2.8 Model Evaluation Metrics

Lake simulations were compared to observations of lake water temperature using mean absolute error (MAE), root mean squared error (RMSE), and bias calculated for each water depth available in the observed data. Scenario data was compared to baseline simulations using the Z-score. Z-score is calculated as the mean of *Z's* for the scenario for a given model variable (e.g. water temperature at 1 m depth), where each *Z* is calculated as the difference of the model variable, x, and the mean of baseline, μ, divided by the standard deviation of the baseline, σ, following Eq. (1):

$$Z = \frac{x-\mu}{\sigma}. \quad (1)$$

2.0 Model Water Resolution

We investigated the effects of the vertical resolution of the water column on the accuracy of modeled water temperatures using LAKE. We investigated both Atqasuk lake, a shallow lake (~2.5m) where the mixed layer extends to the bottom of the lake, and Toolik lake, a deep lake (26m) where stratification can occur. For Atqasuk we tested resolutions of 1.0m, 0.1m, 0.065m, and 0.025m. For Toolik we tested resolutions of 1.0m, 0.65m, 0.5m, and 0.25m.

2.10 Model Soil Vertical Resolution

We investigated the effects of the vertical resolution of the soil column on the accuracy of modeled water temperatures using LAKE. We investigated Atqasuk lake using a 10m deep soil column and vertical resolutions of 2.0m, 1.0m, 0.2m, and 0.1m.

2.11 Model Temporal Resolution

We investigated the effects of temporal resolution of the meteorological data on the accuracy of modeled water temperatures using LAKE. We investigated Atqasuk lake and Fox Den lake using hourly and daily data. For Atqasuk hourly data was averaged to daily data. For Fox Den an hourly and a daily dataset were obtained from NOAA and NASA (see methods).

3.4 Sediment temperature and heat flux

Lake sediment temperatures for all three lakes were modeled to a depth of 10 m (Fig. C2). Observed sediment temperatures were not available for model validation. Shallow sediments (< 3 m) experience more warming in the shallow lakes (Atqasuk, Fox Den) as compared to the deeper lake (Toolik). All three lakes had near constant deeper (> 5 m) sediment temperatures over the several years of simulation. Downward heat flux at the bottom of the lake showed strong seasonal patterns of stronger flux during the thaw period and low fluxes during the frozen period (Fig. C3). The shallower lakes (Atqasuk, Fox Den) had larger magnitude fluxes (compared to the deeper lake, Toolik) due to turbulent mixing in the thaw season.

3.5 Water Column Vertical Resolution

For Atqasuk 0.3m water depth, 1m water resolution had the lowest modeled water temperature error (RMSE=6.87) although higher resolutions were very similar (Figure D1 & Table D1). For

Atqasuk 2.5m water depth, 0.1m, 0.065m, and 0.025m had the lowest modeled water temperature error (RMSE=1.44) although 1m resolution was slightly greater (RMSE=1.64, Figure D2 & Table D1). Vertical resolution of the water column had little effect on Atqasuk snow and ice layer thickness (Figure D3).
For Toolik 1m and 3m water depths, 0.65m water resolution had the lowest modeled water temperature error (RMSE=1.29, RMSE=2.02) although higher resolutions were very similar for Toolik (Figure D4, Figure D5, & Table D2). For Toolik 5m, 10m, and 19m water depths, 1m water resolution had the lowest modeled water temperature error (RMSE=2.01, RMSE=1.84, RMSE=1.78), while 0.65m, 0.5m, and 0.25m resolutions were similar (Figure D6, Figure D7, Figure D8, & Table D2). Vertical resolution had little effect on Toolik snow and ice layer thickness (Figure D9).

3.6 Soil Column Vertical Resolution

For Atqasuk 0.3m water depth, 1m soil resolution had the lowest modeled water temperature error (RMSE=7.15) although higher resolutions were very similar (Figure E1 & Table E1). For Atqasuk 2.5m water depth, all soil resolutions had the same modeled water temperature error (RMSE=1.44). Vertical resolution of the soil column had little effect on snow and ice layer thickness (Figure E3).

3.7 Temporal Resolution

For both Atqasuk and Fox Den, hourly and daily temporal meteorological data resolutions had similar water temperature errors, but daily resolution had a lower error value (Table F1). Temporal resolution had little effect on lake temperatures (Figures F1, F2 & F3; Table F1). Daily temporal resolution appears to reduce ice layer thickness (Figures F4 & F5). However, for Fox Den there is an apparent disagreement between the daily and hourly precipitation data. Observed data are unavailable to compare to modeled snow depth or ice thickness.

Table 1. Parameters of lake simulations.

Lake name	Atqasuk	Fox Den	Toolik
Latitude	70.452497	66.55877	68.631496
Longitude	-156.951984	-164.45670	-149.607404
Area, m2	2,732,050	17,861	1,492,898
Time span of integration	2013-08-12 to 2015-08-10	2009-06-10 to 2013-06-10	2013-05-17 to 2016-09-18
Time step, seconds	20	20	20
Maximal lake depth, m	2.54	1.5	26.0
Vertical grid water column	40 layers, refined near boundaries	40 layers, refined near boundaries	40 layers, refined near boundaries
Vertical resolution water column, m	0.0635	0.0375	0.65
Inital temperature at bottom of soil column, C	5.0	-4.0	4.0
Calibrated temperature at bottom of soil column, C	2.0	-1.0	3.6
Sediment (soil) type	Silt loam	Silt loam	Silt loam
Depth of sediment columns, m	10.0	10.0	10.0
Vertical resolution of soil column, m	1.0	1.0	1.0
Number of columns of sediments	5	5	5
Vertical grid in columns of sediments	10 layers, exponentially compacting towards sediments top	10 layers, exponentially compacting towards sediments top	10 layers, exponentially compacting towards sediments top
Albedo for visible radiation	0.06	0.06	0.06
Fraction of near-infrared energy in shortwave flux	0.35	0.35	0.35
Water surface emissivity	0.98	0.98	0.98
Extinction coefficient for shortwave radiation, m-1	0.58	0.58	0.58
Modal wind fetch, m	1000	1000	1000

Table 2. LAKE model performance for Atqasuk, Fox Den, and Toolik lake. Mean absolute error (MAE), root mean squared error (RMSE) and bias (Bias) for total simulation period (annual), frozen, and thawed seasons. See methods for details. The frozen season was defined as days with ice thickness > 0.0 m.

		Season								
		Annual			Frozen			Thawed		
Lake	Depth	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias
Atqasuk	0.3m	5.07	7.15	-4.74	6.16	8.09	-6.01	2.74	4.54	-2.02
	2.5m	1.3	1.44	-0.448	1.32	1.38	-0.399	1.23	1.55	-0.573
	All	**3.18**	**5.16**	**-2.59**	**3.74**	**5.8**	**-3.2**	**1.99**	**3.39**	**-1.3**
FoxDen	1.5m	1.51	2.43	-1.11	1.04	1.41	-0.713	2.4	3.66	-1.87
	All	**1.51**	**2.43**	**-1.11**	**1.04**	**1.41**	**-0.713**	**2.4**	**3.66**	**-1.87**
Toolik	1m	1.12	1.29	-1.01	-	-	-	1.12	1.29	-1.01
	3m	1.59	2.02	0.464	1.46	1.64	1.26	1.89	2.69	-1.39
	5m	1.78	2.24	0.964	1.73	1.98	1.35	1.87	2.7	0.15
	10m	1.9	2.35	1	1.65	1.99	1.13	2.42	2.96	0.742
	19m	1.77	2.19	1.07	1.65	1.99	1.32	2.01	2.56	0.527
	All	**1.63**	**2.08**	**0.666**	**1.55**	**1.85**	**1.21**	**1.78**	**2.41**	**-0.235**

Appendix B: Toolik Lake simulations comparing simulations with and without inflow and outflow

Table B1. Toolik model lake temperature error (RMSE) for inflow and no-inflow simulations.

Depth, m	Toolik inflow	Toolik no inflow
1	1.29	1.22
2	1.13	1
3	2.02	1.71
4	1.91	1.81
5	2.24	2.29
6	2.44	2.61
7	2.52	2.52
8	2.4	2.34
9	2.36	2.13
10	2.35	1.91
11	2.26	1.7
12	2.21	1.59
13	2.18	1.5
14	2.14	1.46
15	2.14	1.43
16	2.13	1.4
17	2.14	1.36
18	2.09	1.38
19	2.19	1.31
20	2.15	1.2
21	2.25	1.12
22	2.1	0.675
23	1.07	0.582
All	**2.18**	**1.73**

Table B2. Toolik model lake temperature errors for inflow and no-inflow simulations split by season.

		Season			
		Frozen		Thawed	
	Scenario	Inflow	No inflow	Inflow	No inflow
Metric	Depth, m				
RMSE	1	-	-	1.29	**1.22**
	3	1.64	**0.877**	2.69	**2.88**
	4	1.98	**1.12**	**2.70**	3.77
	10	1.99	**0.996**	**2.96**	3.09
	19	1.99	**0.931**	2.56	**1.89**
	All	1.85	**0.967**	**2.41**	2.59
MAE	1	-	-	1.12	1.03
	3	1.46	0.628	1.89	2.00
	4	1.73	0.865	1.87	2.81
	10	1.65	0.772	2.42	2.57
	19	1.65	0.863	2.01	1.80
	All	1.55	0.76	1.78	1.92
Bias	1	-	-	-1.01	-0.999
	3	1.26	-0.361	-1.39	-0.797
	4	1.35	-0.636	0.15	2.32
	10	1.13	-0.54	0.742	2.31
	19	1.32	-0.836	0.527	1.52
	All	1.21	-0.563	-0.235	0.725

Appendix C: Modeled water temperature, soil temperature, and heat flux.

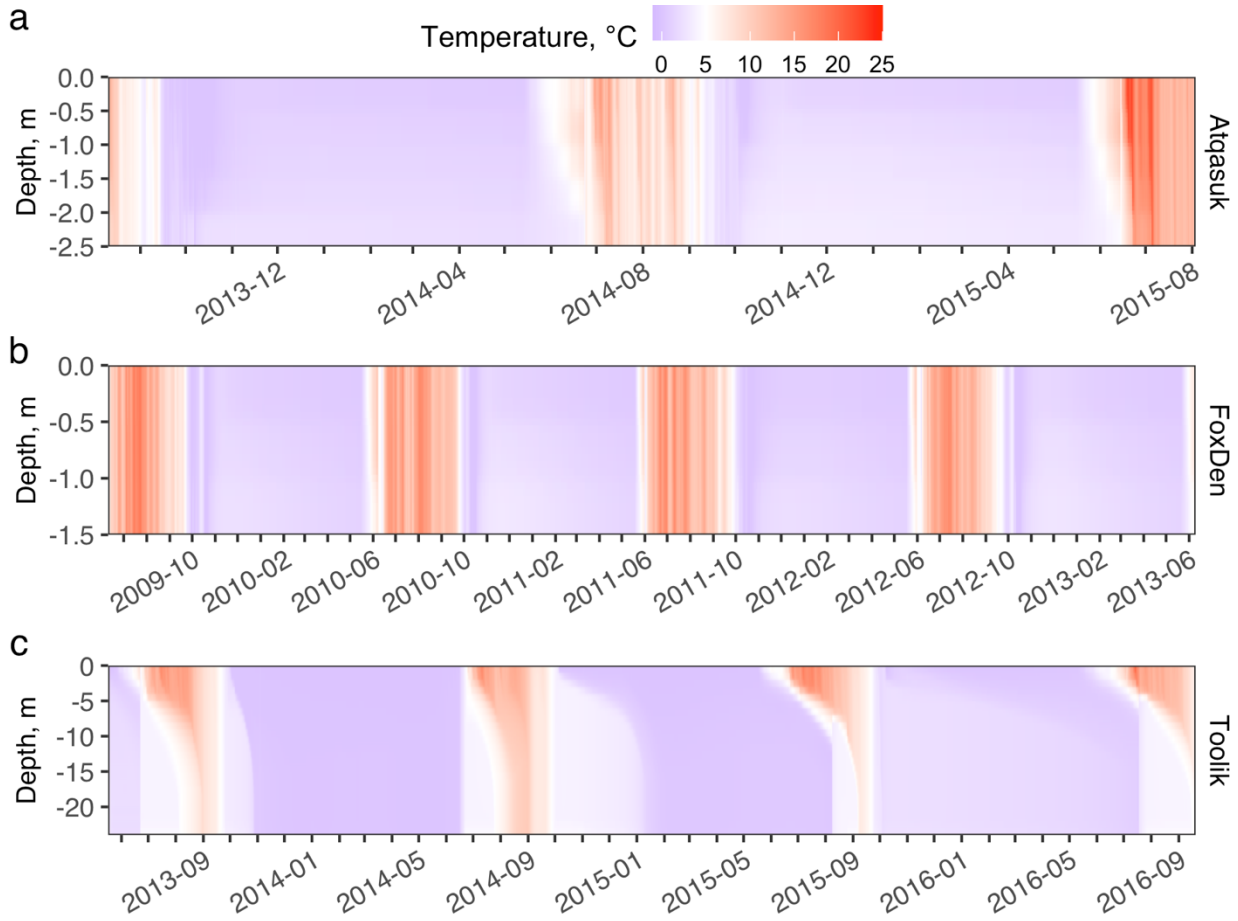

Figure C1. Lake water temperature over the simulation period for Atqasuk (a), Fox Den (b), and Toolik Lake (c).

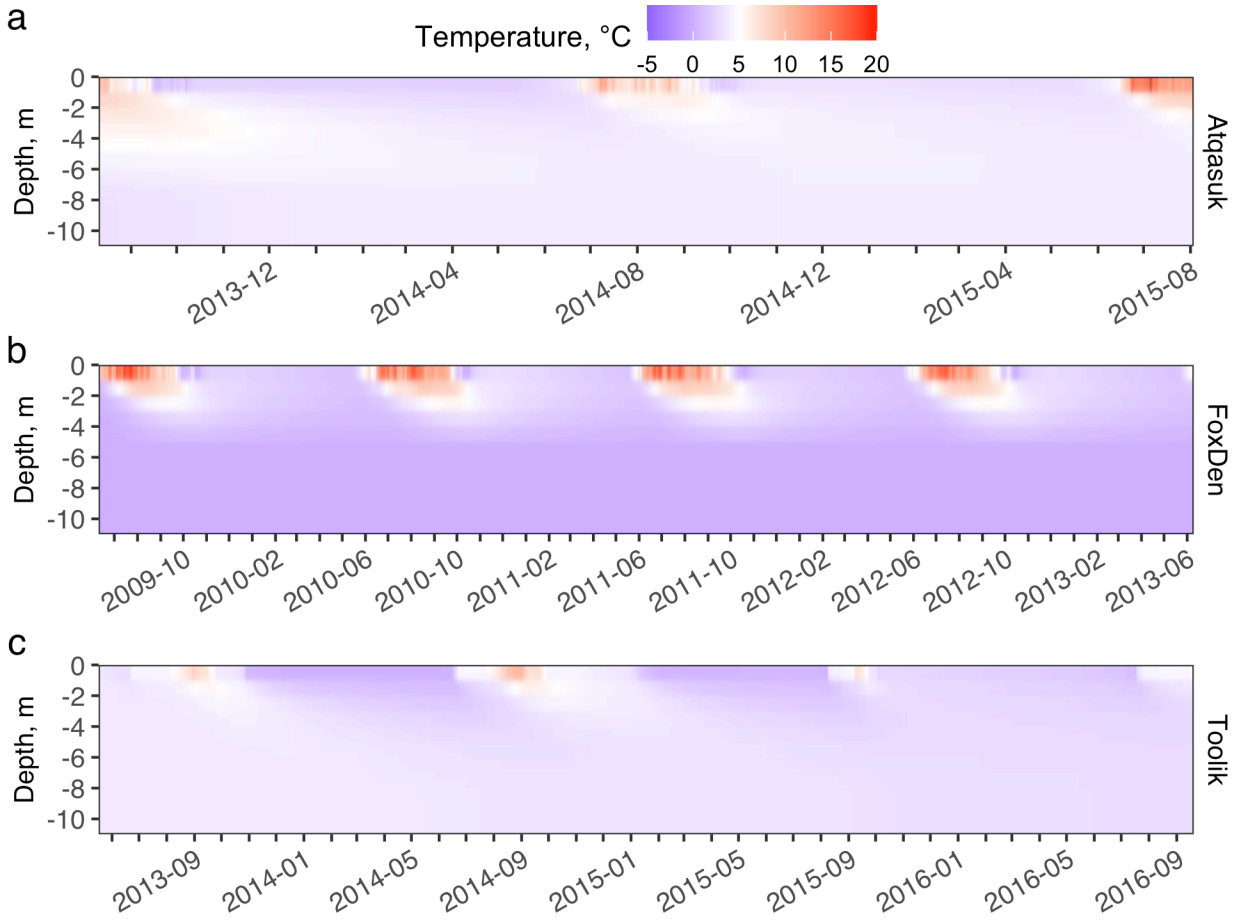

Figure C2. Lake sediment temperature over the simulation period for Atqasuk (a), Fox Den (b), and Toolik Lake (c) from 0 m to 10 m depth.

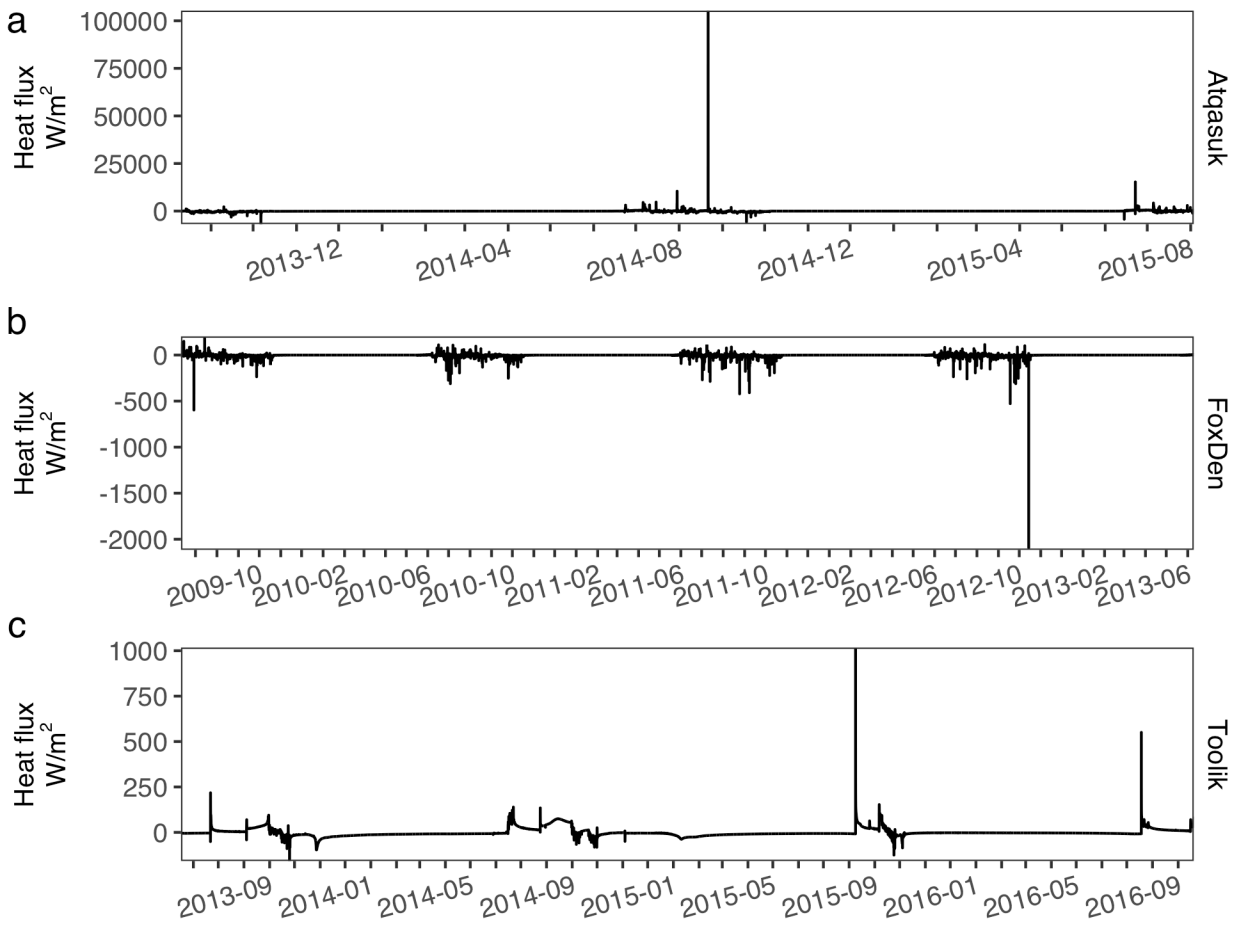

Figure C3. Downward heat flux at the bottom of the water column over the simulation period for Atqasuk (a), Fox Den (b), and Toolik Lake (c).

Appendix D:Water vertical resolution sensitivity analysis

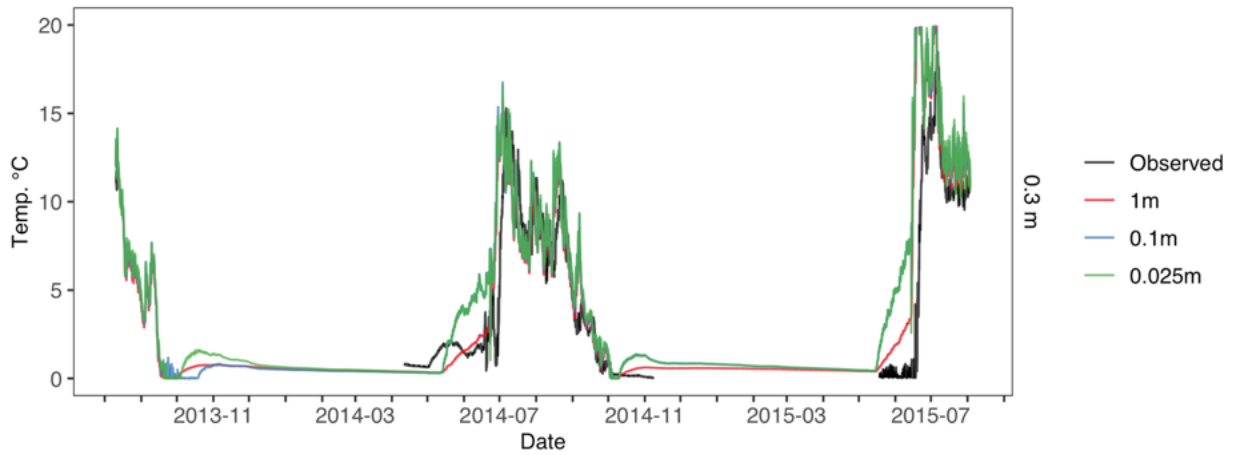

Fig. D1 Atqasuk 0.3m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.1m vertical resolution in blue, 0.025m vertical resolution in green. Note 0.1m and 0.025m overlap for most of the time series.

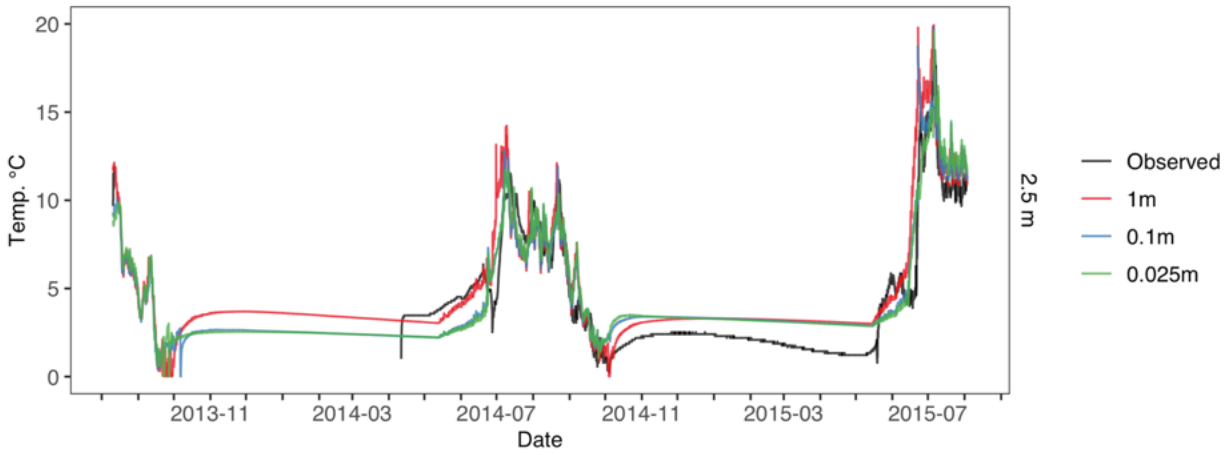

Fig. D2 Atqasuk 2.5m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.1m vertical resolution in blue, 0.025m vertical resolution in green. Note 0.1m and 0.025m overlap for most of the time series.

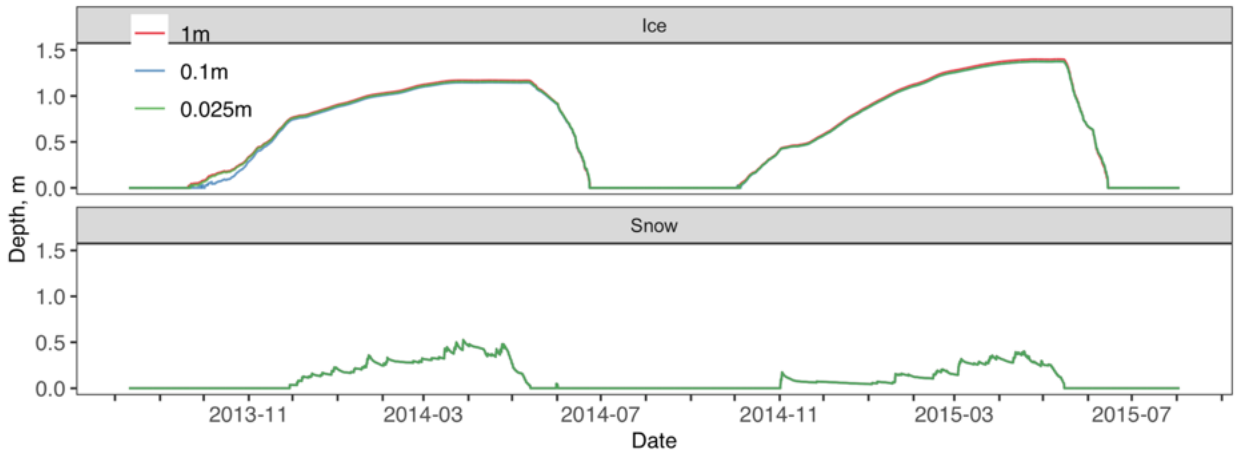

Fig. D3 Atqasuk ice and snow depth. No observed data available, 1m vertical resolution in red, 0.1m vertical resolution in blue, 0.025m vertical resolution in green. Note data overlaps for most of the time series.

Table D1. Atqasuk modeled water temperature errors calculated from measured water temperature time series at 0.3m and 2.5m. Bold rows are the lowest RMSE for a given water depth. Mean absolute error (MAE), root mean square error (RMSE), mean average error (Bias), percent Bias, mean absolute percent error (MAPE), mean z-score (zScore mean), and median z-score (zScore median).

Scenario	Depth (m)	Water Res. (m)	MAE	RMSE	Bias
Atgasuk_w0_025m	0.3	0.025m	5.07	7.16	-4.77
Atgasuk_w0_065m		0.065m	5.07	7.15	-4.74
Atgasuk_w0_1m		0.1m	5.06	7.14	-4.72
Atgasuk_w1m		**1.0m**	**4.58**	**6.87**	**-4.15**
Atgasuk_w0_025m	**2.5**	**0.025m**	**1.3**	**1.44**	**-0.481**
Atgasuk_w0_065m		**0.065m**	**1.29**	**1.44**	**-0.448**
Atgasuk_w0_1m		**0.1m**	**1.29**	**1.44**	**-0.491**
Atgasuk_w1m		1.0m	1.21	1.64	-0.834

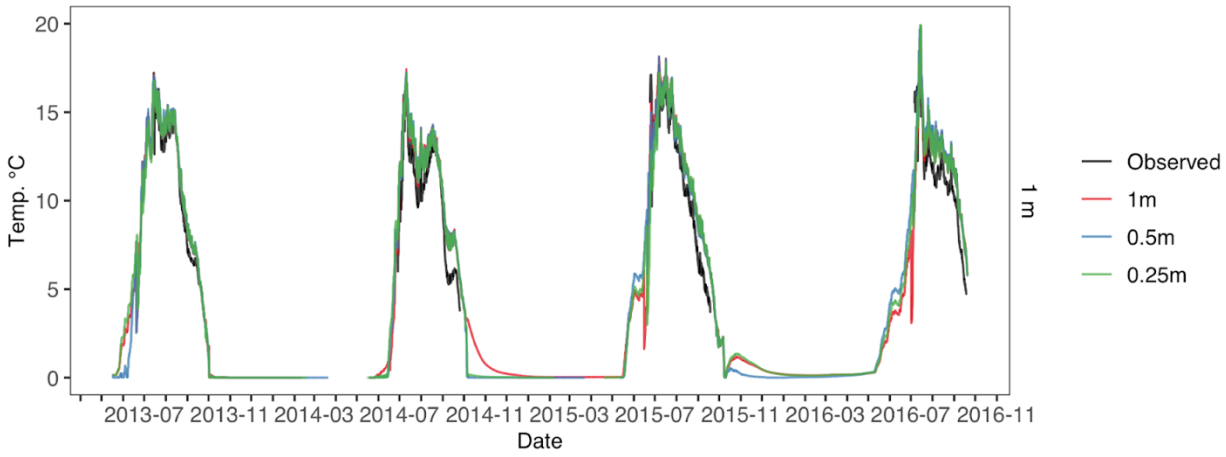

Fig. D4. Toolik 1m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.5m vertical resolution in blue, 0.25m vertical resolution in green.

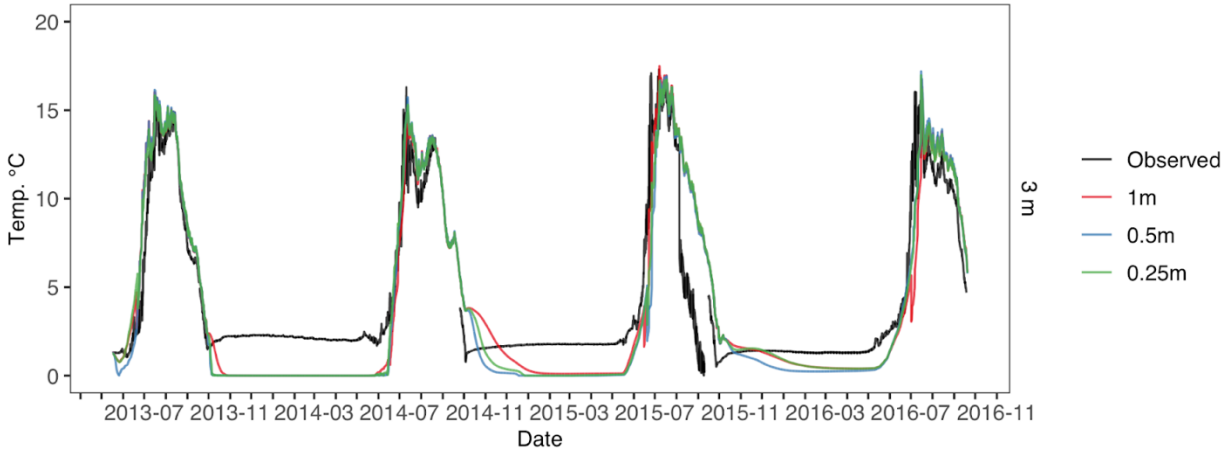

Fig. D5. Toolik 3m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.5m vertical resolution in blue, 0.25m vertical resolution in green.

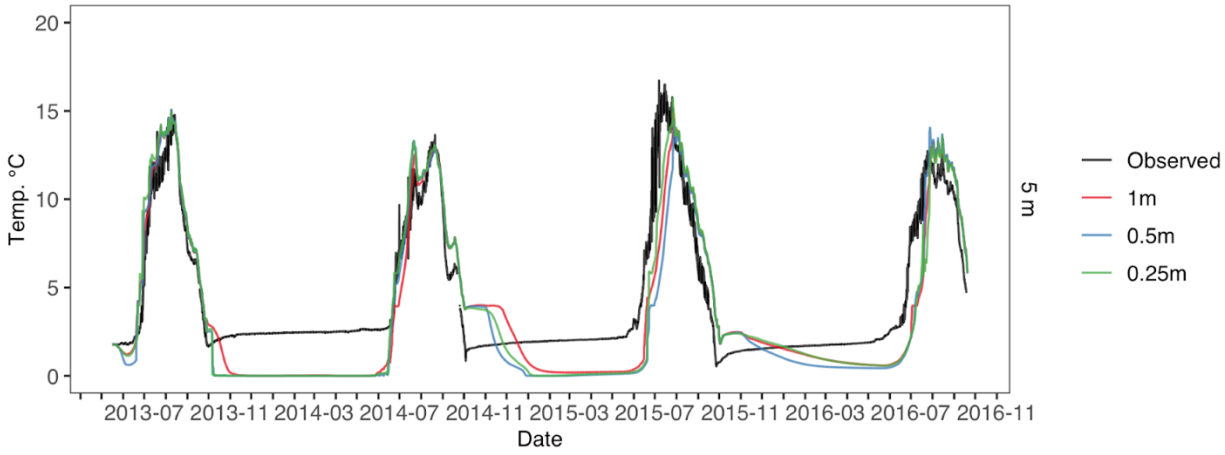

Fig. D6. Toolik 5m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.5m vertical resolution in blue, 0.25m vertical resolution in green.

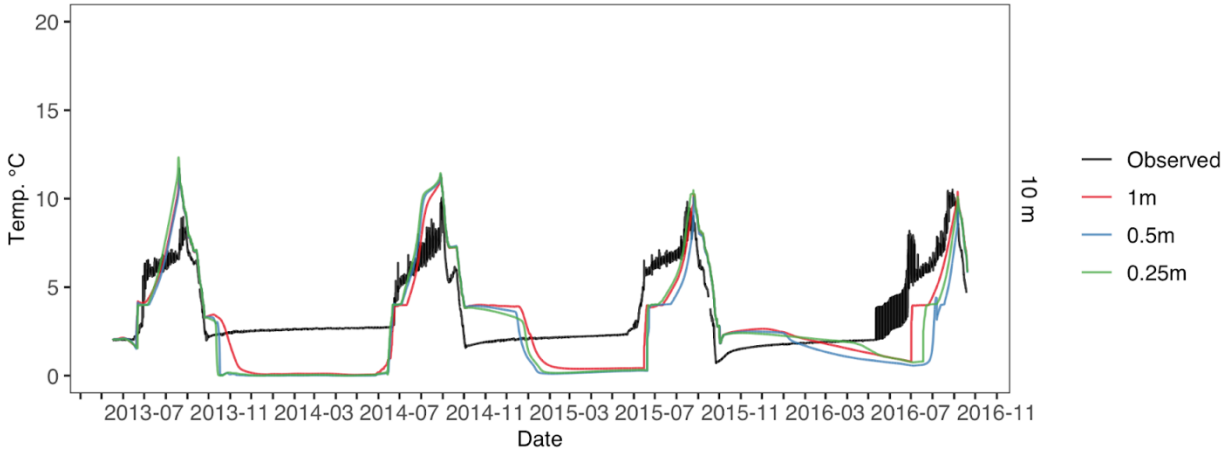

Fig. D7. Toolik 10m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.5m vertical resolution in blue, 0.25m vertical resolution in green.

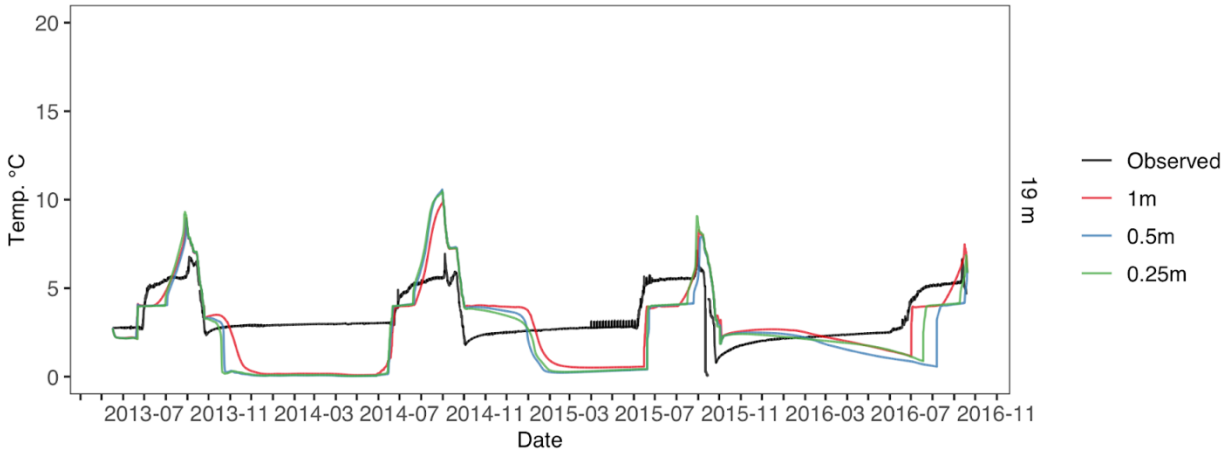

Fig. D8 Toolik 19m water temperature. Observed water temperature in black, 1m vertical resolution in red, 0.5m vertical resolution in blue, 0.25m vertical resolution in green.

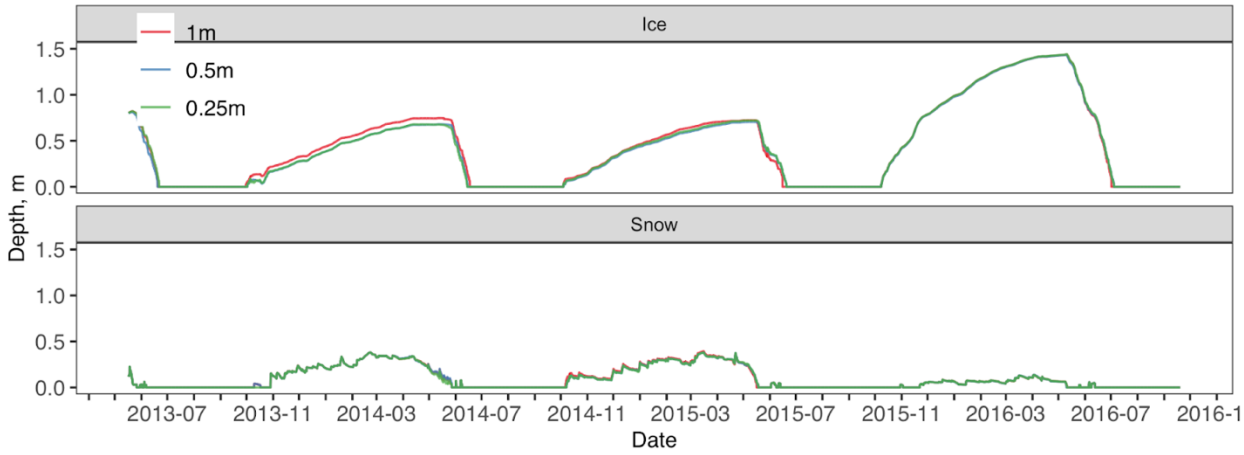

Fig. D9 Toolik ice and snow depth. No observed data available, 1m vertical resolution in red, 0.5m vertical resolution in blue, 0.25m vertical resolution in green.

Table D2. Toolik water temperature error calculated from measured water temperatures at multiple depths. Bold rows are the lowest RMSE for a given water depth. Mean absolute error (MAE), root mean square error (RMSE), mean average error (Bias), percent Bias, mean absolute percent error (MAPE), mean z-score (zScore mean), and median z-score (zScore median).

Scenario	Depth (m)	Water Res. (m)	MAE	RMSE	Bias
Too_w0_25m	1	0.25	1.19	1.42	-0.905
Too_w0_5m		0.5	1.18	1.36	-1
Too_w0_65m		**0.65**	**1.12**	**1.29**	**-1.01**
Too_w1m		1.0	1.15	1.35	-0.976
Too_w0_25m	3	0.25	1.61	2.08	0.469
Too_w0_5m		0.5	1.69	2.15	0.61
Too_w0_65m		**0.65**	**1.59**	**2.02**	**0.464**
Too_w1m		1.0	1.59	2.1	0.421
Too_w0_25m	**5**	0.25	**1.7**	**2.01**	**0.772**
Too_w0_5m		0.5	1.82	2.23	1
Too_w0_65m		0.65	1.78	2.24	0.964
Too_w1m		**1.0**	**1.67**	**2.01**	**0.76**
Too_w0_25m	10	0.25	1.74	2.06	0.69
Too_w0_5m		0.5	1.89	2.2	0.931
Too_w0_65m		0.65	1.9	2.35	1
Too_w1m		**1.0**	**1.6**	**1.84**	**0.57**
Too_w0_25m	19	0.25	1.63	1.98	0.808
Too_w0_5m		0.5	1.76	2.09	0.951
Too_w0_65m		0.65	1.77	2.19	1.07
Too_w1m		**1.0**	**1.49**	**1.78**	**0.594**

Appendix E. Soil vertical resolution sensitivity analysis

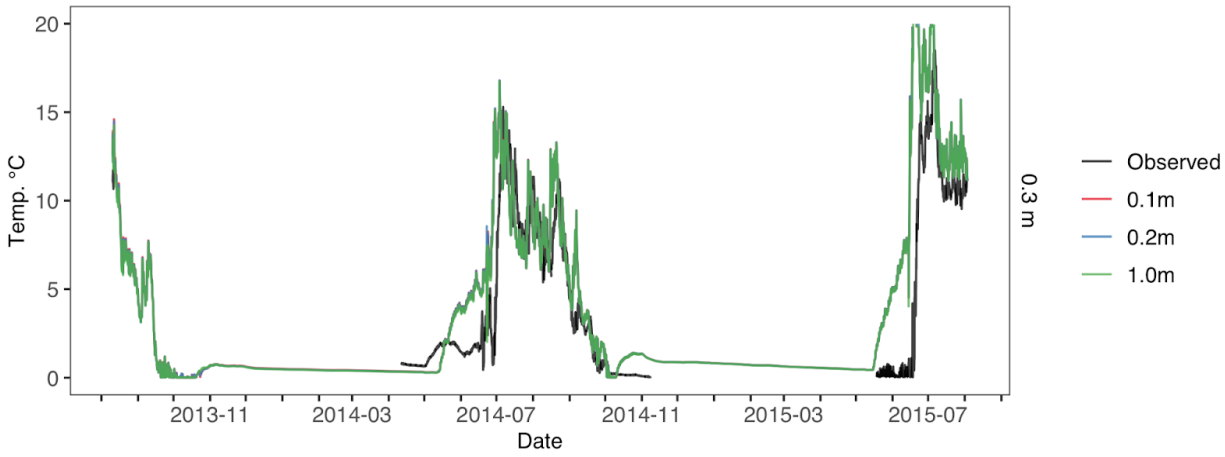

Fig. E1 Atqasuk 0.3m water temperature. Observed water temperature in black, 0.1m vertical resolution in red, 0.2m vertical resolution in blue, 1.0m vertical resolution in green.

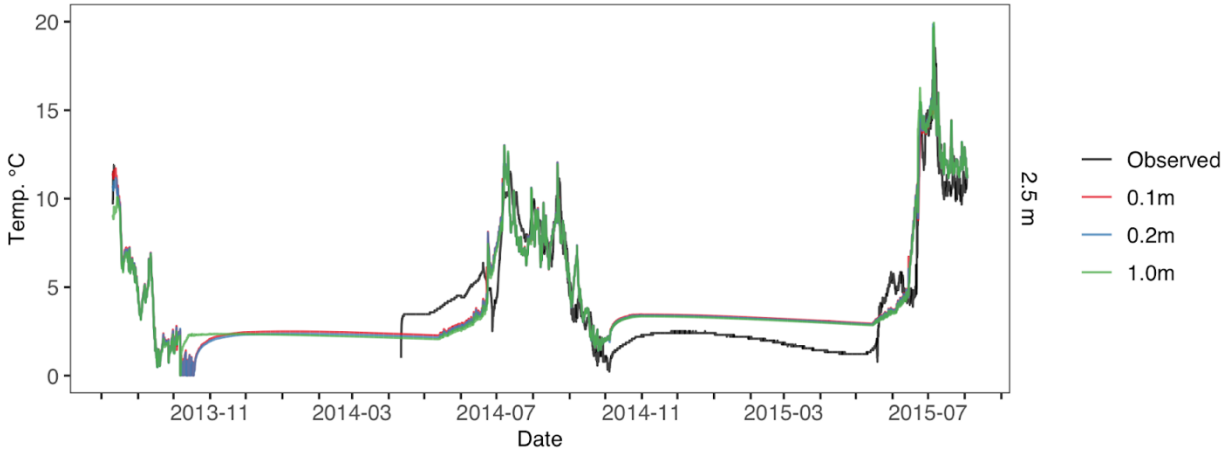

Fig. E2 Atqasuk 2.5m water temperature. Observed water temperature in black, 0.1m vertical resolution in red, 0.2m vertical resolution in blue, 1.0m vertical resolution in green.

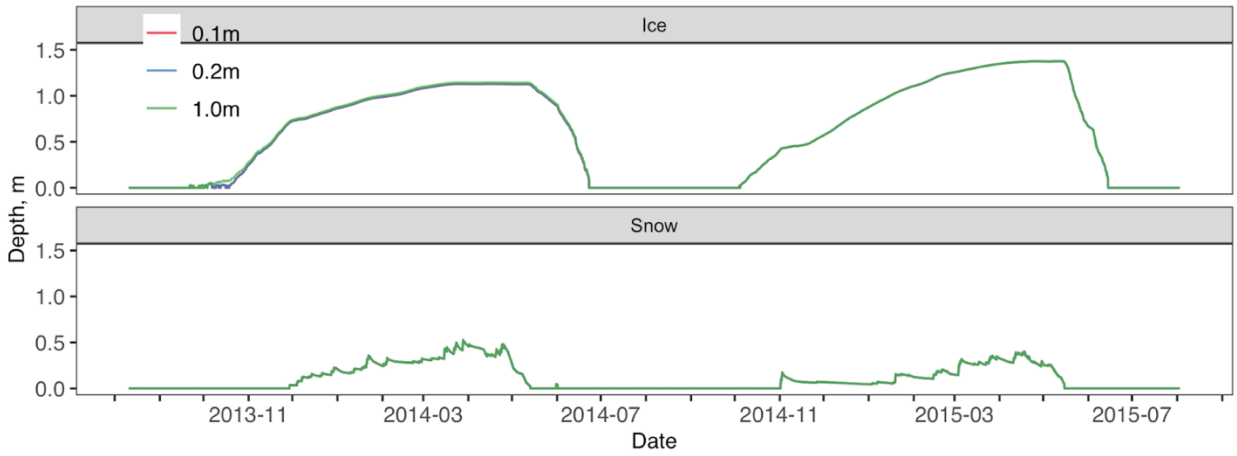

Fig. E3 Atqasuk ice and snow depth. No observed data available, 0.1m vertical resolution in red, 0.2m vertical resolution in blue, 1.0m vertical resolution in green.

Table E1. Atqasuk modeled water temperature errors calculated from measured water temperature time series at 0.3m and 2.5m. Bold rows are the lowest RMSE for a given water depth. Mean absolute error (MAE), root mean square error (RMSE), mean average error (Bias), percent Bias, mean absolute percent error (MAPE), mean z-score (zScore mean), and median z-score (zScore median).

Scenario	Depth (m)	Soil Res. (m)	MAE	RMSE	Bias
Atgasuk_s0_1m	0.3	0.1	5.09	7.17	-4.76
Atgasuk_s0_5m		0.5	5.08	7.16	-4.76
Atgasuk_s1m		**1.0**	**5.07**	**7.15**	**-4.74**
Atgasuk_s0_1m	**2.5**	**0.1**	**1.3**	**1.44**	**-0.541**
Atgasuk_s0_5m		**0.5**	**1.29**	**1.44**	**-0.5**
Atgasuk_s1m		**1.0**	**1.3**	**1.44**	**-0.448**

Appendix F. Temporal resolution sensitivity analysis

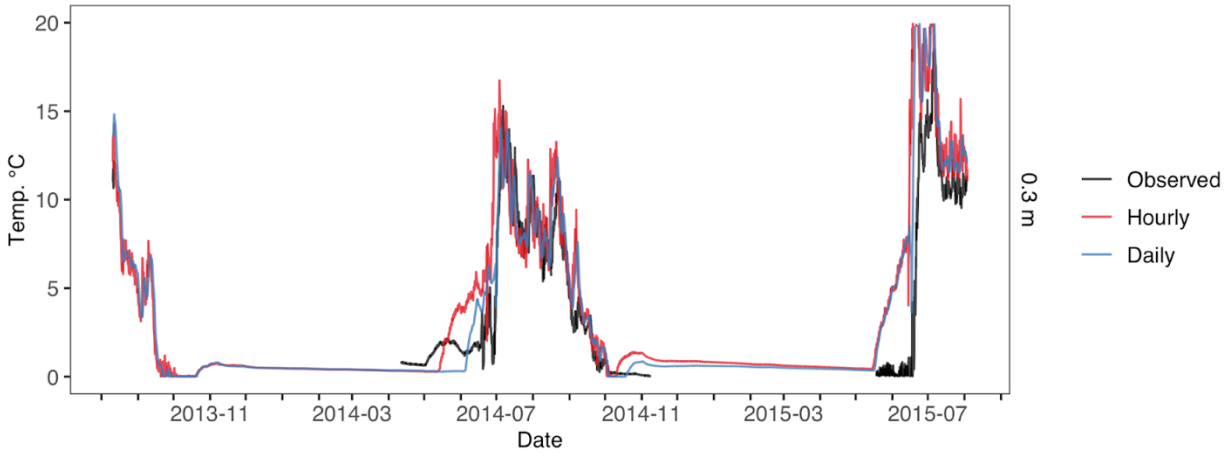

Fig. F1 Atqasuk 0.3m water temperature. Observed water temperature in black, hourly temporal resolution in red and daily temporal resolution in blue.

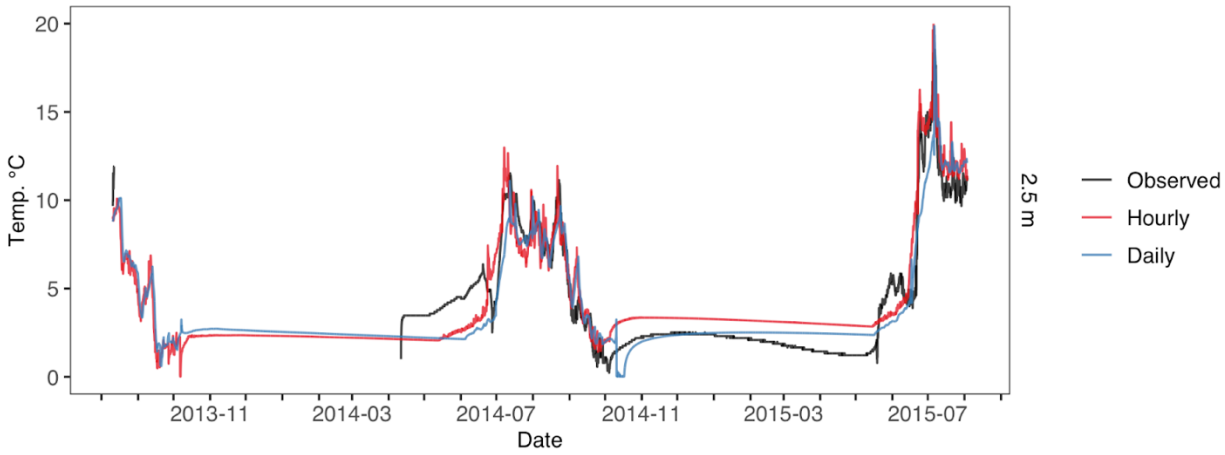

Fig. F2 Atqasuk 2.5m water temperature. Observed water temperature in black, hourly temporal resolution in red and daily temporal resolution in blue.

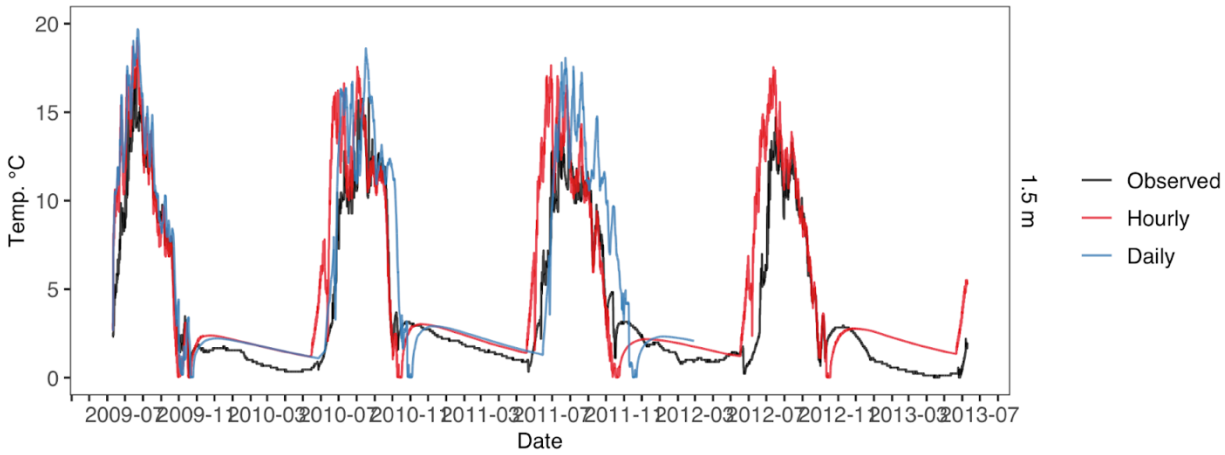

Fig. F3 Fox Den 1.5m water temperature. Observed water temperature in black, hourly temporal resolution in red and daily temporal resolution in blue.

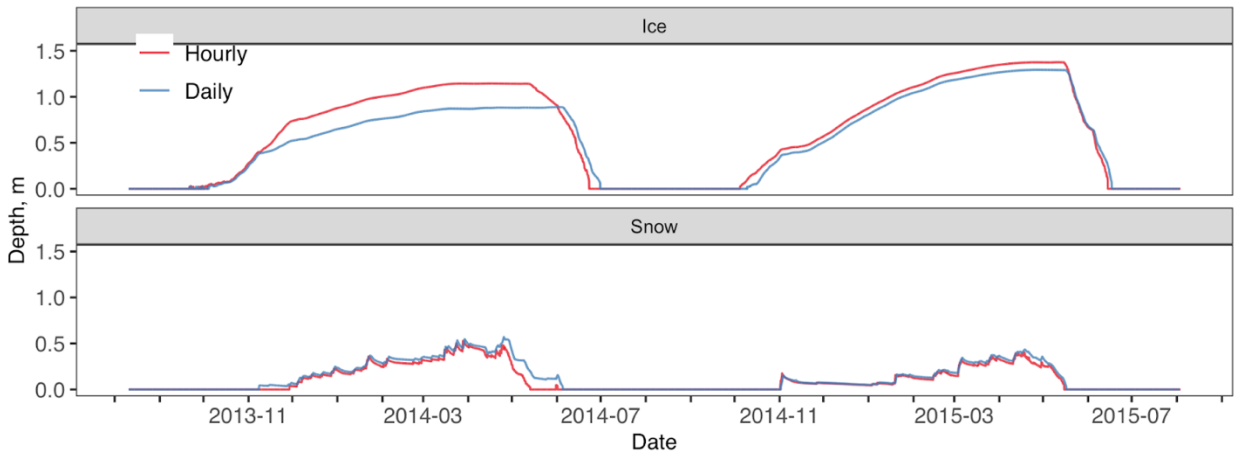

Fig. F4 Atqasuk ice and snow depth. No observed data available, hourly temporal resolution in red and daily temporal resolution in blue.

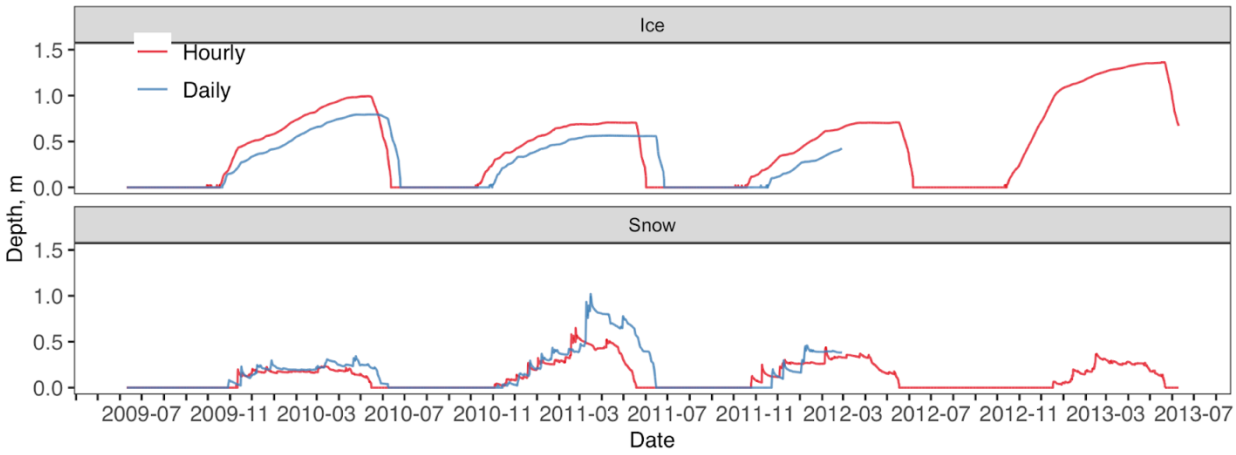

Fig. F5 Fox Den ice and snow depth. No observed data available, hourly temporal resolution in red and daily temporal resolution in blue.

Table F1. Atqasuk modeled water temperature errors calculated from measured water temperature time series at 0.3m and 2.5m. Bold rows are the lowest RMSE for a given water depth. Mean absolute error (MAE), root mean square error (RMSE), mean average error (Bias), percent Bias, mean absolute percent error (MAPE), mean z-score (zScore mean), and median z-score (zScore median).

Site	Temporal Res.	Depth, m	MAE	RMSE	Bias
Atqasuk	Hourly	0.3	5.07	7.15	-4.74
	Daily		**4.7**	**6.83**	**-4.33**
	Hourly	2.5	1.29	1.44	-0.448
	Daily		**1.02**	**1.31**	**0.24**
Fox Den	Hourly	1.5	1.69	2.47	-1.27
	Daily	**1.5**	**1.45**	**2.41**	**-0.953**