Supplementary

Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning

Helen Weierbach [1,†], Aranildo R. Lima [2], Jared Willard[1,4], Valerie C. Hendrix [3], Danielle S. Christianson [3], Michaelle Lubich[1,5], and Charuleka Varadharajan [1*]

Citation: Weierbach, H.; Lima, A.R.; M; Willard, J; Hendrix, V.C.; Christianson, D.S.; Lubich, M.; and Varadharajan, C. Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning. *Water* **2022**, *14*, 1032. https://doi.org/10.3390/w14071032

1 Earth and Environmental Sciences Area, Lawrence Berkeley National Lab
2 Aquatic Informatics
3 Computing Sciences Area, Lawrence Berkeley National Lab
4 University of Minnesota
5 University of California, Berkeley
* Correspondence: cvaradharajan@lbl.gov
† Current address: 1 Cyclotron Rd, Berkeley, CA 94720

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Submitted to *Water* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/ 4.0/).

1. Station Information

Table S1 presents station metadata for each of the 78 stations used in this study. Latitude (Lat, °N), longitude (Long, °E) and elevation (Elev, m) were obtained using BASIN-3D. Drainage area (A, m^2), number of major dams (N Dams), and (R)/non-reference (NR) classes were acquired from the GAGES-II dataset (stations with 'NA' values are not present in GAGES-II). Mean values of air temperature ($\overline{T_a}$, °C), stream temperature ($\overline{T_s}$, °C) and discharge ($\overline{Q}$, m^3/s) were calculated using each station's full data record.

Table S1: Information on stations in the HUC02 and HUC17 regions used in this study.

Station ID	Station Name	Class	Date Range	N Months	Lat[1]	Lon[1]	Elev[2]	A[2]	N Dams	$\overline{T_a}$[1]	$\overline{T_s}$[1]	$\overline{Q}$[1]
			HUC 02									
01362230	DIVERSION FROM SCHOHARIE RESERVOIR NY	NA	1997-2020	268	42.11	-74.36	294	NA	NA	8.49	9.07	0.17
01362500	ESOPUS CREEK AT COLDBROOK NY	NR	1996-2020	284	42.01	-74.27	189	493	0	9.28	9.9	0.58
01388000	Ramapo River at Pompton Lakes NJ	NR	1989-2020	375	40.99	-74.28	52	421	1	11.32	13.49	0.24
01417500	EAST BRANCH DELAWARE RIVER AT HARVARD NY	NR	1980-2020	485	42.02	-75.12	307	1184	1	7.19	8.92	0.35
01420500	BEAVER KILL AT COOKS FALLS NY	NR	1987-2020	317	41.95	-74.98	352	627	0	7.7	9.59	0.5
01421000	EAST BRANCH DELAWARE RIVER AT FISHS EDDY NY	NR	1980-2020	473	41.97	-75.18	291	2030	1	8.32	10.04	0.99
01425000	WEST BRANCH DELAWARE RIVER AT STILESVILLE NY	NR	1980-2020	469	42.07	-75.40	302	1181	1	8.18	7.21	0.49
01426500	WEST BRANCH DELAWARE RIVER AT HALE EDDY NY	NR	1980-2020	490	42.00	-75.38	288	1539	2	7.99	8.09	0.71
01427510	DELAWARE RIVER AT CALLICOON NY	NR	1980-2020	408	41.76	-75.06	224	4725	3	8.73	10.63	2.47
01428500	DELAWARE R ABOVE LACKAWAXEN R NEAR BARRYVILLE NY	NR	1980-2020	486	41.51	-74.99	183	5243	3	9.06	11.45	2.69
01436690	NEVERSINK RIVER AT BRIDGEVILLE NY	NR	1992-2020	328	41.64	-74.62	322	443	2	8.42	9.63	0.23
01447500	Lehigh River at Stoddartsville, PA	NR	1980-2020	286	41.13	-75.63	446	240	0	8.12	9.95	0.16
01447720	Tobyhanna Creek near Blakeslee, PA	NR	1980-2020	319	41.08	-75.61	461	308	1	7.83	10.21	0.22
01449360	Pohopoco Creek at Kresgeville, PA	NR	1980-2020	332	40.90	-75.50	201	130	0	9.91	10.19	0.08
01463500	Delaware River at Trenton NJ	NR	1980-2020	490	40.22	-74.78	0	17580	35	12.18	13.36	9.97
01466500	McDonalds Branch in Byrne State Forest NJ	R	1980-2020	260	39.88	-74.51	36	5	0	12.28	10.7	0.0
01480400	Birch Run near Wagontown, PA	NR	1996-2020	294	40.03	-75.84	154	12	1	11.66	11.88	0.01
01480500	West Branch Brandywine Creek at Coatesville, PA	NR	1995-2020	311	39.99	-75.83	93	119	1	12.0	12.32	0.05
01480617	West Branch Brandywine Creek at Modena, PA	NR	1981-2020	410	39.96	-75.80	80	144	1	13.53	14.85	0.06
01480870	East Branch Brandywine Creek below Downingtown...	NR	1980-2020	424	39.97	-75.67	59	232	2	13.89	14.73	0.12
01481000	Brandywine Creek at Chadds Ford, PA	NR	1980-2020	431	39.87	-75.59	46	755	4	14.11	14.98	0.35
01646500	POTOMAC RIVER NEAR WASH, DC LITTLE FALLS PUMP STA	NR	1988-2019	365	38.95	-77.13	11	29952	123	13.46	15.29	10.19
02011400	JACKSON RIVER NEAR BACOVA, VA	R	1980-2020	439	38.04	-79.88	500	408	0	11.16	12.85	0.14
02011500	BACK CREEK NEAR MOUNTAIN GROVE, VA	NR	1980-2020	443	38.07	-79.90	519	348	2	10.85	12.68	0.15
			HUC 17									
12056500	NF SKOKOMISH R BL STAIRCASE RPDS NR HOODSPORT, WA	R	1989-2020	272	47.51	-123.33	232	147	0	9.45	7.13	0.41
12115000	CEDAR RIVER NEAR CEDAR FALLS, WA	R	1997-2020	267	47.37	-121.63	475	103	0	7.5	6.65	0.2
12116500	CEDAR RIVER AT CEDAR FALLS, WA	NR	1999-2020	259	47.42	-121.79	275	220	1	9.67	9.16	0.26
12117600	CEDAR RIVER BELOW DIVERSION NEAR LANDSBURG, WA	NR	2001-2020	236	47.38	-121.98	149	346	1	10.67	9.23	0.47
12119000	CEDAR RIVER AT RENTON, WA	NR	1992-2020	321	47.48	-122.20	5	448	1	11.83	10.48	0.55
12138160	SULTAN RIVER BELOW POWERPLANT NEAR SULTAN, WA	NR	1995-2020	273	47.91	-121.82	81	244	2	10.31	7.9	0.67
12147500	NORTH FORK TOLT RIVER NEAR CARNATION, WA	R	1995-2016	244	47.71	-121.79	183	103	0	9.73	7.72	0.29
12147600	SOUTH FORK TOLT RIVER NEAR INDEX, WA	R	1994-2020	303	47.71	-121.60	564	14	0	8.49	6.15	0.05
12148000	SOUTH FORK TOLT RIVER NEAR CARNATION, WA	NR	1994-2020	315	47.69	-121.71	396	52	0	9.34	7.79	0.07
12148300	SF TOLT RIVER BL REGULATING BASIN NR CARNATION...	NR	1994-2020	305	47.70	-121.79	204	77	1	10.12	8.27	0.12
12181000	SKAGIT RIVER AT MARBLEMOUNT, WA	NA	1986-2020	374	48.53	-121.43	93	NA	NA	10.09	7.73	5.04
12323770	Warm Springs Creek at Warm Springs MT	NR	2000-2020	246	46.18	-112.79	1466	414	1	6.05	6.76	0.05
12362500	S F Flathead River nr Columbia Falls MT	NR	1999-2020	259	48.36	-114.04	927	4318	1	6.4	7.28	2.8
12363000	Flathead River at Columbia Falls MT	NA	1980-2020	479	48.36	-114.18	909	NA	NA	6.34	6.9	7.61
12398600	PEND OREILLE RIVER AT INTERNATIONAL BOUNDARY	NA	2000-2020	243	49.00	-117.35	518	NA	NA	8.16	10.47	19.9
13334300	SNAKE RIVER NEAR ANATONE, WA	NA	1994-2020	303	46.10	-116.98	246	NA	NA	12.69	11.74	26.56
13340000	CLEARWATER RIVER AT OROFINO ID	R	1993-2020	255	46.48	-116.26	302	14269	0	10.52	9.4	6.71
13340600	NF CLEARWATER RIVER NR CANYON RANGER STATION ID	R	1999-2020	251	46.84	-115.62	506	3355	0	8.41	7.61	2.58
13342500	CLEARWATER RIVER AT SPALDING ID	NR	1999-2020	254	46.45	-116.83	235	24065	3	11.45	8.32	11.71
14138720	BULL RUN RIVER AT LOWER FLUME NR BRIGHTWOOD, OR	NR	1995-2020	304	45.47	-121.87	866	13	1	7.62	6.11	0.02
14138850	BULL RUN RIVER NEAR MULTNOMAH FALLS, OR	NR	1980-2020	489	45.50	-122.01	329	124	1	9.98	7.51	0.32
14138870	FIR CREEK NEAR BRIGHTWOOD, OR	R	1980-2020	492	45.48	-122.03	451	14	0	8.87	7.14	0.03
14138900	NORTH FORK BULL RUN RIVER NEAR MULTNOMAH FALLS...	R	1980-2020	489	45.49	-122.04	323	22	0	10.13	6.94	0.06
14139800	SOUTH FORK BULL RUN RIVER NEAR BULL RUN, OR	R	1980-2020	492	45.44	-122.11	302	41	0	9.75	7.84	0.09
14145500	MF WILLAMETTE RIVER ABV SALT CRK, NEAR OAKRIDG...	NR	1980-2020	348	43.72	-122.44	368	1017	1	11.18	8.76	0.92
14150000	MIDDLE FORK WILLAMETTE RIVER NEAR DEXTER, OR	NR	1980-2020	427	43.95	-122.84	181	2605	3	11.96	11.13	2.32
14151000	FALL CREEK BLW WINBERRY CREEK, NEAR FALL CREEK...	NR	1985-2020	370	43.94	-122.77	194	481	1	12.01	10.06	0.44
14152000	MIDDLE FORK WILLAMETTE RIVER AT JASPER, OR	NR	1986-2020	255	44.00	-122.91	156	3491	4	12.18	11.35	2.98
14159200	SO FK MCKENZIE RIVER ABV COUGAR LAKE NR RAINBO...	R	1985-2020	260	44.05	-122.22	521	414	0	10.15	7.27	0.5
14159500	SOUTH FORK MCKENZIE RIVER NEAR RAINBOW, OR	NR	1985-2020	403	44.14	-122.25	377	540	1	10.9	8.32	0.66
14162200	BLUE RIVER AT BLUE RIVER, OR	NR	1985-2020	396	44.16	-122.33	322	228	1	11.04	8.39	0.36
14162500	MCKENZIE RIVER NEAR VIDA, OR	NR	1980-2020	311	44.12	-122.47	261	2403	4	11.18	8.69	3.22
14166000	WILLAMETTE RIVER AT HARRISBURG, OR	NR	1985-2020	258	44.27	-123.17	88	8895	10	11.72	11.82	8.41
14178000	NO SANTIAM R BLW BOULDER CRK, NR DETROIT, OR	NR	1985-2020	289	44.71	-122.10	485	558	0	9.74	7.51	0.8
14179000	BREITENBUSH R ABV FRENCH CR NR DETROIT, OR.	R	1985-2020	290	44.75	-122.13	480	272	0	9.46	7.48	0.44
14180300	BLOWOUT CREEK NEAR DETROIT, OR	R	1998-2020	267	44.65	-122.13	561	67	0	9.43	8.07	0.09
14181500	NORTH SANTIAM RIVER AT NIAGARA, OR	NR	1985-2020	397	44.75	-122.30	333	1171	2	9.61	8.15	1.8
14187200	SOUTH SANTIAM RIVER NEAR FOSTER, OR	NR	1985-2020	361	44.41	-122.69	162	1445	2	11.64	9.51	2.23
14206950	FANNO CREEK AT DURHAM, OR	NR	1993-2020	249	45.40	-122.75	35	81	0	12.28	12.64	0.04
14211400	JOHNSON CREEK AT REGNER ROAD, AT GRESHAM, OR	NR	1999-2020	264	45.49	-122.42	93	40	0	11.83	11.72	0.03
14211499	KELLEY CREEK AT SE 159TH DRIVE AT PORTLAND, OR	NR	2000-2020	248	45.48	-122.50	75	12	0	11.9	11.36	0.01
14211500	JOHNSON CREEK AT SYCAMORE, OR	NR	1998-2020	273	45.48	-122.51	70	68	0	12.06	12.1	0.04
14211550	JOHNSON CREEK AT MILWAUKIE, OR	NR	1998-2020	272	45.45	-122.64	0	137	0	12.57	12.71	0.06
14316495	BOULDER CREEK NEAR TOKETEE FALLS, OR	R	1999-2016	212	43.30	-122.53	500	79	0	11.02	9.63	0.08
14318000	LITTLE RIVER AT PEEL, OR	R	1999-2020	258	43.25	-123.03	252	460	1	12.51	11.08	0.33
14330000	ROGUE RIVER BELOW PROSPECT, OR	NR	1985-2020	419	42.73	-122.52	599	986	2	11.36	6.99	1.1
14337600	ROGUE RIVER NEAR MCLEOD, OR	NR	1985-2020	242	42.66	-122.72	454	2438	4	12.53	8.52	1.59
14338000	ELK CREEK NEAR TRAIL, OR	NR	1985-2020	420	42.68	-122.74	455	336	1	12.35	12.04	0.15
14339000	ROGUE RIVER AT DODGE BRIDGE, NEAR EAGLE POINT, OR	NR	1985-2020	416	42.52	-122.84	388	3155	6	12.69	9.16	1.86
14359000	ROGUE RIVER AT RAYGOLD NEAR CENTRAL POINT, OR	NR	1996-2020	288	42.44	-122.99	342	5310	17	12.75	10.61	2.26
14362000	APPLEGATE RIVER NEAR COPPER, OR	NR	1985-2020	419	42.06	-123.11	533	580	1	12.3	9.68	0.31
14366000	APPLEGATE RIVER NEAR APPLEGATE, OR	NR	1986-2020	419	42.24	-123.14	392	1253	1	12.67	11.12	0.38
14369500	APPLEGATE RIVER NEAR WILDERVILLE, OR	NR	1985-2020	414	42.35	-123.41	289	1810	1	13.01	12.82	0.5
14372300	ROGUE RIVER NEAR AGNESS, OR	NR	1986-2020	331	42.58	-124.06	35	10196	19	13.08	13.22	4.56

[1] Latitude, longitude, $\overline{T_a}$, $\overline{T_s}$, and $\overline{Q}$ were rounded to the nearest two decimal places.
[2] Elevation and drainage area were rounded to the nearest integer.

2. Additional Results

Additional model metrics for optimal model configurations for the SS, regional, and ensemble PUB scenarios are presented in figures S1,S3, and S5 where MAE, NSE, and Bias are represented by the following equations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{1}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{t=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

$$Bias = \overline{\mathbf{y} - \hat{\mathbf{y}}} \tag{3}$$

Additionally, error metrics for the temporal single region (SR) and predictions in unmonitored basins (PUB) scenarios are presented in Table S2.

Table S2: Root mean squared error (RMSE) metrics for different model configurations tested within the single region (SR) and PUB scenarios (described in Table 1). For each configuration, we present the number of training stations (N) with RMSE at Q1 (25th percentile), μ/Q2, Q3 (75th percentile) and median performance at extremes (Ext). The optimal configuration for each scenario used for further error metrics is presented in bold.

Configuration Name	**N**	**MLR**				**SVR**				**XGB**			
		Q1	μ/Q2	Q3	Ext[1]	Q1	μ/Q2	Q3	Ext[1]	Q1	μ/Q2	Q3	Ext[1]
Temporal Predictions–Regional (SR)													
02_noAtt_HPO	24	1.12	1.43/1.24	1.44	1.85	0.84	1.24/0.92	1.34	1.17	1.12	1.29/1.22	1.32	1.17
02_Att_HPO	24	1.12	1.43/1.24	1.44	1.85	0.76	1.17/0.86	1.24	1.07	1.00	1.19/1.10	1.25	1.07
02_noAtt_noHPO	24	1.12	1.43/1.24	1.44	1.85	0.94	1.26/1.01	1.30	1.42	1.04	1.20/1.17	1.30	1.42
02_Att_noHPO	**24**	1.12	1.43/1.24	1.44	1.85	**0.83**	**1.15/0.93**	**1.17**	**1.33**	1.01	1.18/1.10	1.27	1.33
02_Att_Drain_noHPO	23	1.09	1.34/1.21	1.41	1.66	0.78	1.00/0.91	1.09	1.34	0.97	1.09/1.06	1.17	1.34
17_noAtt_HPO	54	0.76	1.24/1.25	1.45	1.40	0.71	1.09/1.02	1.30	1.04	0.81	1.18/1.10	1.43	1.04
17_Att_HPO	54	0.76	1.24/1.25	1.45	1.40	0.72	1.05/0.94	1.25	1.03	0.80	1.15/1.08	1.42	1.03
17_noAtt_noHPO	54	0.76	1.24/1.25	1.45	1.40	0.66	1.05/0.98	1.30	1.01	0.78	1.04/1.02	1.27	1.01
17_Att_noHPO	**54**	0.76	1.24/1.25	1.45	1.40	**0.59**	**0.98/0.92**	**1.23**	**0.98**	0.65	0.97/0.94	1.23	0.98
17_Att_Drain_noHPO	50	0.74	1.22/1.20	1.41	1.35	0.59	0.93/0.81	1.15	0.97	0.69	0.98/0.89	1.23	0.97
Spatial Predictions–PUB													
PUB_02	**24**	1.16	1.45/1.32	1.49	1.72	0.75	1.04/0.89	1.17	1.13	**0.53**	**0.63/0.60**	**0.66**	**1.13**
PUB_17	**54**	0.79	1.23/1.17	1.47	1.38	0.64	0.97/0.88	1.13	0.98	**0.44**	**0.64/0.63**	**0.72**	**0.98**

[1] Extremes defined as T_s <0.05 percentile or >0.95 percentile

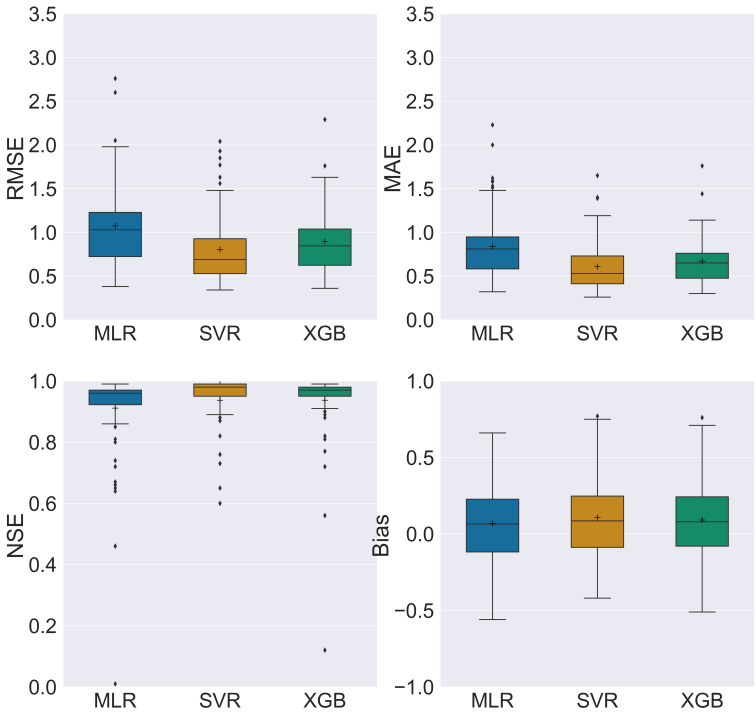

Figure S1. Single station RMSE (a) MAE (b) NSE (c) and Bias (d) distributions averaged over the test period for the SS_noHPO scenario. Means for all error metrics are denoted by (+).

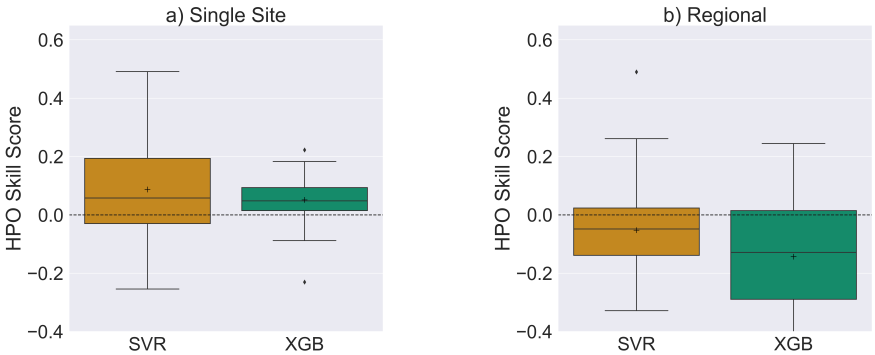

Figure S2. HPO skill scores for best case (a) single station scenario and (b) regional scenario.

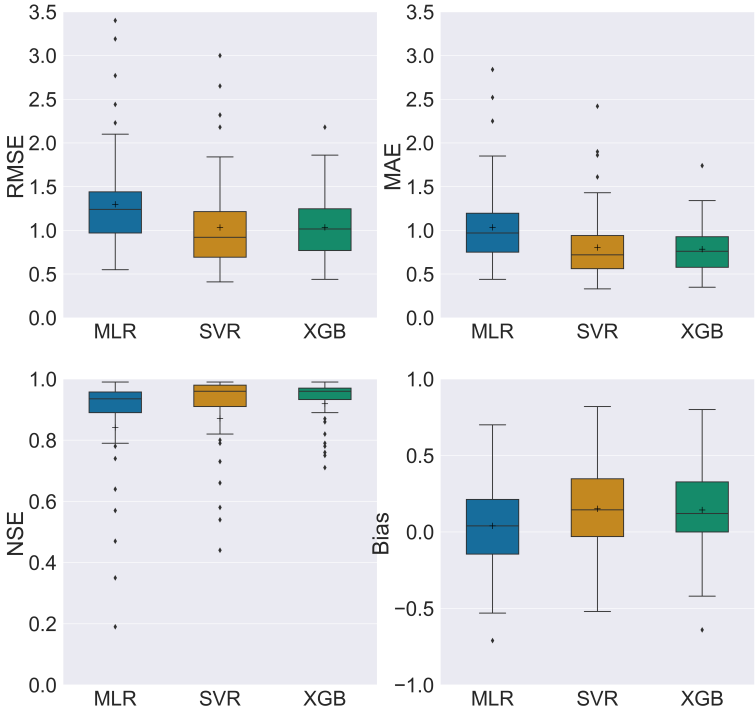

Figure S3. Regional RMSE (a) MAE (b) NSE (c) and Bias (d) distributions averaged over the test period for all 78 stations (02_17_SR_Att_noHPO scenario). Means of each error metric are denoted by '+'.

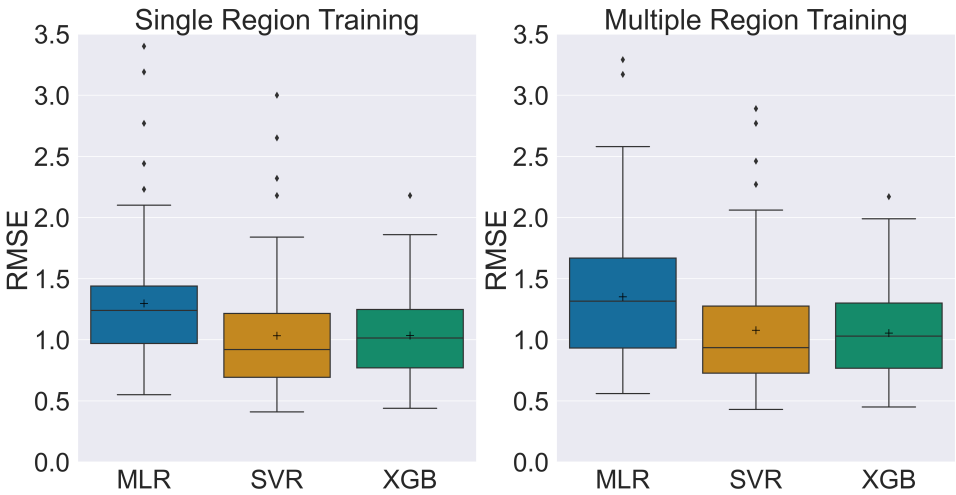

Figure S4. RMSE for regional predictions with HUC 02, HUC17 comparing training with single regions (02_17_SR_Att_noHPO, left) and training with multiple regions (02_17_MR_Att_noHPO, right).

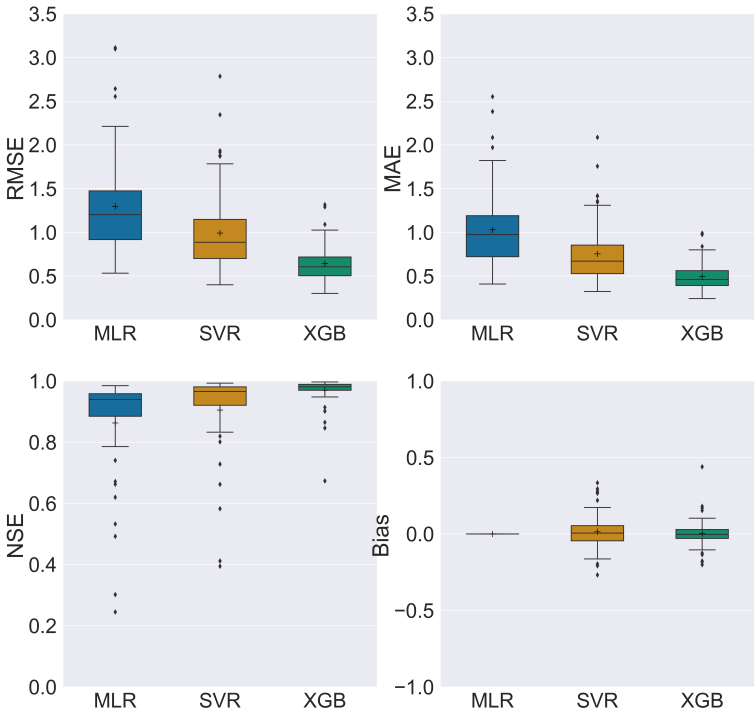

Figure S5. PUB RMSE (a) MAE (b) NSE (c) and Bias (d) distributions averaged over the test period for all 78 stations (PUB_02_17_SR scenario). Means for all error metrics are denoted by (+).

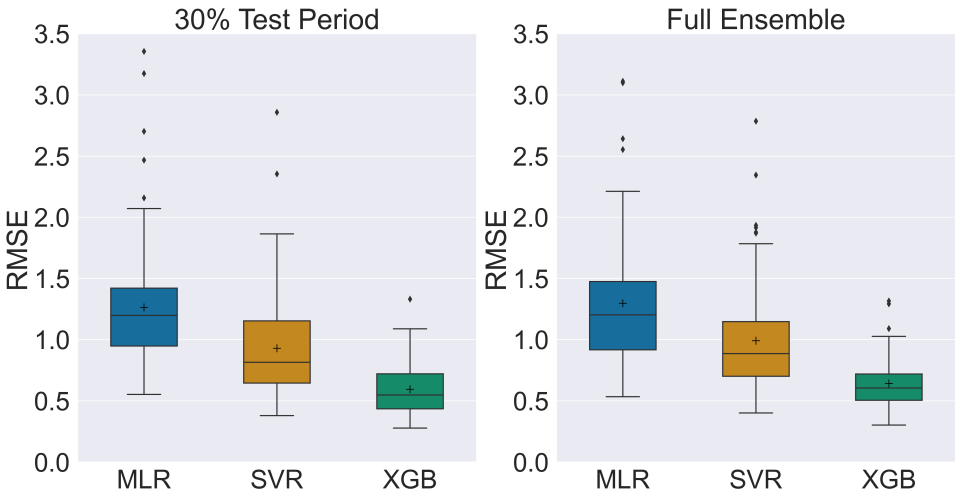

Figure S6. Sensitivity of the ensemble PUB model RMSE to changing the testing period, with the last 30% of the data record scored on the left and the full record scored on the right.

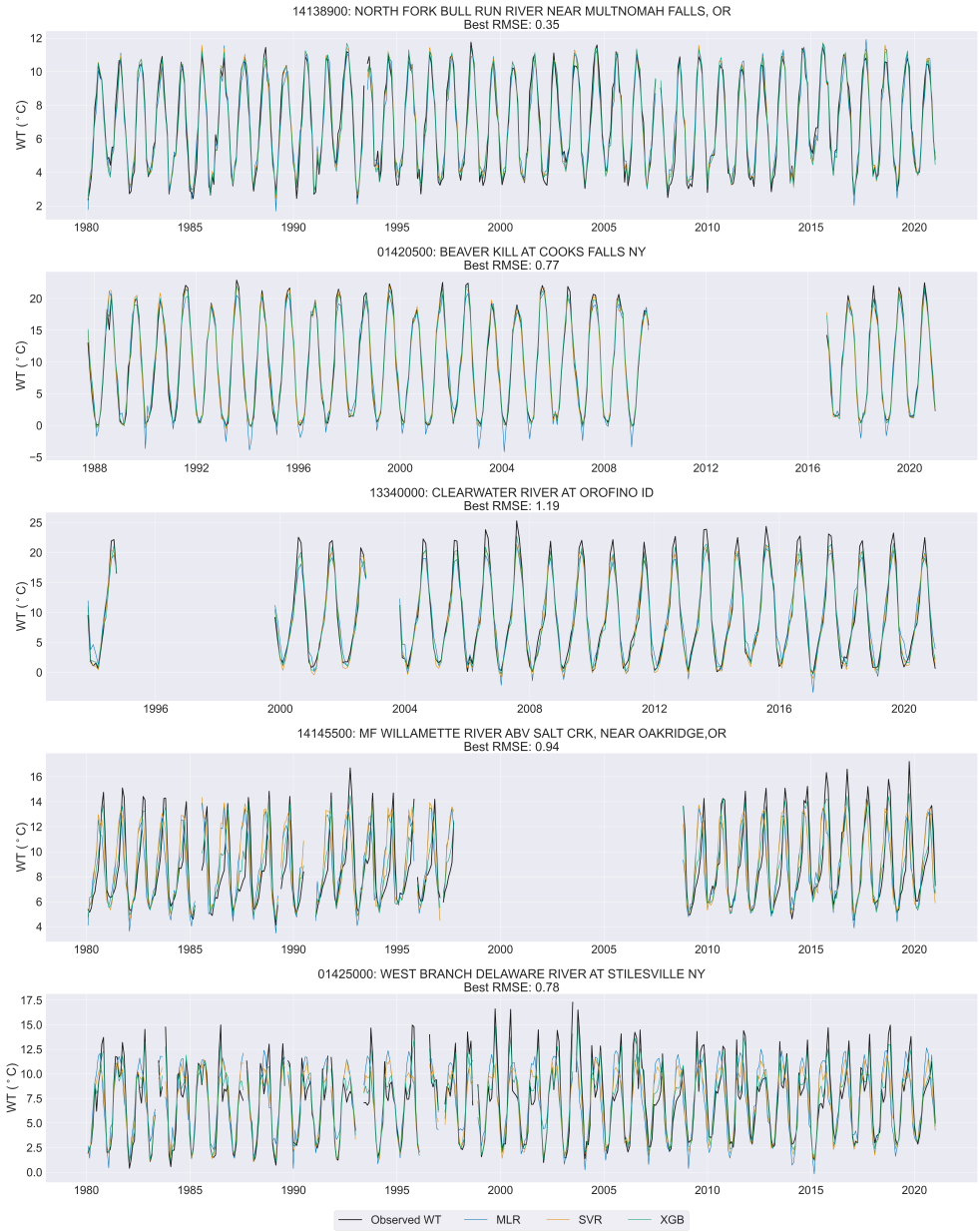

Figure S7. Examples of time series predictions from the ensemble PUB models for the five representative stations shown in Figure 7.

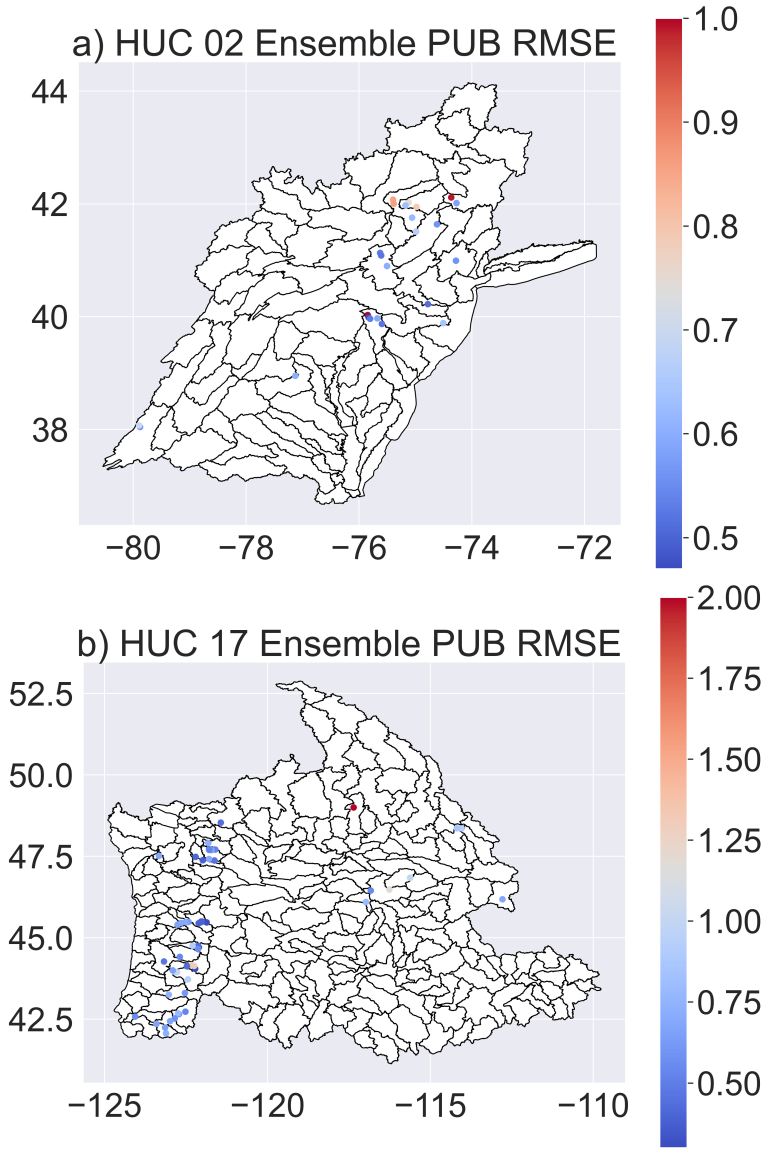

Figure S8. Spatial distribution of the RMSE for the ensemble PUB XGB models (PUB_02_17_SR) for all station locations in a) HUC 02 and b) HUC 17.