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

Figures S1 to S.6
Tables S1 to S.5

Additional Supporting Information (Files uploaded separately)

Table S.3: Literature_flux_data.csv

Introduction

This file provides supplementary figures and tables.

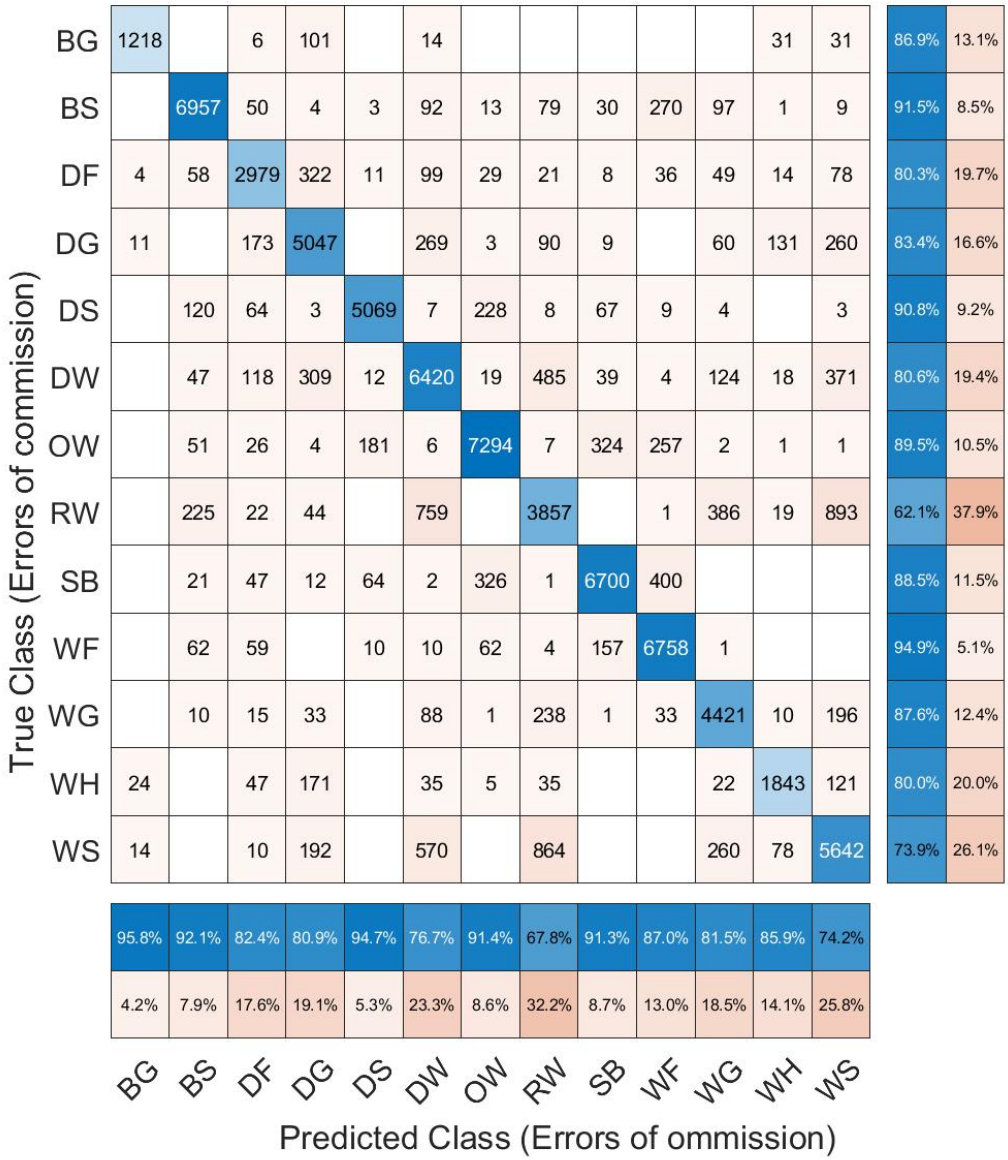

Figure S.1. Confusion matrix for the classifier used for the PAD, YF and CSB study areas. The classifier has an overall accuracy of 84.0% and kappa coefficient of 0.824.

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.

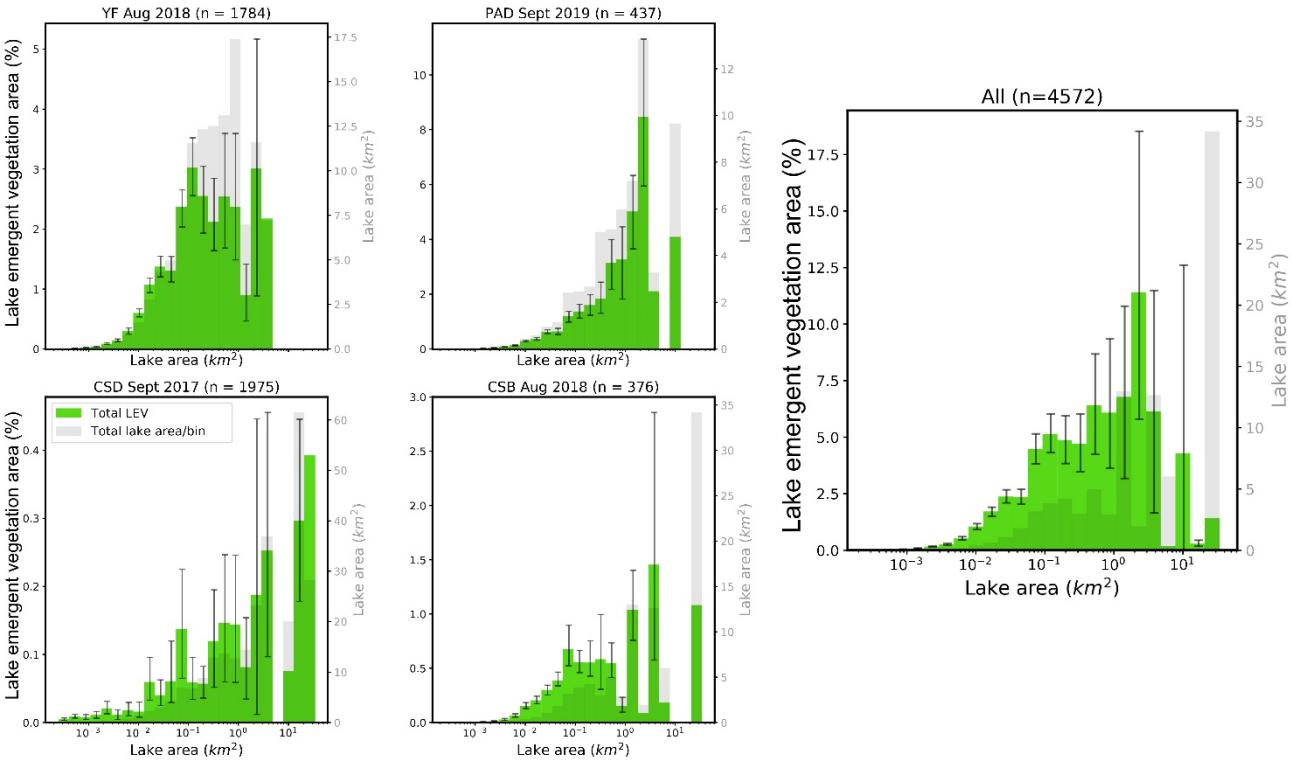

Figure S.2. Lake emergent vegetation (LEV) area summed by logarithmically-spaced lake area bins, in contrast with **Figure 4**, which uses bin means. Most LEV area comes from the largest size bins for each region. When combined (**right plot**), the trend still holds, although of the 10 lakes comprising 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.

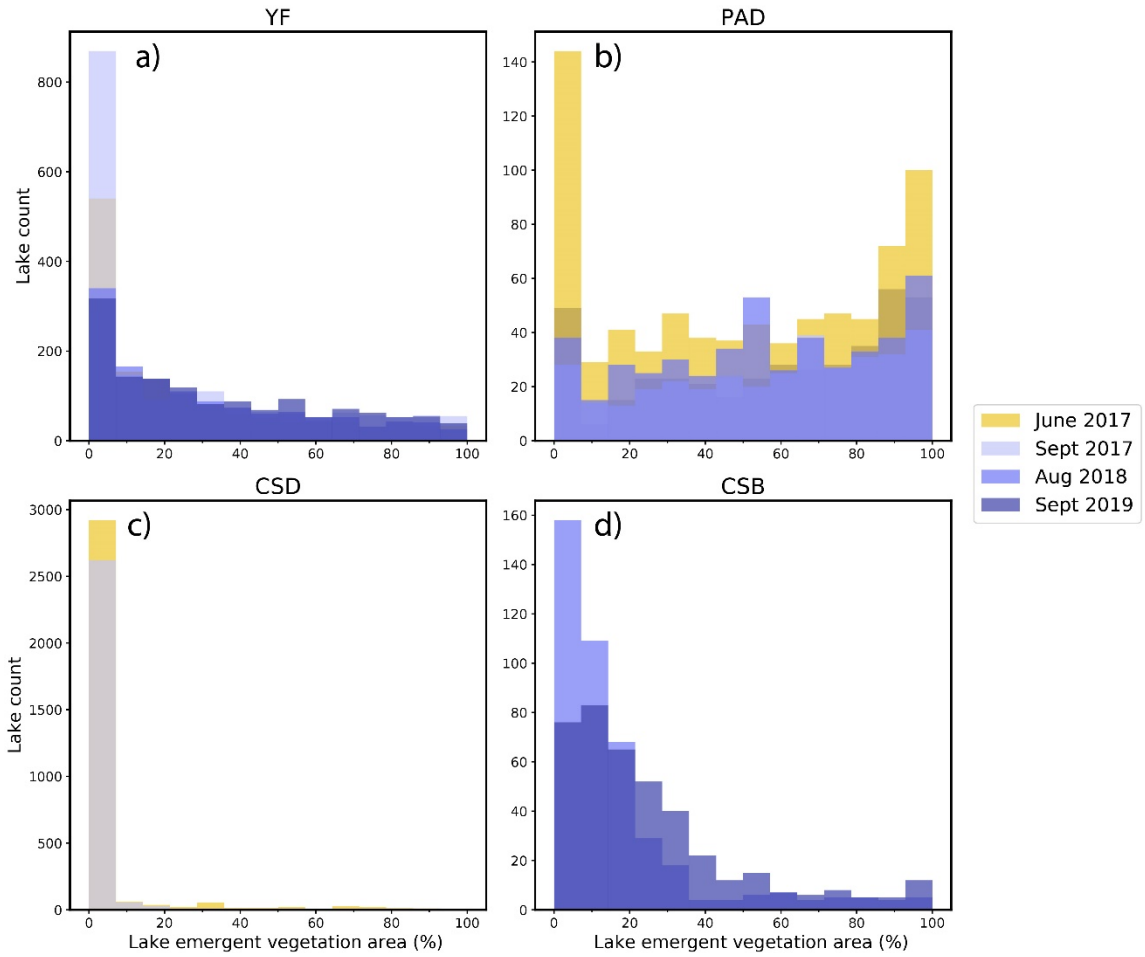

Figure S.3 Although the overall lake count changes across seasons and years as water bodies merge during high water seasons, the distributions of lake emergent vegetation (LEV) coverage remain similar. Histograms are made with 25 equally-spaced bins for each UAVSAR acquisition date for each region. Early 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.

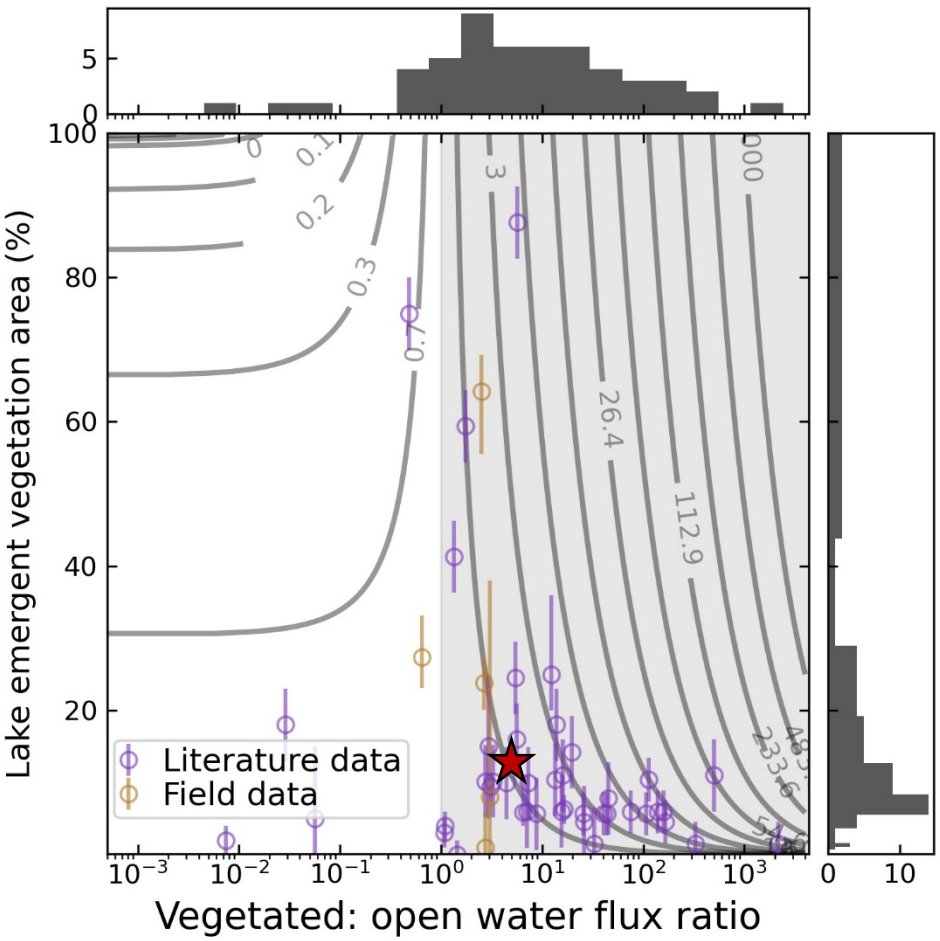

Figure S.4. Scatter plot of data from PAD and published literature showing the vegetated: open water methane flux ratio plotted against lake emergent vegetation (LEV) coverage as a percentage of each lake. The distributions of both variables are shown as histograms along the relevant axes. Vertical error 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