AirSWOT Color-Infrared Imagery Over Alaska and Canada, 2017. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1643

Kyzivat, E. D., Smith, L. C., Pitcher, L. H., Fayne, J. V., Cooley, S. W., Cooper, M. G., … Pavelsky, T. M. (2019). A high-resolution airborne color-infrared camera water mask for the NASA ABoVE campaign. *Remote Sensing*, *11*, 1–28. https://doi.org/10.3390/rs11182163

Kyzivat, E. D., Smith, L. C., Pitcher, L. H., Fayne, J. V., Cooley, S. W., Cooper, M. G., Topp, S., Langhorst, T., Harlan, M. E., Gleason, C. J., & Pavelsky, T. M. (2020). ABoVE: AirSWOT Water Masks from Color-Infrared Imagery over Alaska and Canada, 2017 (Version 1). ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1707

Kyzivat, E. D., Smith, L. C., Huang, C., Wang, C., Langhorst, T., Fayne, J. V., Harlan, M.E., Ishitsuka, Y., Feng, D., Dolan, W., Pitcher, L.H, Pavelsky, T. M. (2021a). ABoVE: Lake and Wetland Classification from L-band SAR, Alaska and Canada, 2017-2019. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1883

Kyzivat, E., F. Garcia Tigreros, T. Langhorst, J.V. Fayne, M.E. Harlan, Y. Ishitsuka, D. Feng, K.P. Wickland, M.M. Dornblaser, R.G. Striegl, D.E. Butman, and C.J. Gleason. 2021b. Methane and carbon dioxide fluxes from vegetated and open water zones of lakes in the Peace-Athabasca Delta, Alberta, Canada, 2019 ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/1e0cadadd8024c8fabc692ee21dc1f57 (Accessed 2022-03-04).

Laanbroek, H. J. (2009). Methane emission from natural wetlands: interplay between emergent macrophytes and soil microbial processes. A mini-review.

Langenegger, T., Vachon, D., Donis, D., & McGinnis, D. F. (2019). What the bubble knows: Lake methane dynamics revealed by sediment gas bubble composition. *Limnology and Oceanography*, *64*(4), 1526–1544. https://doi.org/10.1002/LNO.11133

Larmola, T., Alm, J., Juutinen, S., Huttunen, J. T., Martikainen, P. J., & Silvola, J. (2004). Contribution of vegetated littoral zone to winter fluxes of carbon dioxide and methane from boreal lakes. *J. Geophys. Res*, *109*. https://doi.org/10.1029/2004JD004875

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*, 1–22. https://doi.org/10.1016/j.jhydrol.2004.03.028

Li, M., Peng, C., Zhu, Q., Zhou, X., Yang, G., Song, X., & Zhang, K. (2020). The significant contribution of lake depth in regulating global lake diffusive methane emissions. *Water Research*, *172*, 115465. https://doi.org/10.1016/j.watres.2020.115465

Loken, L. C., Crawford, J. T., Schramm, P. J., Stadler, P., Desai, A. R., & Stanley, E. H. (2019). Large Spatial and Temporal Variability of Carbon Dioxide and Methane in a Eutrophic Lake. *Journal of Geophysical Research: Biogeosciences*, *124*(7), 2248–2266. https://doi.org/10.1029/2019JG005186

Mäkelä, S., Huitu, E., & Arvola, L. (2004). Spatial patterns in aquatic vegetation composition and environmental covariates along chains of lakes in the Kokemäenjoki watershed (S. Finland). *Aquatic Botany*, *80*(4), 253–269. https://doi.org/10.1016/j.aquabot.2004.08.006

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., … Kaplan, J. O. (2013). Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences*, *10*, 753–788. https://doi.org/10.5194/bg-10-753-2013

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*, 1–11. https://doi.org/10.1038/ncomms13603

Michmerhuizen, C. M., Striegl, R. G., & McDonald, M. E. (1996). Potential methane emission from north-temperate lakes following ice melt. *Limnology and Oceanography*, *41*(5), 985–991. https://doi.org/10.4319/lo.1996.41.5.0985

Miller, C., Griffith, C. P., Goetz, S. J., Hoy, E. E., Pinto, N., McCubbin, I. B., … Margolis, H. A.

(2019). An overview of ABoVE airborne campaign data acquisitions and science opportunities. *Environmental Research Letters*, *14*(8). https://doi.org/10.1088/1748-9326/ab0d44

NASA/JPL. (2017-2019). UAVSAR_POLSAR [Data set]. NASA Alaska Satellite Facility DAAC. https://doi.org/10.5067/7PEQV8SVR4DM

Natchimuthu, S., Sundgren, I., Gålfalk, M., Klemedtsson, L., Crill, P., Danielsson, Å., & Bastviken, D. (2016). Spatio-temporal variability of lake CH 4 fluxes and its influence on annual whole lake emission estimates. *Limnology and Oceanography*, *61*(S1), S13–S26. https://doi.org/10.1002/lno.10222

Nelson, S. A. C., Cheruvelil, K. S., & Soranno, P. A. (2006). Satellite remote sensing of freshwater macrophytes and the influence of water clarity. *Aquatic Botany*, *85*(4), 289–298. https://doi.org/10.1016/j.aquabot.2006.06.003

David Olefeldt, Mikael Hovemyr, McKenzie Kuhn, David Bastviken, Theodore Bohn, et al. 2021. The fractional land cover estimates from the Boreal-Arctic Wetland and Lake Dataset (BAWLD), 2021b. Arctic Data Center. https://doi.org/10.18739/A2C824F9X

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. I., Heffernan, L., Helbig, M., Hugelius, G., Hutchins, R., Juutinen, S., Lara, M. J., Malhotra, A., … Watts, J. D. (2021a). The Boreal-Arctic Wetland and Lake Dataset (BAWLD) D. Olefeldt et al.: The Boreal-Arctic Wetland and Lake Dataset. *Earth Syst. Sci. Data*, *13*, 5127–5149. https://doi.org/10.5194/essd-13-5127-2021

Parks Canada. (2019). *Wood Buffalo National Park World Heritage Site Action Plan* (Cat. No.: 978-0-660-27537-6). Retrieved from https://pcacdn.azureedge.net/-/media/pn-np/nt/woodbuffalo/2021-changes/02_11-action-plan-PDFs/WoodBuffalo-WHS-Action-Plan_EN.pdf

Pavelsky, T. M., & Smith, L. C. (2008). Remote sensing of hydrologic recharge in the Peace-Athabasca Delta, Canada. *Geophysical Research Letters*. https://doi.org/10.1029/2008GL033268

Rey-Sanchez, A. C., Morin, T. H., Stefanik, K. C., Wrighton, K., & Bohrer, G. (2018). Determining total emissions and environmental drivers of methane flux in a Lake Erie estuarine marsh. *Ecological Engineering*, *114*, 7–15. https://doi.org/10.1016/j.ecoleng.2017.06.042

Ribaudo, C., Bartoli, M., Longhi, D., Castaldi, S., Neubauer, S. C., & Viaroli, P. (2012). CO2 and CH4 fluxes across a Nuphar lutea (L.) Sm. stand. *Journal of Limnology*, *71*(1), 200–210.

Rosentreter, J. A., Borges, A. v, Deemer, B. R., Holgerson, M. A., Liu, S., Song, C., Melack, J., Raymond, P. A., Duarte, C. M., Allen, G. H., Olefeldt, D., Poulter, B., Battin, T. I., & Eyre, B. D. (2021). Half of global methane emissions come from highly variable aquatic

ecosystem sources. *Nature Geoscience*. https://doi.org/10.1038/s41561-021-00715-2

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., … Zhuang, Q. (2020). The Global Methane Budget 2000–2017. *Earth Syst. Sci. Data*, *12*, 1561–1623. https://doi.org/10.5194/essd-12-1561-2020

Schmiedeskamp, M., Stephanie, L., Praetzel, E., Bastviken, D., & Knorr, K.-H. (2021). Whole-lake methane emissions from two temperate shallow lakes with fluctuating water levels: Relevance of spatiotemporal patterns. *Limnol. Oceanogr*, *9999*, 1–15. https://doi.org/10.1002/lno.11764

Seekell, D., Cael, B., Norman, S., & Byström, P. (2021). Patterns and Variation of Littoral Habitat Size Among Lakes. *Geophysical Research Letters*, *48*(20), e2021GL095046. https://doi.org/10.1029/2021GL095046

Simard, M., Riel, B. V., Denbina, M., & Hensley, S. (2016). Radiometric Correction of Airborne Radar Images over Forested Terrain with Topography. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4488–4500. https://doi.org/10.1109/TGRS.2016.2543142

Slaymaker, O. (2016). *Landscapes and Landforms of Western Canada*. Springer International Publishing.

Smith, L. K., & Lewis, W. M. (1992). Seasonality of methane emissions from five lakes and associated wetlands of the Colorado Rockies. *Global Biogeochemical Cycles*, *6*(4), 323–338. https://doi.org/10.1029/92GB02016

Smith, V. H., & Wallsten, M. (1986). Prediction of emergent and floating-leaved macrophyte cover in Central Swedish lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *43*(12), 2519–2523. https://doi.org/10.1139/f86-311

Spence, C., & Woo, M. (2006). Hydrology of subarctic Canadian Shield: heterogeneous headwater basins. *Journal of Hydrology*, *317*(1–2), 138–154. https://doi.org/10.1016/J.JHYDROL.2005.05.014

Stephanie, L., Praetzel, E., Plenter, N., Schilling, S., Schmiedeskamp, M., Broll, G., … Praetzel@uni-Muenster, L. De. (2020). Organic matter and sediment properties determine in-lake variability of sediment CO 2 and CH 4 production and emissions of a small and shallow lake. *Biogeosciences*, *17*, 5057–5078. https://doi.org/10.5194/bg-17-5057-2020

Striegl, R. G., Dornblaser, M. M., McDonald, C. P., Rover, J. R., & Stets, E. G. (2012). Carbon dioxide and methane emissions from the Yukon River system. *Global Biogeochemical Cycles*, *26*(4). https://doi.org/10.1029/2012GB004306

Striegl, R. G., & Michmerhuizen, C. M. (1998). Hydrologic influence on methane and carbon dioxide dynamics at two north-central Minnesota lakes. *Limnology and Oceanography*, *43*(7), 1519–1529. https://doi.org/10.4319/lo.1998.43.7.1519

Ström, L., Mastepanov, M., & Christensen, T. R. (2005). Species-specific Effects of Vascular

Plants on Carbon Turnover and Methane Emissions from Wetlands. *Biogeochemistry*, *75*(1), 65–82. https://doi.org/10.1007/s10533-004-6124-1

Thornton, B. F., Wik, M., & Crill, P. M. (2016). Double-counting challenges the accuracy of high-latitude methane inventories. *Geophysical Research Letters*, *43*(24), 12,569-12,577. https://doi.org/10.1002/2016GL071772

Timoney, K. P. (2013). *The Peace-Athabasca Delta: Portrait of a Dynamic Ecosystem*. Edmonton: University of Alberta Press.

Töyrä, J., & Pietroniro, A. (2005). Towards operational monitoring of a northern wetland using geomatics-based techniques. Remote Sensing of Environment, 97(2), 174–191. https://doi.org/10.1016/J.RSE.2005.03.012

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., … Weyhenmeyer, G. A. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, *54*(6part2), 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298

Ulander, L. M. H. (1996). Radiometrie slope correction of synthetic-aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(5), 1115–1122. https://doi.org/10.1109/36.536527

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, *41*(18), 6396–6402. https://doi.org/10.1002/2014GL060641

Villa, J. A., Ju, Y., Yazbeck, T., Waldo, S., Wrighton, K. C., & Bohrer, G. (2021). Ebullition dominates methane fluxes from the water surface across different ecohydrological patches in a temperate freshwater marsh at the end of the growing season. *Science of the Total Environment*, *767*, 144498. https://doi.org/10.1016/j.scitotenv.2020.144498

Wang, J. A., Sulla-menashe, D., Woodcock, C. E., Sonnentag, O., Keeling, R. F., & Friedl, M. A. (2019). ABoVE: Landsat-derived Annual Dominant Land Cover Across ABoVE Core Domain, 1984-2014. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1691

Wang, J. A., Sulla-Menashe, D., Woodcock, C. E., Sonnentag, O., Keeling, R. F., & Friedl, M. A. (2019). Extensive land cover change across Arctic–Boreal Northwestern North America from disturbance and climate forcing. *Global Change Biology*, *00*, 1–16. https://doi.org/10.1111/gcb.14804

Ward, E. M., & Gorelick, S. M. (2018). Drying drives decline in muskrat population in the Peace-Athabasca Delta, Canada. *Environmental Research Letters*, 13, 124026. https://doi.org/10.1088/1748-9326/aaf0ec

West, W. E., Creamer, K. P., & Jones, S. E. (2016). Productivity and depth regulate lake contributions to atmospheric methane. Limnology and Oceanography, 61(S1), S51–S61.

https://doi.org/10.1002/LNO.10247

Wetzel, R. G. (1990). Land-water interfaces: Metabolic and limnological regulators. *Verh. Internat. Verein. Limnol.*, *24*(September), 6–24.

Wetzel, R. G. (2001). *Limnology: Lake and River Ecosystems* (Third). Boston: Academic Press.

Wik, M., Crill, P. M., Varner, R. K., & Bastviken, D. (2013). Multiyear measurements of ebullitive methane flux from three subarctic lakes. *Journal of Geophysical Research: Biogeosciences*, *118*(3), 1307–1321. https://doi.org/10.1002/jgrg.20103

Wik, M., Varner, R. K., Anthony, K. W., MacIntyre, S., & Bastviken, D. (2016a). Climate-sensitive northern lakes and ponds are critical components of methane release. *Nature Geoscience*, *9*(2), 99–105. https://doi.org/10.1038/ngeo2578

Wik, M., Thornton, B. F., Bastviken, D., Uhlbäck, J., & Crill, P. M. (2016b). Biased sampling of methane release from northern lakes: A problem for extrapolation. *Geophysical Research Letters*, *43*(3), 1256–1262. https://doi.org/10.1002/2015GL066501

Wolfe, B. B., Hall, R. I., Last, W. M., Edwards, T. W. D., English, M. C., Karst-Riddoch, T. L., … Palmini, R. (2006). Reconstruction of multi-century flood histories from oxbow lake sediments, Peace-Athabasca Delta, Canada. Hydrological Processes, 20(19), 4131–4153. https://doi.org/10.1002/hyp.6423

Zhang, B., Tian, H., Lu, C., Chen, G., Pan, S., Anderson, C., & Poulter, B. (2017). Methane emissions from global wetlands: An assessment of the uncertainty associated with various wetland extent data sets. *Atmospheric Environment*, *165*, 310–321. https://doi.org/10.1016/J.ATMOSENV.2017.07.001

Zhang, S., Foerster, S., Medeiros, P., de Araújo, J. C., & Waske, B. (2018). Effective water surface mapping in macrophyte-covered reservoirs in NE Brazil based on TerraSAR-X time series. *International Journal of Applied Earth Observation and Geoinformation*, *69*, 41–55. https://doi.org/10.1016/j.jag.2018.02.014

Zhang, Y., Jeppesen, E., Liu, X., Qin, B., Shi, K., Zhou, Y., … Deng, J. (2017). Global loss of aquatic vegetation in lakes. *Earth-Science Reviews*, *173*, 259–265. https://doi.org/https://doi.org/10.1016/j.earscirev.2017.08.013

Zhang, Z., Ni, W., Sun, G., Huang, W., Ranson, K. J., Cook, B. D., & Guo, Z. (2017). Biomass retrieval from L-band polarimetric UAVSAR backscatter and PRISM stereo imagery. *Remote Sensing of Environment*, *194*, 331–346. https://doi.org/10.1016/J.RSE.2017.03.034

Zhang, Z., Fluet-Chouinard, E., Jensen, K., Mcdonald, K., Hugelius, G., Gumbricht, T., … Poulter, B. (2021). Development of the global dataset of Wetland Area and Dynamics for Methane Modeling (WAD2M). Earth Syst. Sci. Data, 13. https://doi.org/10.5194/essd-13-2001-2021

JGR: Biogeosciences

Supporting Information for

The Importance of Lake Emergent Aquatic Vegetation for Estimating Arctic-Boreal Methane Emissions

Ethan D. Kyzivat[1], Laurence C. Smith[1], Fenix Garcia-Tigreros[2], Chang Huang[1, 3], Chao Wang[4], Theodore Langhorst[4], Jessica V. Fayne[5], Merritt E. Harlan[6], Yuta Ishitsuka[6], Dongmei Feng[10], Wayana Dolan[4], Lincoln H Pitcher[5, 8], Kimberly P. Wickland[7], Mark M. Dornblaser[7], Robert G. Striegl[7], Tamlin M. Pavelsky[4], David E. Butman[2,9], and Colin J. Gleason[6]

[1]Department of Earth, Environmental & Planetary Sciences and Institute at Brown for Environment & Society, Brown University, Providence, RI, 02912 USA

[2]School of Environmental and Forest Sciences, University of Washington, Seattle, WA, 98195 USA

[3]School of Urban and Environmental Sciences, Northwest University, Xi'an, Shaanxi, 710127 China

[4]Department of Earth, Marine and Environmental Sciences, University of North Carolina, Chapel Hill, NC, 27599 USA

[5]Department of Geography, University of California-Los Angeles, Los Angeles, CA, 90095 USA

[6]Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, MA, 01003 USA

[7]U.S. Geological Survey, Water Resources Mission Area, Boulder, CO, 80303 USA

[8]Cooperative Institute for Research in Environmental Sciences (CIRES). University of Colorado, Boulder. Boulder, CO, 80309, USA.

[9]School of Engineering and Environmental Sciences, University of Washington, Seattle, WA, 98195 US

[10]Department of Chemical and Environmental Engineering, University of Cincinnati, OH, 45221 USA

Contents of this file

Additional Supporting Information (Files uploaded separately)

Table S.3: Literature_flux_data.csv

Introduction

This file provides supplementary figures and tables.

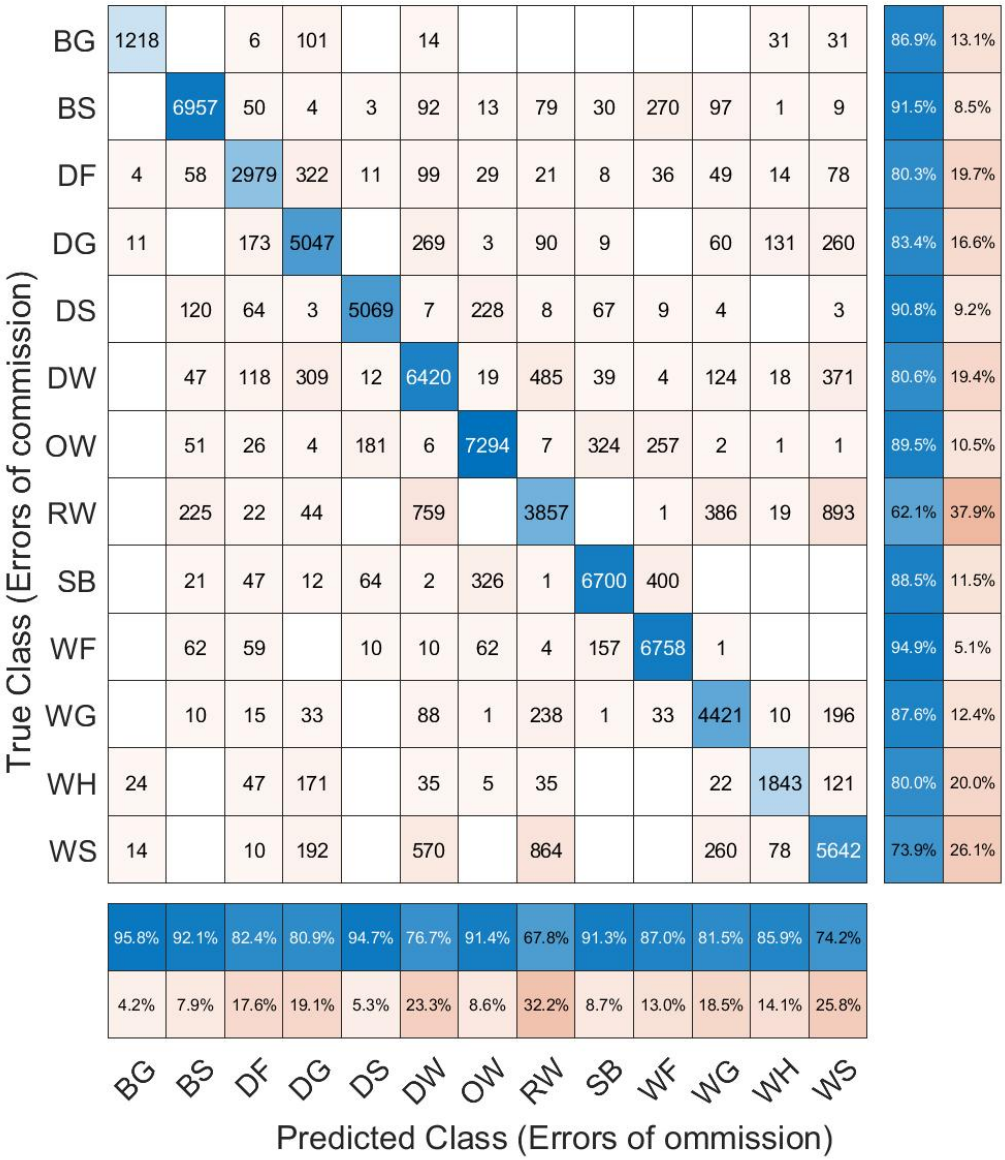

Figure S.1. Confusion matrix for the classifier used for the PAD, YF and CSB study areas. The classifier has an overall accuracy of 84.0% and kappa coefficient of 0.824.

Study area	Date	Scene(s) used
CSB	08/21/18	bakerc_16008_18047_005_180821_L090_CX_02
CSB	09/04/19	bakerc_16008_19059_012_190904_L090_CX_01
CSD	06/14/17	daring_21405_17063_010_170614_L090_CX_01

CSD	09/09/17	daring_21405_17094_010_170909_L090_CX_01
PAD	09/04/19	padelE_36000_19059_003_190904_L090_CX_01
PAD	06/13/17	PADELT_18035_17062_004_170613_L090_CX_01 PADELT_36000_17062_003_170613_L090_CX_01
PAD	09/08/17	padelE_36000_17093_007_170908_L090_CX_01 padelW_18035_17093_008_170908_L090_CX_01
PAD	08/21/18	padelE_36000_18047_000_180821_L090_CX_01 padelW_18035_18047_001_180821_L090_CX_01
YF	06/21/17	yflats_04707_17069_010_170621_L090_CX_01 yflats_21508_17069_009_170621_L090_CX_01
YF	09/16/17	ftyuko_04707_17098_007_170916_L090_CX_01 yflatE_21609_17098_008_170916_L090_CX_01 yflatW_21508_17098_006_170916_L090_CX_01
YF	08/27/18	ftyuko_04707_18051_008_180827_L090_CX_01 yflatE_21609_18051_009_180827_L090_CX_01
YF	09/14/19	ftyuko_04707_19064_006_190914_L090_CX_01 yflatE_21609_19064_007_190914_L090_CX_01

Table S.1. UAVSAR scenes used.

Feature creation parameters		
Parameter	**Value**	**Description**
Minimum incidence angle	0.5 radians	Minimum incidence angle to mask in radians
Maximum incidence angle	Infinity	Maximum incidence angle to mask in radians
Offset filter dimensions	3x3 px	Offset filter is simply a Gaussian smoothing filter applied to a center pixel a given offset away, used as input to classifier
Offset filter orientation	Parallel and anti-parallel to look angle	Direction relative to look angle
Offset filter gaussian width	2 px	Determines effective radius of filter, used as classifier input
Guided filter	5x5 px	Edge-preserving smoothing for classifier input

Standard deviation filter dimensions	5x5 px	Texture metric for classifier input
Use raw image	True	Use the raw, unfiltered image as a feature for classifier input.
Classifier parameters		
Parameter	Value	Description
Out-of-bag prediction error	0.167	Not a parameter, but a result
Number of trees	40	Number of decision trees
Minimum leaves per tree	30	Nodes per tree

Table S.2 Land cover classification filter parameters and random forests classifier parameters.

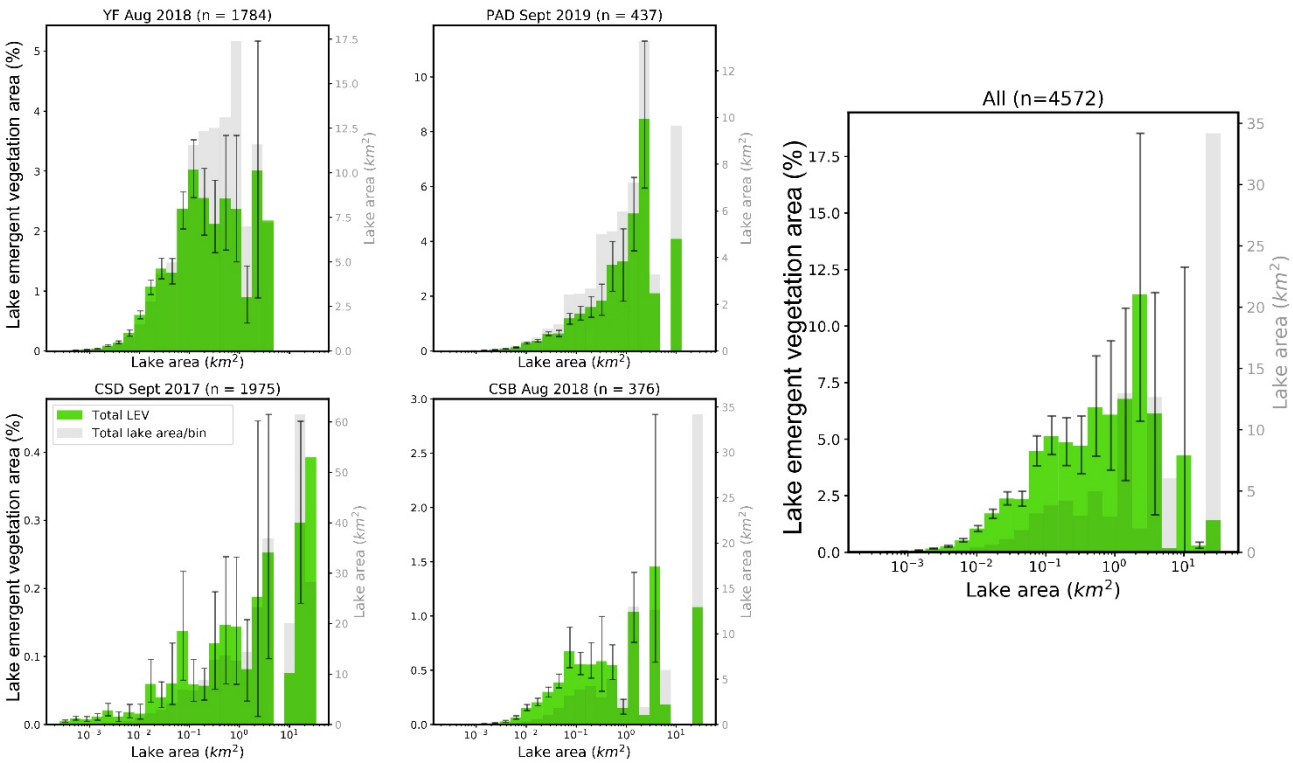

Figure S.2. Lake emergent vegetation (LEV) area summed by logarithmically-spaced lake area bins, in contrast with **Figure 4**, which uses bin means. Most LEV area comes from the largest size bins for each region. When combined (**right plot**), the trend still holds, although of the 10 lakes comprising the final four bins, all but one come from Canadian Shield lakes, so they are not showing a domain-wide trend. This situation, combined with the lesser macrophyte coverage in the Shield and correspondingly different y-axis scaling causes the outlier behaviour in the final four bins of the combined plot.

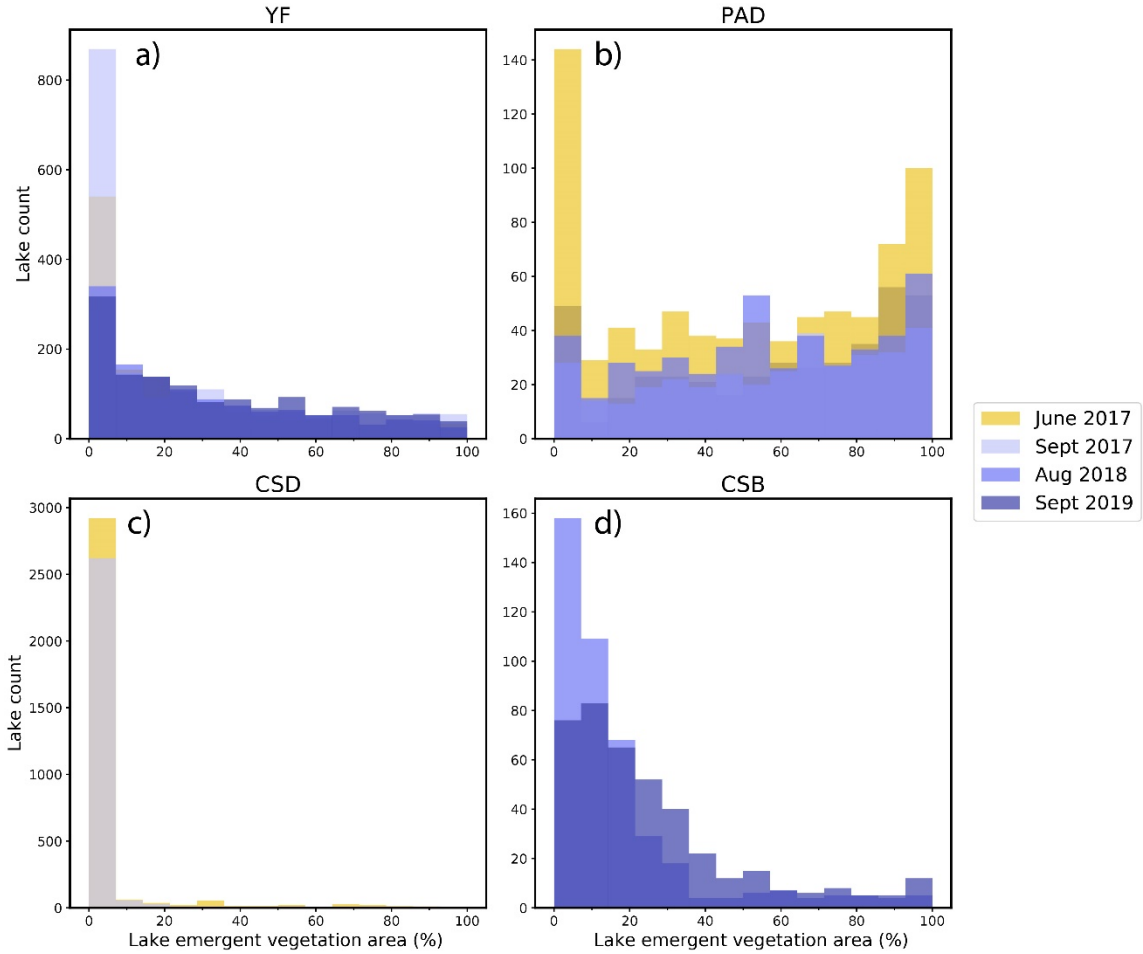

Figure S.3 Although the overall lake count changes across seasons and years as water bodies merge during high water seasons, the distributions of lake emergent vegetation (LEV) coverage remain similar. Histograms are made with 25 equally-spaced bins for each UAVSAR acquisition date for each region. Early summer dates (high water season) are plotted in gold and late summer in shades of purple, with intersections in shades of purple-grey. CSD was only acquired in June and September 2017 and CSB in August 2018 and September 2019.

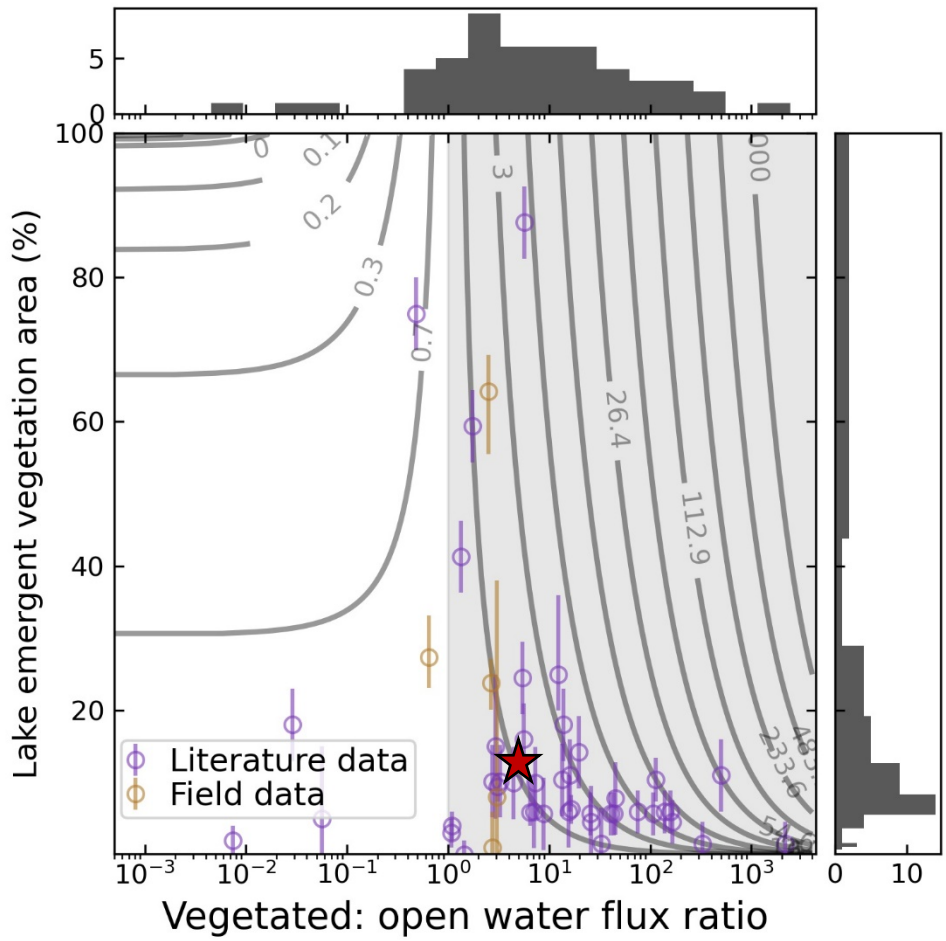

Figure S.4. Scatter plot of data from PAD and published literature showing the vegetated: open water methane flux ratio plotted against lake emergent vegetation (LEV) coverage as a percentage of each lake. The distributions of both variables are shown as histograms along the relevant axes. Vertical error bars show the temporal range in coverage for the field data (orange) and the estimated mapping uncertainty for the literature data (purple). Points falling in the shaded region come from lakes that would have higher calculated fluxes if their LEV zones are accounted for separately from open water. Contour lines show how much higher this calculated flux would be (*I*) and are logarithmically spaced to achieve uniform separation in a log-log space. Using the corrected median flux ratio (4.6) and area-weighted mean macrophyte coverage (16%) leads to fluxes 21% times greater (located at the red star). Note the logarithmically-scaled x-axis and linearly-scaled y-axis.

Figure S.6: Photos of emergent macrophyte (left) and open water (right) chamber flux collection.

Table S.3. See supplementary file "Literature_flux_data.csv" for a table showing collection dates and locations for field flux measurements at 15 lakes in the Peace-Athabasca Delta, July-August 2019. Methane fluxes are given in units of $mgCH_4/m^2/day$ for the type of lake zone considered (LEV = lake emergent vegetation, OW = open water, S = shallow, D = deep) and include attributes for confidence intervals or ranges, if given; flux pathway(s); emergent macrophyte delineation method uncertainty, and percentage; total macrophyte percentage, if applicable; and citation. The flux ratio is calculated based on the lake zone division of the paper (LEV versus OW or S versus D).

Additional data published on the Environmental Data Initiative (EDI, https://doi.org/10.6073/pasta/1e0cadadd8024c8fabc692ee21dc1f57) contains a table showing collection dates and locations for field flux measurements at 15 lakes in the Peace-Athabasca Delta, July-August 2019. Fluxes are given in units of $mol/m^2/day$ for both methane and carbon dioxide and include attributes for location and vegetation type, if applicable, as well as a quality flag that indicates if the data were used.

		UAVSAR		
		Land	**Lake**	**LEV**
GLO-WABO	**Not lake**	32.9%	19.6%	83.0%
	Lake	67.1%	80.4%	**17.0%**

		UAVSAR		
		Land	**Lake**	**LEV**
Hydro-Lakes	**Not lake**	15.0%	3.5%	45.0%
	Lake	85.0%	96.5%	55.0%

Table S.4. Confusion matrices between two global lake datasets and our lake classification from UAVSAR, normalized by column totals. From the total area of lake emergent vegetation (LEV) considered in the analysis, these matrices show that only 17.0% (GLOWABO; Verpoorter et al. 2014) to 55.0% (HydroLakes; Messager et al., 2016) coincides with global dataset lakes, which are commonly used to distinguish between lakes and wetlands for methane modelling. Therefore, we use the mean value of 0.36 as the scalar *c* that corrects for double-counting between our mapped lake emergent macrophytes and areas that are already considered (high-emitting) wetlands in global datasets.

	HydroLakes			GLOWABO		
	Not Lake	**Lake**	**Total**	**Not Lake**	**Lake**	**Total**
Lake	700206	19550038	20250244	4113557	16837558	20951115
WG	129790	163736	293526	301827	63198	365025
WS	5575	2148	7723	8778	577	9355
WF	93	4	97	104	1	105
Wetland WG	874	8195	9069	755	4285	5040
Wetland WS	7	107	114	5	8	13
Wetland WF	0	0	0	0	0	0
Other	483521	2737609	3221130	771296	1570154	2341450
Total	1320066	22461837	23781903	5196322	18475781	23672103
WG/LEV	**95.8%**	**98.7%**	**97.4%**	**97.1%**	**99.1%**	**97.5%**

Table S.5. More detailed confusion matrices between two global lake datasets and our lake classification from UAVSAR. Unlike **Table S.4**, lake emergent vegetation (LEV) is broken out into wet graminoid (WG), wet shrub (WS), and wet forest (WF) classes to facilitate comparing their relative proportions after comparing to the global datasets. As mentioned in **Section 2.3.8**, prior to the comparison to the global datasets, lakes from HydroLakes/GLOWABO or our classification were excluded if they didn't overlap at least partly with a lake in the comparison dataset. This step generally removed the smallest lakes and most of the WS and WF classes. The confusion matrix shows that regardless of global dataset or agreement with its lake classes, most of the remaining LEV is WG.

	Region	Baker	Daring	PAD	YF	Late summer mean	Late summer area-weighted mean
	Area (km^2)	1164.9	3035.0	1556.2	5365.3	2780.3	11121.3
BAWLD land cover (%)	**GLA**	0.0	0.0	0.0	0.0	0.0	0.0
	ROC	14.0	15.6	12.3	0.3	10.6	7.6
	TUN	2.2	35.9	0.4	25.1	15.9	22.2
	BOR	36.2	13.7	39.5	46.2	33.9	35.3
	PEB	3.2	5.4	4.3	7.7	5.1	6.1
	WTU	1.0	2.3	1.0	4.6	2.2	3.1
	MAR	1.0	0.1	3.1	2.2	1.6	1.6
	BOG	1.5	0.0	8.3	1.1	2.7	1.9
	FEN	1.9	0.3	9.3	3.1	3.6	3.1
	LAL	24.4	10.7	13.4	0.0	12.1	7.4
	MPL	2.4	0.9	2.3	1.5	1.8	1.5
	MYL	0.0	0.0	0.0	1.3	0.3	0.6
	MGL	10.0	11.5	2.4	1.3	6.3	5.2
	SPL	0.6	0.2	1.0	0.6	0.6	0.6
	SYL	0.0	0.0	0.0	0.5	0.1	0.2
	SGL	1.6	3.2	0.6	0.1	1.4	1.2
	RIV	0.1	0.1	2.2	4.4	1.7	2.5
	Total	100.0	100.0	100.0	100.0	100.0	100.0
BAWLD land cover summary (%)	**LAK**	39.0	26.6	19.6	5.3	22.6	**16.6**
	WET	8.5	8.1	26.0	18.8	15.3	15.8
	MAR + FEN	2.9	0.4	12.5	5.3	5.3	**4.7**
UAVSAR land cover (%)	**Open lake**	24.5	25.9	12.3	6.1	16.9	**16.6**
	LEV	2.8	0.5	10.7	2.1	3.6	2.7
	WF	0.0	0.0	0.1	0.1	0.0	0.0
	WS	0.6	0.0	3.7	0.4	1.0	0.7
	WG	2.1	0.5	7.0	1.5	2.5	1.9
	WEV	1.5	0.1	2.1	1.7	1.3	1.1
	LEV + WEV	4.3	0.6	12.9	3.8	4.9	**3.8**

Table S.6. Comparison between UAVSAR land cover classification and the Boreal and Arctic Wetland and Lake Dataset (BAWLD, Olefeldt et al., 2021a, Olefeldt et al., 2021b). Relevant BAWLD classes are permafrost bogs (PEB), tundra wetlands (WTU),

marshes (MAR), bogs (BOG), fen (FEN), mid-sized peatland lakes (MPL), mid-sized yedoma lakes (MYL), mid-sized glacial lakes (MGL), small peatland lakes (SPL), small yedoma lakes (SYL), small glacial lakes (SGL), rivers (RIV), total lakes (LAK, defined as lentic open-water ecosystems), and total wetlands (WET). Since individual wetland classes are not equivalent between datasets, we suggest comparing total BAWLD MAR and FEN with total UAVSAR lake emergent vegetation (LEV) and wetland emergent vegetation (WEV) as roughly equivalent open water wetland classes. Study area-weighted mean open lake coverage shows remarkable agreement between the datasets (16.6% in both with variability based on study area), and the equivalent emergent vegetation and/or wetland classes are 24% greater in BAWLD (3.8% from UAVSAR, 4.7% from BAWLD). In summary, current methods show good agreement in detecting open water (including submerged vegetation) lakes, and poor agreement in detecting wetlands or total inundation, even when lake and wetland classes are mutually exclusive within each dataset.

Works cited

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*, 1–11. https://doi.org/10.1038/ncomms13603

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. I., Heffernan, L., Helbig, M., Hugelius, G., Hutchins, R., Juutinen, S., Lara, M. J., Malhotra, A., … Watts, J. D. (2021a). The Boreal-Arctic Wetland and Lake Dataset (BAWLD) D. Olefeldt et al.: The Boreal-Arctic Wetland and Lake Dataset. *Earth Syst. Sci. Data*, *13*, 5127–5149. https://doi.org/10.5194/essd-13-5127-2021

David Olefeldt, Mikael Hovemyr, McKenzie Kuhn, David Bastviken, Theodore Bohn, et al. 2021. The fractional land cover estimates from the Boreal-Arctic Wetland and Lake Dataset (BAWLD), 2021b. Arctic Data Center. https://doi.org/10.18739/A2C824F9X

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, *41*(18), 6396–6402. https://doi.org/10.1002/2014GL060641