Supplemental Information for Geospatial analysis of Alaskan lakes indicates wetland fraction and surface water area are useful predictors of methane ebullition

Contents

1. Supplementary Tables

2. Supplementary Figures

3. Citations

1. Supplementary Tables

Table SI1. Compiled regression results from past ebullition regression studies.

Reference	Model	R²	Adj. R	rmse	aic	f	n	P<0.005?
Bastviken et al. 2004	log(Area)		0.78				17	yes
	log(Area) + log(TP)		0.89				13	yes, no
	log(TP)		0.46				13	no
DelSontro et al. 2016	Sediment temp (ponds)	0.59					77	yes
	Sediment temp (lakes)	0.015					83	no
	log(TP) + Sediment temp + log(TP)*Sediment temp	0.52					89	yes
Wik et al. 2016	Waterbody type					8.857	51	yes
	Sediment type					8.937		yes
	log(depth)					0.047		no
	waterbody type*sediment type					0.086		no
	waterbody type*log(depth)					8.491		no
	sediment type*log(depth)					5.238		no
	Waterbody type					0.532	51	no
	Sediment type					0.946		no
	log(area)					3.444		no
	waterbody type*sediment type					3.405		no
	waterbody type*log(area)					2.955		no
	sediment type*log(area)					0.444		no
DelSontro, Beaulieu, and Downing 2018	log(TP)	0.292		0.648			101	yes
	log(TN)	0.311		0.647			47	yes
	log(Chl a)	0.317		0.630			65	yes
	log(area)	0.013		0.775			137	no
	log(area)*log(TP)	0.292		0.648			101	yes
	log(area)*log(TN)	0.387		0.610			47	yes
	log(area)*log(chl a)	0.280		0.634			64	yes
Sanches et al. 2019	Estimation method	0.85					78	yes
	Climatic zone							yes
	Min temp							no
	Avg. temp							no
	Estimation method*max temp							no
	Estimation method*avg temp							no

Study	Predictor							
	Climatic zone*max temp							no
	Climatic zone*avg temp							no
	Climatic zone*year							no
	Precipitation							no
	Climatic zone*min precip							no
	Min temp*avg temp							no
	Min temp*min precip							no
	Max temp*avg temp							yes
	Area	0.98					46	no
	Landscape							yes
	Min temp							yes
	Avg temp							yes
	Area*landscape							yes
	Max precip	0.91					19	no
	Avg temp							yes
	Year precip*DOC							no
Kuhn et al. 2021	DOC		0.14			12.25	72	yes
	Area		0.08			13.88	165	yes
	Latitude		0.03			5.38	161	no
	Water temp		0.06			5.55	68	no
	Depth					0.02	152	no
	Area		0.21			19.85	69	yes
Deemer and Holgerson, 2021	Water temp	0.09					134	yes
	Latitude	0.04					218	yes
	ln(Area)	0.00					216	no
	ln(Max depth)	0.07					135	yes
	ln(Mean depth)	0.00					87	no
	ln(DOC)	0.00					85	no
	ln(Chl a)	0.18					143	yes
	Waterbody type + latitude + chl a + area		0.29		481.69		130	
	Waterbody type*chl a + latitude + area		0.29		483.17		130	

Table SI2. All compiled environmental variables with sources, variable descriptions, shorthand variable names, data formats, and spatial resolutions. Selected representative variables which appear in Tables 1-3 of the main text are shown in bold. Representative variable selection was based on the highest adj. R^2 and lowest AIC within each dataset, except in the cases of the climate (Thornton et al. 2020) and soil carbon (Poggio et al. 2021) datasets, where variables were so highly correlated (Figures SI4 and SI5) that physical meaningfulness of the variables was also taken into account.

Source	Variable Description	Variable Name	Data Format	Resolution	n
Engram, Walter Anthony, and Meyer 2020	Region (Atqasuk, Barrow, Fairbanks, Seward, or Toolik)	**Region**	Categorical	0.005 km²	5,143
	Methane ebullition flux (mg m⁻² d⁻¹)	**MassFlxCH4**	Vector	0.005 km²	5,143
	Lake area (km²)	**AreaSqkm**			
Derived from Engram, Walter Anthony, and Meyer 2020	Lake perimeter (km)	**perimeter**	Tabular	0.005 km²	
	Buffer ratio	bf_pr_wi			
	Perimeter-to-area (P/A) ratio	p_a_ratio			
Messager et al. 2016	Shoreline development	Shore_dev	Vector	0.1 km²	1,224
	Lake volume (mcm)	Vol_total			
	Lake depth (m)	**Depth_avg**			
	Watershed area (km²)	Wshd_area			
Pastick et al. 2015	Mean permafrost probability within 1 m of the surface	**Pfrst_mean**	Raster	30 m	5,143
Wang et al. 2019	Fraction of combined wetland class (LWF) in lake and surrounding wetlands within 100m of lake	lbf_100n	Raster	30 m	5,139
	Fraction of littoral zone in lake and surrounding wetlands within 100m of lake	littoral_100n			
	Fraction of bog in lake and surrounding wetlands within 100m of lake	bog_100n			
	Fraction of fen in lake and surrounding wetlands within 100m of lake	fen_100n			
	Fraction of sparsely vegetated land among land pixels within 100m of lake	sparseveg_100n			
	Fraction of deciduous forest among land pixels within 100m of lake	decid_100n			

	Fraction of evergreen forest among land pixels within 100m of lake	evrgrn_100n			
	Fraction of barren land among land pixels within 100m of lake	barren_100n			
	Fraction of shrubland among land pixels within 100m of lake	shrub_100n			
	Fraction of herbaceous land among land pixels within 100m of lake	herb_100n			
Thornton et al. 2020	Average winter temperature (˚C)	**Tavg_W**	Raster	1 km	5,141
	Average annual temperature (˚C)	Tavg_yr			
	Average annual vapor pressure (Pa)	Vpavg_yr			
	Maximum annual temperature (˚C)	Tmax_yr			
	Minimum annual temperature (˚C)	Tmin_yr			
	Total annual precipitation (mm)	Precip_yr			
Poggio et al. 2021	Organic carbon stocks (t/ha)	SOCS	Raster	250 m	5,130
	Soil organic carbon content in the fine earth fraction (dg/kg)	**SOC**			
	Proportion of silt particles (≥ 0.002 mm and ≤ 0.05 mm) in the fine earth fraction (g/kg)	Silt			
	Proportion of sand particles (> 0.05 mm) in the fine earth fraction (g/kg)	Sand			
	Soil pH (pHx10)	Soil_pH			
	Organic carbon density (hg/m³)	OCD			
	Total nitrogen (N) (cg/kg)	Nitrogen			
	Cation Exchange Capacity of the soil (mmol(c)/kg)	Cat_Ex			
	Bulk density of the fine earth fraction (cg/cm³)	Bulk_Dens			
	Proportion of clay particles (< 0.002 mm) in the fine earth fraction (g/kg)	Clay			
	Volumetric fraction of coarse fragments (> 2 mm) (cm3/dm3 (vol‰))	Co_Frag			
Kuhn and Butman, 2021	Mean Landsat growing season surface reflectance (Rs) in the green wavelengths	**Greenness**	Tabular	0.1 km²	1,061

Table SI3. Individual regression results for all assembled variables sorted and shaded by dataset (colors indicate divisions between datasets). Selected representative variables which appear in Tables 1-3 of the main text are shown in bold. Representative variable selection was based on the highest adj. R^2 and lowest AIC within each dataset, except in the cases of the climate (Thornton et al. 2020) and soil carbon (Poggio et al. 2021) datasets, where variables were so highly correlated (Figures SI4 and SI5) that physical meaningfulness of the variables was also taken into account.

Variable	rs	r2_adj	aic	n	cond_no	p < 0.005
Region	**0.321**	**0.320**	**3845**	**5143**	**5.99**	**yes**
AreaSqkm	**0.201**	**0.201**	**4678**	**5143**	**5.47**	**yes**
perimeter	**0.176**	**0.176**	**4836**	**5143**	**30.1**	**yes**
bf_per_wi	0.166	0.166	4895	5143	36.6	yes
p_a_ratio	0.175	0.175	4838	5143	11.8	yes
Shore_dev	0.002	0.001	873.2	1224	8.68	no
Vol_total	0.009	0.008	864.5	1224	2.23	yes
Depth_avg	**0.026**	**0.026**	**842.9**	**1224**	**7.27**	**no**
Wshd_area	0.003	0.003	871.6	1224	1.42	no
Pfrst_mean	**0.038**	**0.038**	**5630**	**5143**	**32.3**	**yes**
lbf_100n	**0.211**	**0.211**	**4564**	**5132**	**8.49**	**yes**
barren_100n	0.007	0.006	5794	5139	33.6	yes
bog_100n	0.000	0.000	5826	5139	2760	no
decid_100n	0.028	0.027	5684	5139	92.9	yes
evrgrn_100n	0.000	0.000	5826	5139	24.0	no
fen_100n	0.030	0.030	5672	5139	38.1	yes
herb_100n	0.073	0.073	5438	5139	14.8	yes
littoral_100n	0.112	0.112	5173	5132	13.9	yes
shrub_100n	0.099	0.099	4964	5017	10.4	yes
sparseveg_100n	0.163	0.163	4915	5139	11.2	yes
Tavg_W	**0.157**	**0.157**	**4949**	**5141**	**2670**	**yes**
Tavg_yr	0.147	0.147	5011	5141	2690	yes

VPavg_yr	0.160	0.160	4934	5141	138	yes
Tmax_yr	0.046	0.046	5586	5141	2340	yes
Tmin_yr	0.043	0.042	5606	5141	6040	yes
Precip_yr	0.132	0.132	5103	5141	39.1	yes
SOCS	0.023	0.023	5679	5129	16.2	yes
SOC	**0.060**	**0.059**	**5485**	**5129**	**42.2**	**yes**
Silt	0.140	0.140	5028	5129	23.1	yes
Sand	0.010	0.010	5747	5129	24.4	yes
Soil_pH	0.075	0.075	5399	5129	14.2	yes
OCD	0.080	0.080	5370	5129	28.7	yes
Nitrogen	0.093	0.093	5297	5129	32.8	yes
Cat_Ex	0.071	0.071	5422	5129	26.4	yes
Bulk_Dens	0.116	0.116	5167	5129	12.3	yes
Clay	0.052	0.052	5527	5129	20.7	yes
Co_Frag	0.097	0.097	5278	5129	7.18	yes
Greenness	**0.017**	**0.016**	**574.4**	**1061**	**28.2**	**yes**

1. Supplementary Figures

Figure SI1. Correlation matrix for lake morphometry variables derived from Engram, Walter Anthony, and Meyer (2020).
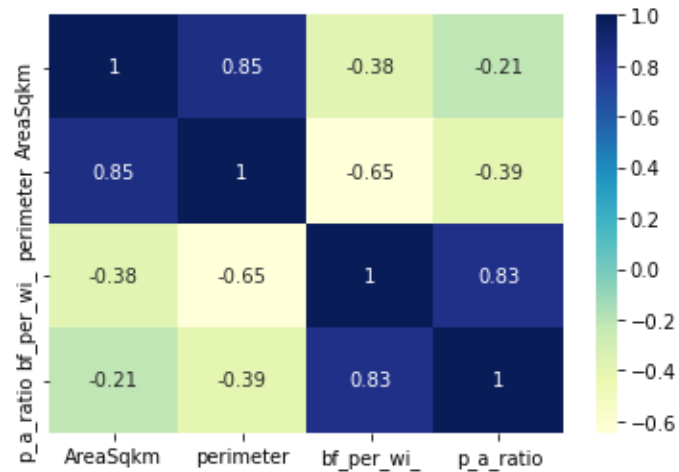

Figure SI2. Correlation matrix for lake morphometry variables from Messager et al. (2016).

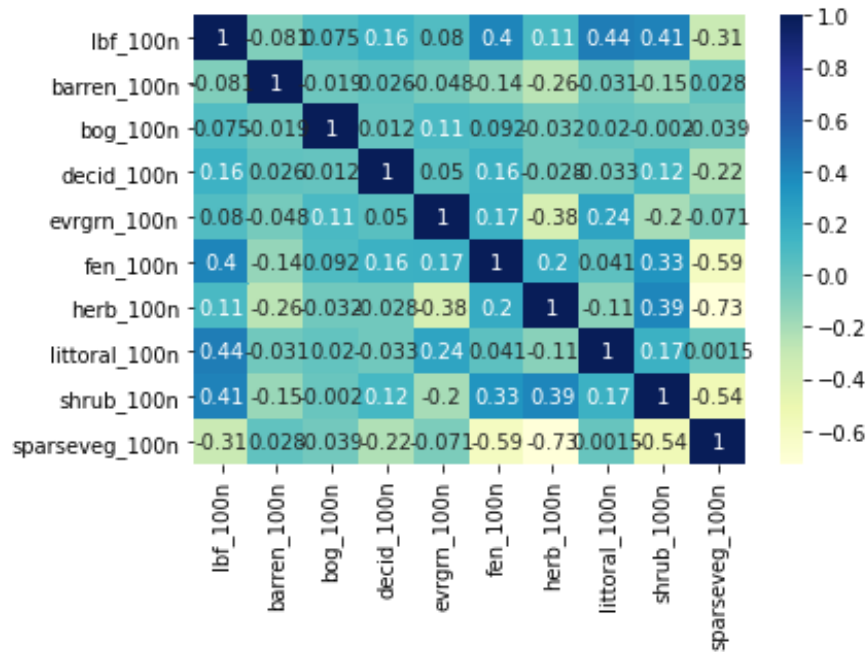Figure SI3. Correlation matrix for land cover variables from Wang et al. (2020).

	lbf_100n	barren_100n	bog_100n	decid_100n	evrgrn_100n	fen_100n	herb_100n	littoral_100n	shrub_100n	sparseveg_100n
lbf_100n	1	-0.081	0.075	0.16	0.08	0.4	0.11	0.44	0.41	-0.31
barren_100n	-0.081	1	0.019	0.026	0.048	-0.14	-0.26	-0.031	-0.15	0.028
bog_100n	0.075	0.019	1	0.012	0.11	0.092	-0.032	0.02	-0.002	0.039
decid_100n	0.16	0.026	0.012	1	0.05	0.16	-0.028	0.033	0.12	-0.22
evrgrn_100n	0.08	-0.048	0.11	0.05	1	0.17	-0.38	0.24	-0.2	-0.071
fen_100n	0.4	-0.14	0.092	0.16	0.17	1	0.2	0.041	0.33	-0.59
herb_100n	0.11	-0.26	-0.032	-0.028	-0.38	0.2	1	-0.11	0.39	-0.73
littoral_100n	0.44	-0.031	0.02	-0.033	0.24	0.041	-0.11	1	0.17	0.0015
shrub_100n	0.41	-0.15	-0.002	0.12	-0.2	0.33	0.39	0.17	1	-0.54
sparseveg_100n	-0.31	0.028	0.039	-0.22	-0.071	-0.59	-0.73	0.0015	-0.54	1

Figure SI4. Correlation matrix for climate variables from Thornton et al. (2020).

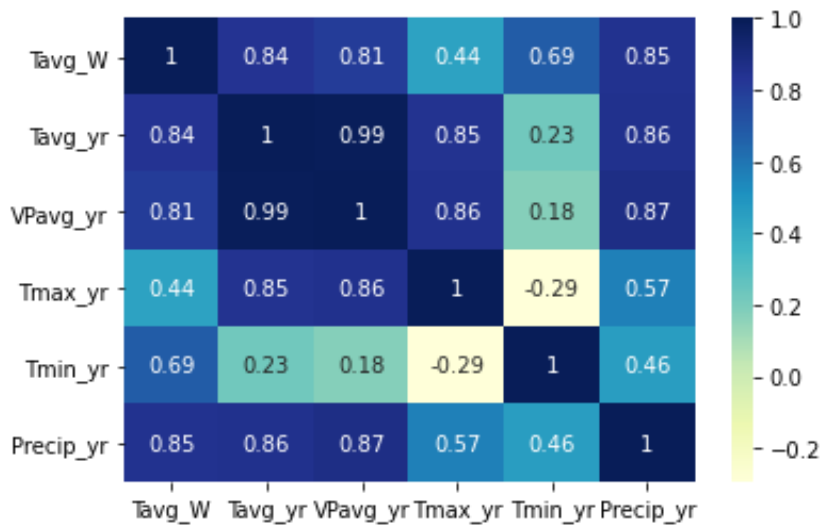|	Tavg_W	Tavg_yr	VPavg_yr	Tmax_yr	Tmin_yr	Precip_yr
Tavg_W	1	0.84	0.81	0.44	0.69	0.85
Tavg_yr	0.84	1	0.99	0.85	0.23	0.86
VPavg_yr	0.81	0.99	1	0.86	0.18	0.87
Tmax_yr	0.44	0.85	0.86	1	-0.29	0.57
Tmin_yr	0.69	0.23	0.18	-0.29	1	0.46
Precip_yr	0.85	0.86	0.87	0.57	0.46	1

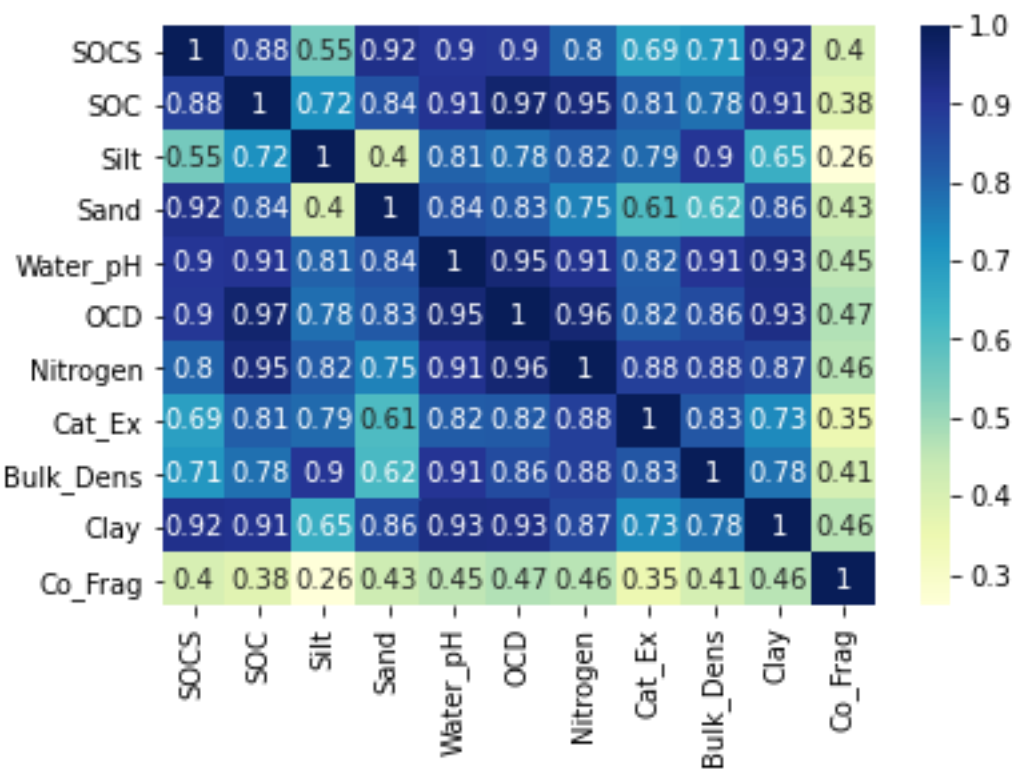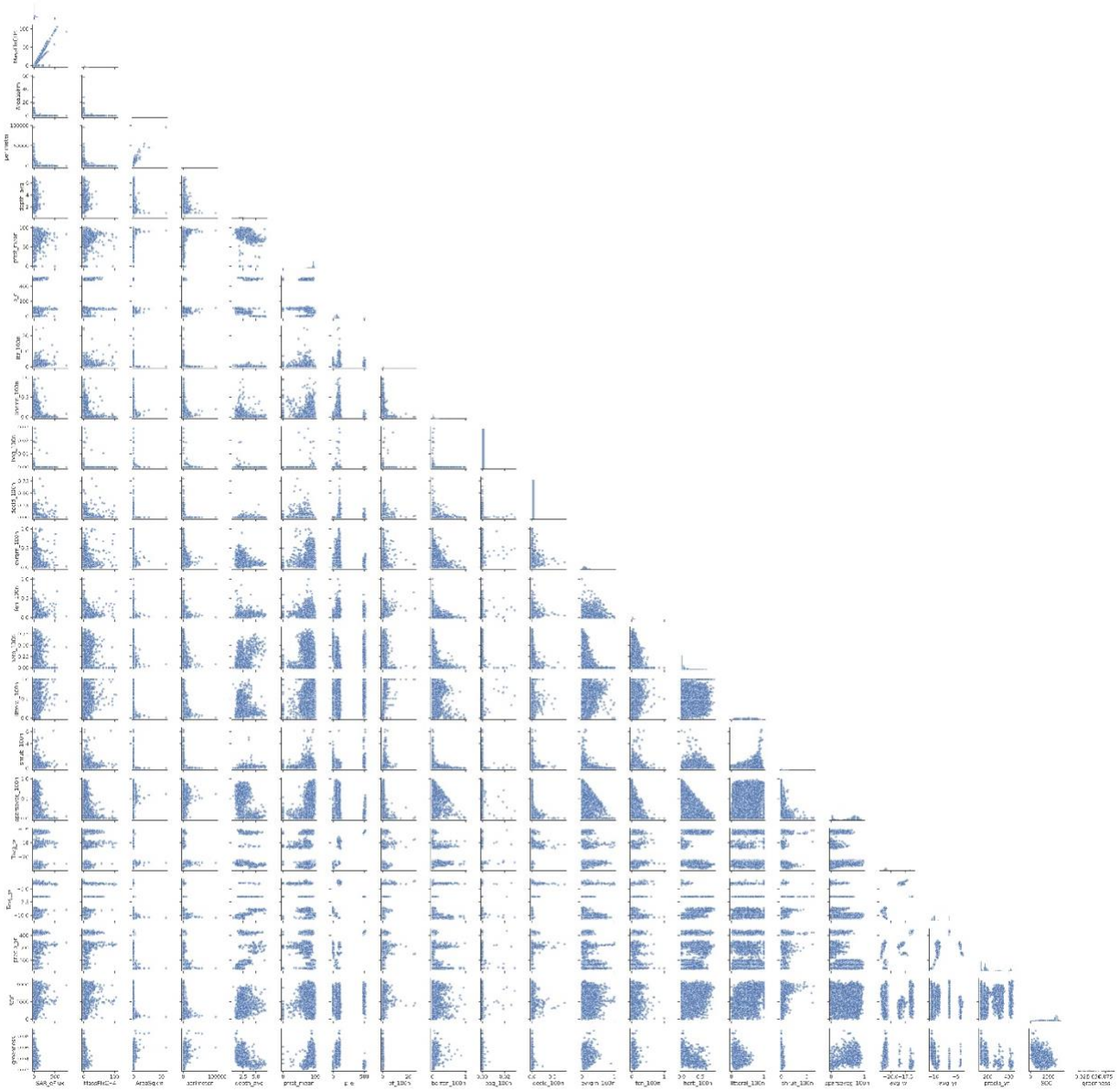Figure SI5. Correlation matrix for soil variables from Poggio et al. (2021).

Figure SI6. Pair plots of volumetric and mass-based ebullitive methane fluxes for selected predictor variables. Histograms are shown on the diagonal elements and scatter plots on the lower diagonal elements. Figure is high-resolution and can be zoomed to view individual plots.

3. References

Bastviken, D., J. Cole, M. Pace, and L. Tranvik. 2004. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. Global biogeochemical cycles, 18(4). doi: 10.1029/2004gb002238.

Deemer, B.R. and M.A. Holgerson. 2021. Drivers of methane flux differ between lakes and reservoirs, complicating global upscaling efforts. Journal of Geophysical Research: Biogeosciences, 126(4), p.e2019JG005600. doi: 10.1029/2019JG005600.

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

DelSontro, T., L. Boutet, A. St-Pierre, P.A. del Giorgio, and Y.T. Prairie. 2016. Methane ebullition and diffusion from northern ponds and lakes regulated by the interaction between temperature and system productivity. Limnology and Oceanography, 61(S1), pp.S62-S77. doi: 10.1002/lno.10335.

Engram, M.J., K. Walter Anthony, and F.J. Meyer. 2020. ABoVE: SAR-based Methane Ebullition Flux from Lakes, Five Regions, Alaska, 2007-2010. ORNL DAAC, Oak Ridge, Tennessee, USA. doi: 10.3334/ORNLDAAC/1790.

Kuhn, C., and D. Butman. 2021. ABoVE: Lake Growing Season Green Surface Reflectance Trends, AK and Canada, 1984-2019. ORNL DAAC, Oak Ridge, Tennessee, USA. doi: 0.3334/ORNLDAAC/1866.

Kuhn, M.A., R.K. Varner, D. Bastviken, P. Crill, S. MacIntyre, M. Turetsky, K. Walter Anthony, A.D. McGuire, and D. Olefeldt. 2021. BAWLD-CH 4: a comprehensive dataset of methane fluxes from boreal and arctic ecosystems. Earth System Science Data, 13(11), pp.5151-5189. doi: 10.5194/essd-13-5151-2021.

Messager, M.L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume
and age of water stored in global lakes using a geo-statistical approach. Nature
communications, 7(1), pp.1-11. doi: 10.1038/ncomms13603. Data is available at
www.hydrosheds.org.

Pastick, N.J., M.T. Jorgenson, B.K. Wylie, S.J. Nield, K.D. Johnson, and A.O. Finley. 2015.
Distribution of near-surface permafrost in Alaska: Estimates of present and future
conditions. Remote Sensing of Environment, 168, pp.301-315. doi:
10.1016/j.rse.2015.07.019.

Poggio, L., L.M. De Sousa, N.H. Batjes, G. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter.
2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial
uncertainty. Soil, 7(1), pp.217-240. doi: 10.5194/soil-7-217-2021.

Sanches, L.F., B. Guenet, C.C. Marinho, N. Barros, and F. de Assis Esteves. 2019. Global
regulation of methane emission from natural lakes. Scientific Reports, 9(1), pp.1-10. doi:
10.1038/s41598-018-36519-5.

Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. 2020. Daymet:
Daily Surface Weather Data on a 1-km Grid for North America, Version 4. ORNL
DAAC, Oak Ridge, Tennessee, USA. doi: 10.3334/ORNLDAAC/1840.

Wang, J.A., D. Sulla-Menashe, C.E. Woodcock, O. Sonnentag, R.F. Keeling, and M.A. Friedl.
2019. ABoVE: Landsat-derived Annual Dominant Land Cover Across ABoVE Core
Domain, 1984-2014. ORNL DAAC, Oak Ridge, Tennessee, USA. doi:
10.3334/ORNLDAAC/1691.

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