Supplement of Hydrol. Earth Syst. Sci., 30, 1077–1095, 2026
https://doi.org/10.5194/hess-30-1077-2026-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

Sensitivity of hydrological machine learning prediction accuracy to information quantity and quality

Minhyuk Jeung et al.

Correspondence to: Kwangsik Yoon (ksyoon@chonnam.ac.kr)

S1. Performance enhancement through mechanistic data integration

This study used continuous daily flow rate and nutrient load data that were derived from mechanistic models. However, raw flow rate and nutrient concentration observations can be valuable as part of the training data for machine learning models. Nutrient concentration records are not typically available at daily intervals; instead, they are usually collected at longer temporal intervals, likely due to cost concerns. In this study, the nutrient concentrations were measured every one or two weeks. In contrast, flow rate data—generated from flow level observations—are generally recorded at finer temporal resolutions (such as hourly or daily) and converted using a rating curve. As a result, nutrient concentration records provide less detailed information compared to nutrient load estimates, which can be obtained from mechanistic model outputs. Consequently, nutrient concentration records may have a more limited impact on improving the predictive accuracy of hydrological machine learning models, compared to the daily nutrient load estimates.

To evaluate this assumption, we conducted seven additional experiments directly incorporating observed flow rates and nutrient concentration records into training data to assess their impact on the predictive accuracy of hydrological machine learning models: 1) WD+O_Flow (with observed flow rates), 2) WD+O_NCon (with observed nitrogen concentrations), 3) WD+O_PCon (with observed phosphorus concentrations), 4) WD+O_Flow+O_NCon (with both observed flow rates and nitrogen concentrations), 5) WD+O_Flow+O_PCon (with both observed flow rates and phosphorus concentrations), 6) WD+S_NLoad (with simulated nitrogen loads using the calibrated SWAT model), and 7) WD+S_PLoad (with simulated phosphorus loads using the calibrated SWAT model). Here, "Flow," "NCon," and "PCon" refer to observed flow rates, nitrogen concentrations, and phosphorus concentrations, respectively. In addition, “S_NLoad” and “S_PLoad” represent simulated nitrogen and phosphorus loads derived from using the calibrated mechanistic model, serving as reference cases that use continuous daily data.

The test results indicated that the WD+O_Flow case performed comparably or even better in predicting TN and TP loads compared to the WD+O_NCon and WD+O_PCon cases (Fig. S12). Previous studies have shown that pollutant loads are primarily influenced by streamflow rather than nutrient concentrations, as demonstrated by regression models relating streamflow to pollutant loads (Lee et al., 2016; Song et al., 2022; Song et al., 2024; Wu et al., 2024). This is likely because the energy, or transport capacity, of flowing runoff primarily drives the detachment of sediment and nutrients from the soil surface and their downstream transport to waterbodies, particularly under transport-limited (rather than supply-limited) conditions (Basu et al., 2010; Wainwright et al., 2015; Song et al., 2024). Additionally, while using all available variables (WD+O_Flow+O_NCon and WD+O_Flow+O_PCon) improved performance relative to cases that only incorporated nutrient concentrations alongside weather records (WD+O_NCon and WD+O_PCon), it still underperformed when compared to using datasets containing continuous daily nutrient load data derived from mechanistic models (WD+S_NLoad and WD+S_PLoad). This result was expected, given the significantly smaller number of observed nutrient concentration records compared to the daily nutrient load estimates produced by mechanistic models. For example, TN concentrations were measured every one to two weeks from July 12, 2013, to December 31, 2017, resulting in 109 to 229 data points per watershed. In contrast, the mechanistic model generated 1,634 daily TN load estimates for the same period. The average information content of the observed concentrations was 7.001 bits for

TN and 6.808 bits for TP, substantially lower than the simulated loads, which contained 8.600 bits for TN and 9.437 bits for TP, on average. These findings highlight the crucial role of data quantity in improving the predictive accuracy of hydrological machine learning models.

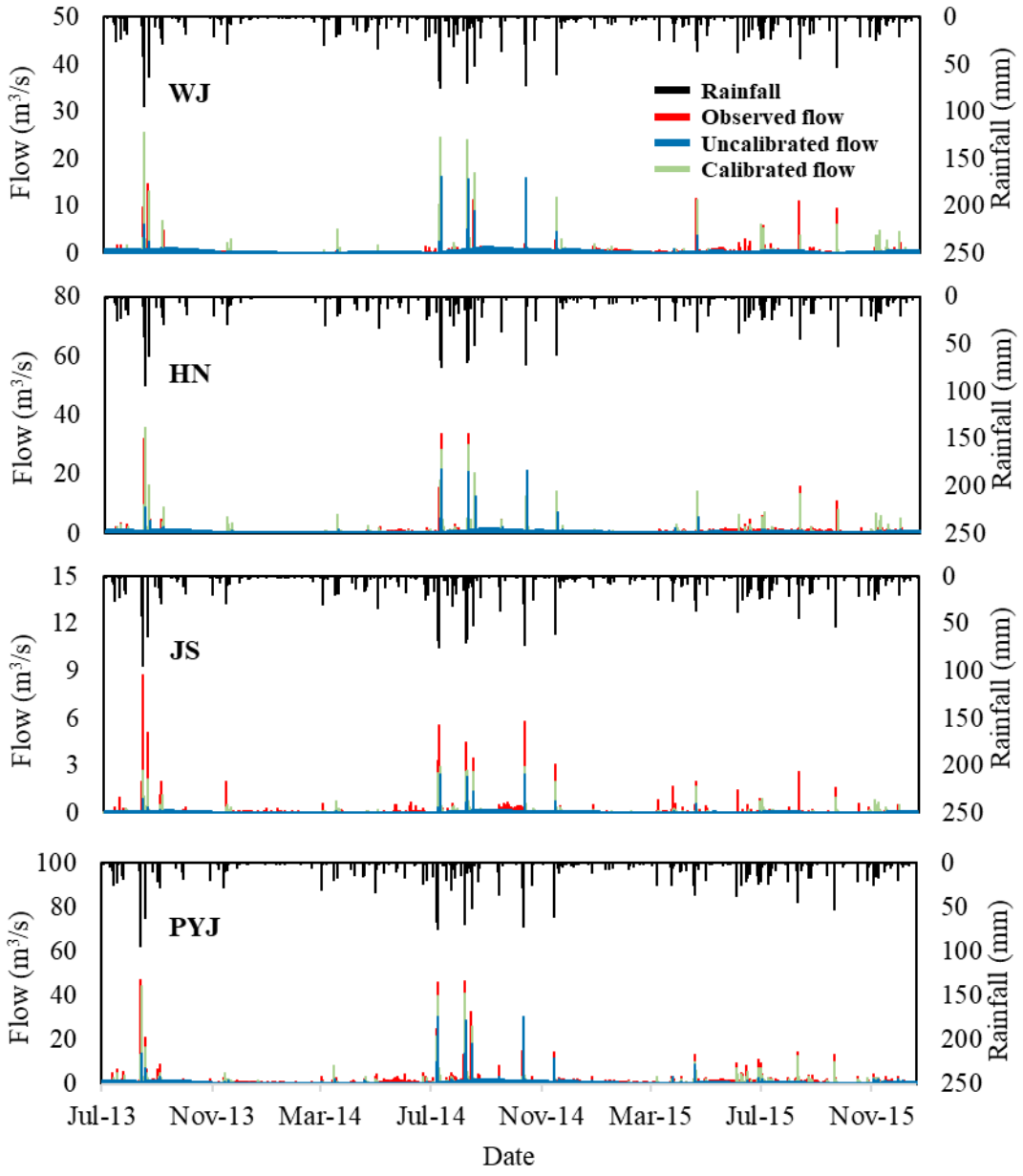


Figure S1. Comparison of daily streamflow predicted using the mechanistic models (i.e., uncalibrated and calibrated SWAT models) and observed during the training period (July 12, 2013, to December 31, 2015).

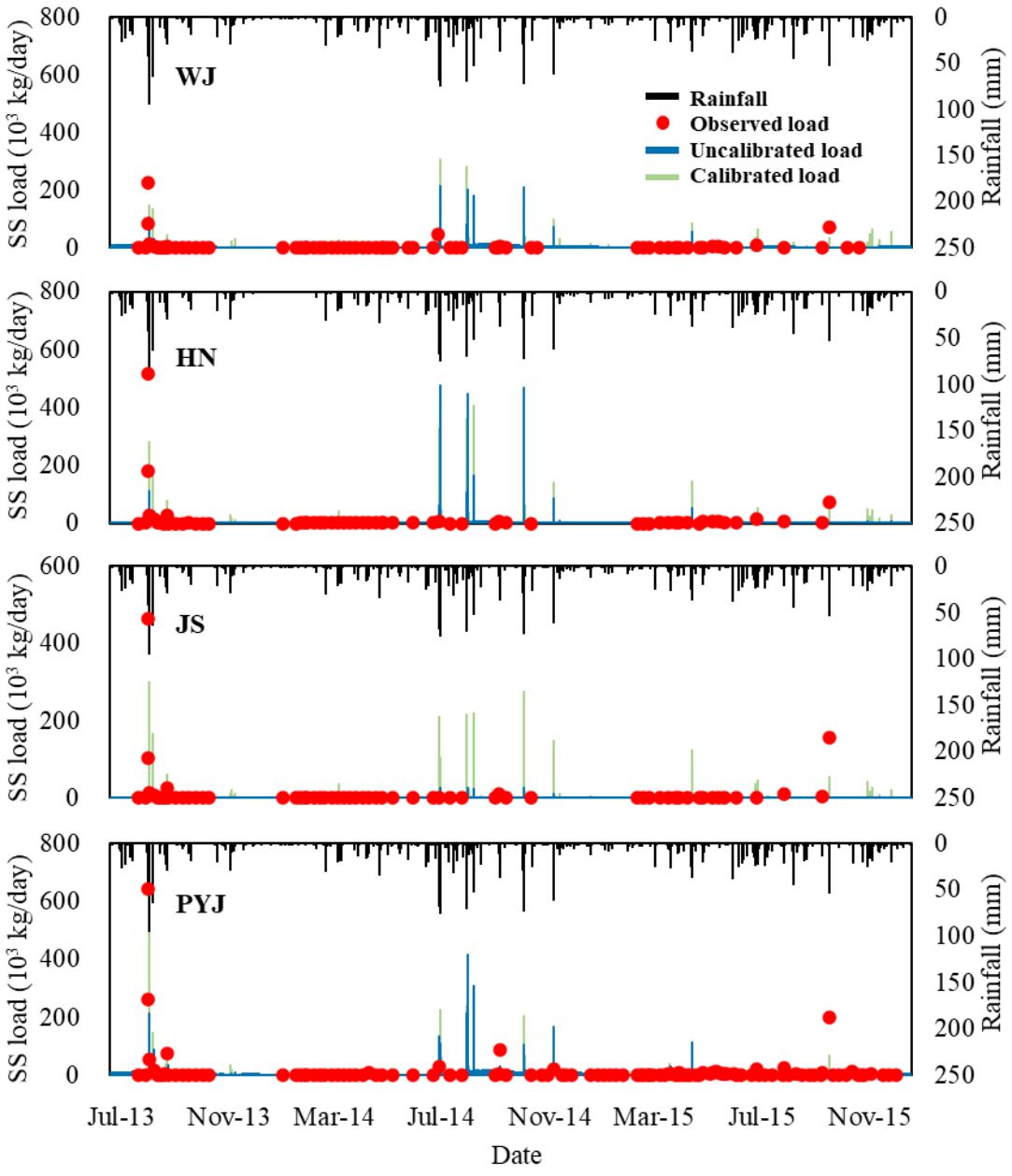


Figure S2. Comparison of daily SS loads predicted using the mechanistic models (i.e., uncalibrated and calibrated SWAT models) and observed during the training period (July 12, 2013, to December 31, 2015).

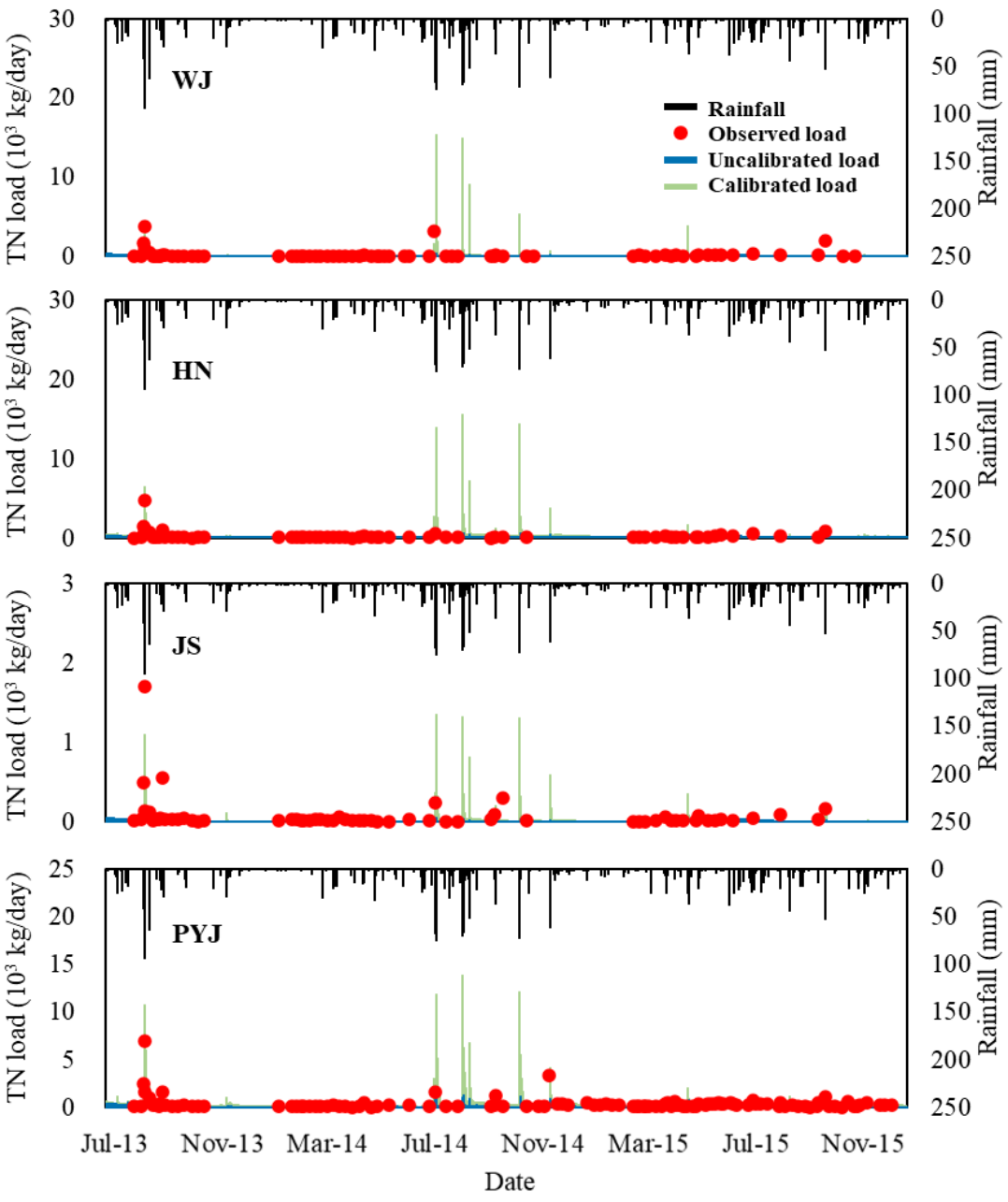


Figure S3. Comparison of daily TN loads predicted using the mechanistic models (i.e., uncalibrated and calibrated SWAT models) and observed during the training period (July 12, 2013, to December 31, 2015).

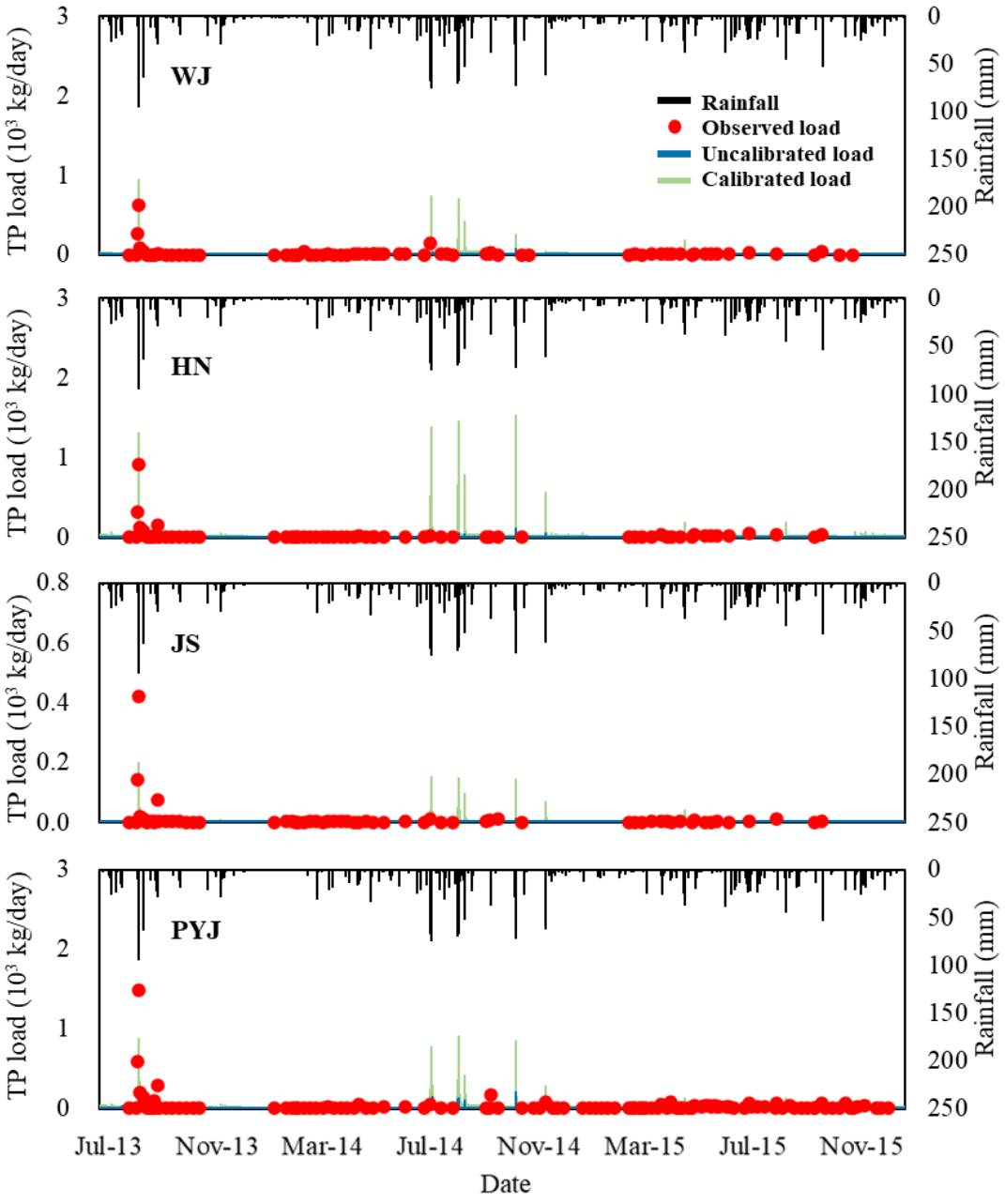


Figure S4. Comparison of daily TP loads predicted using the mechanistic models (i.e., uncalibrated and calibrated SWAT models) and observed during the training period (July 12, 2013, to December 31, 2015).

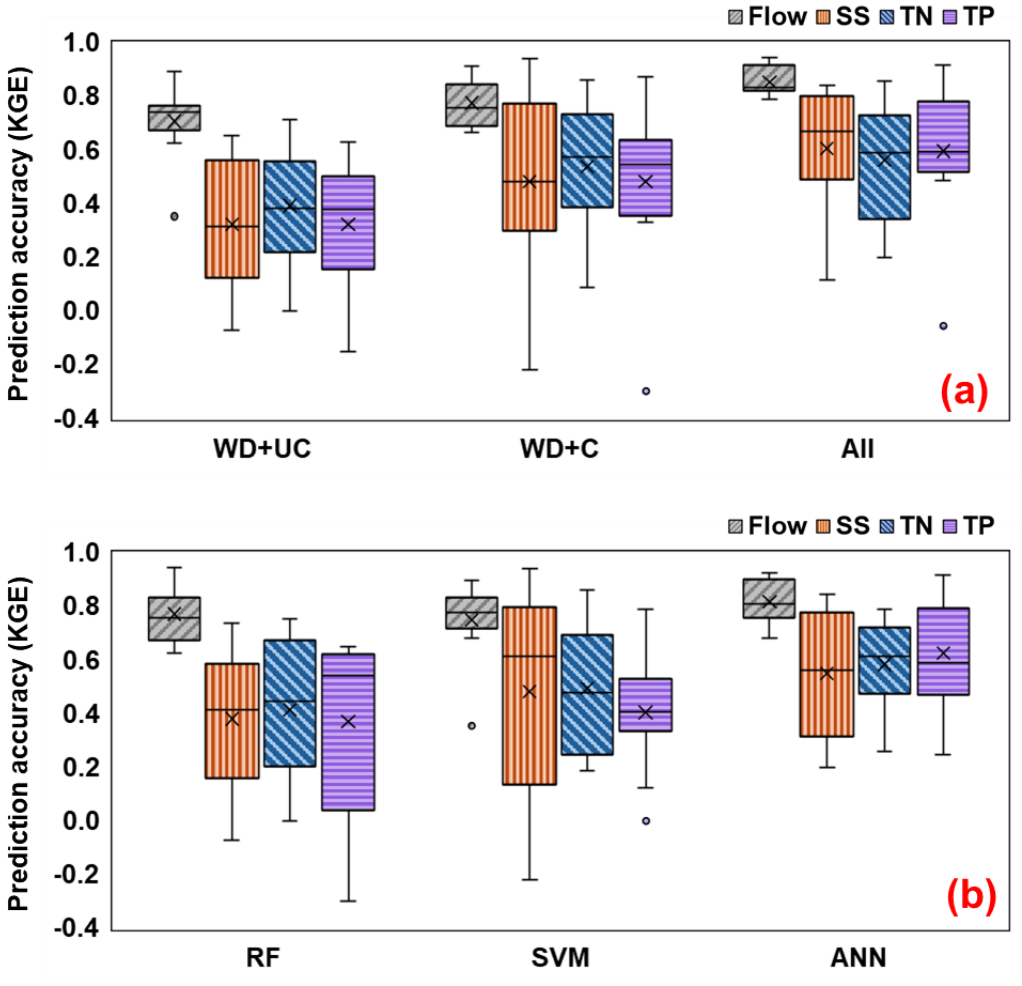
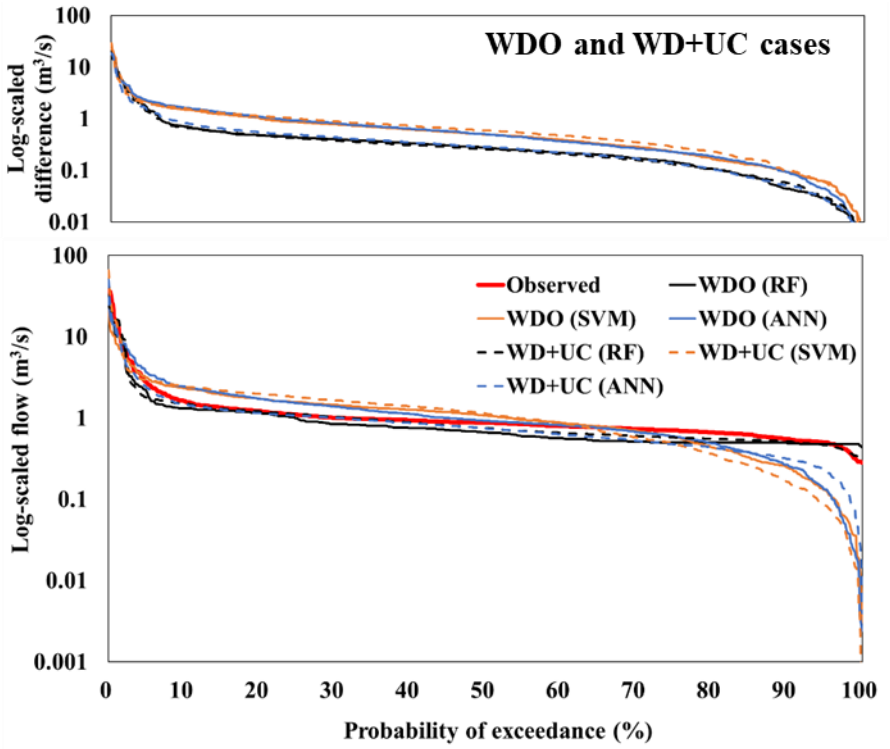
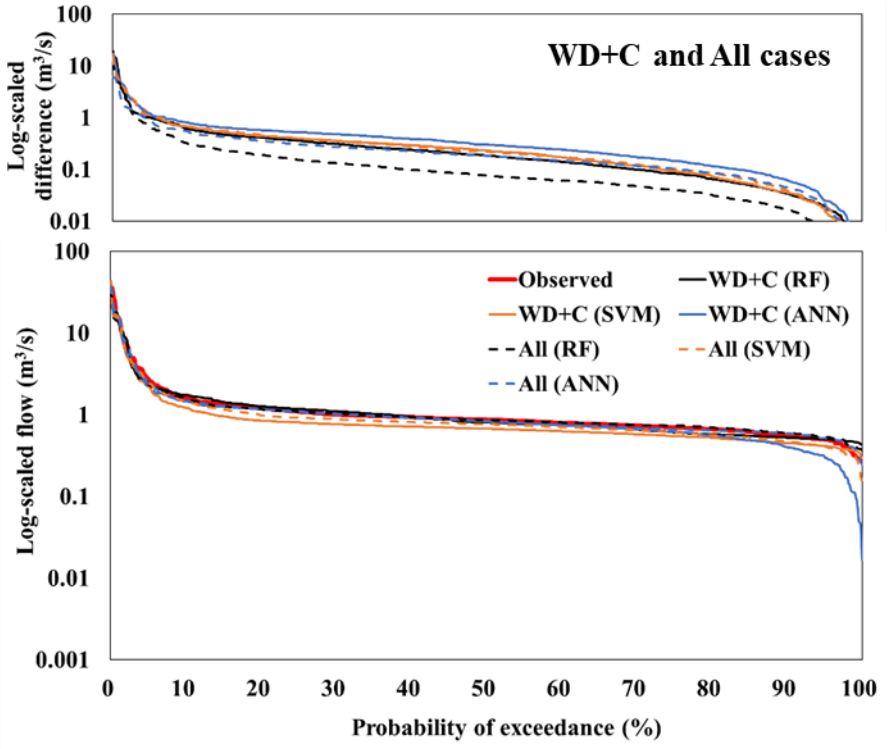
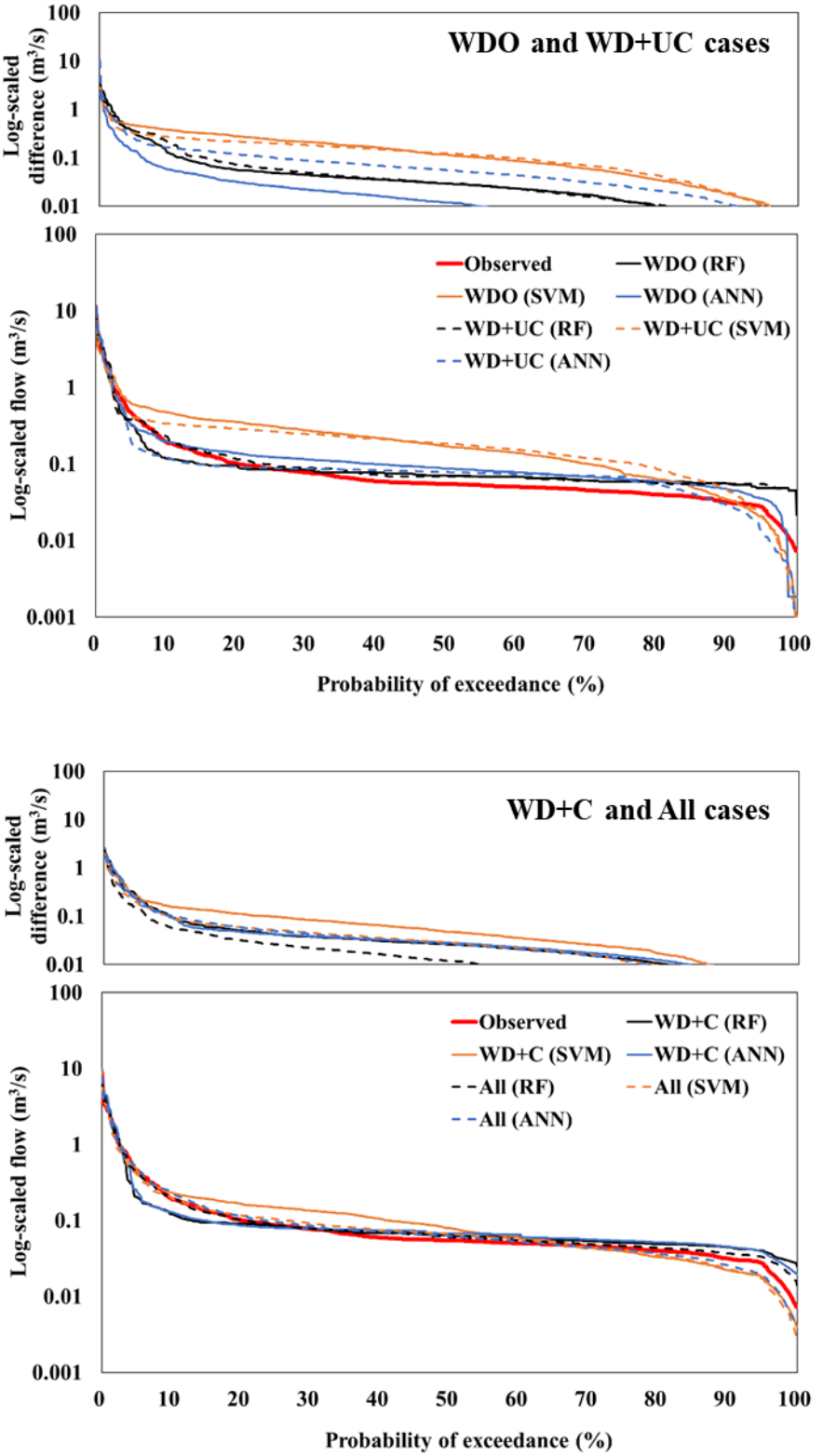


Figure S5. Variations in prediction accuracy (KGE) of the ML models for the watersheds and ML models (a) and for the watersheds and training datasets (b).

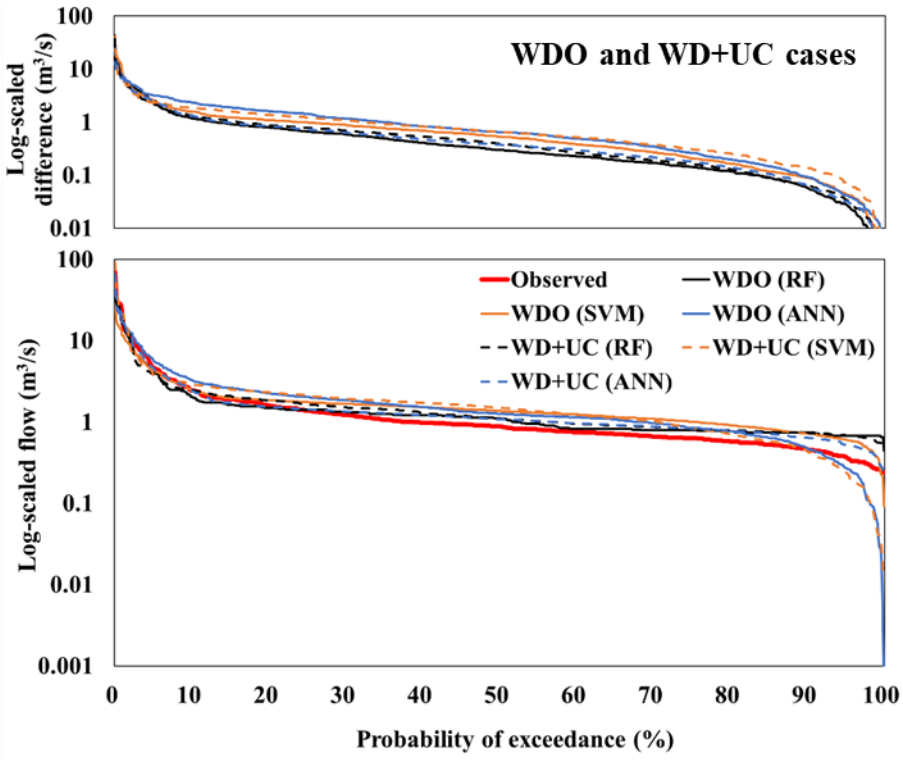


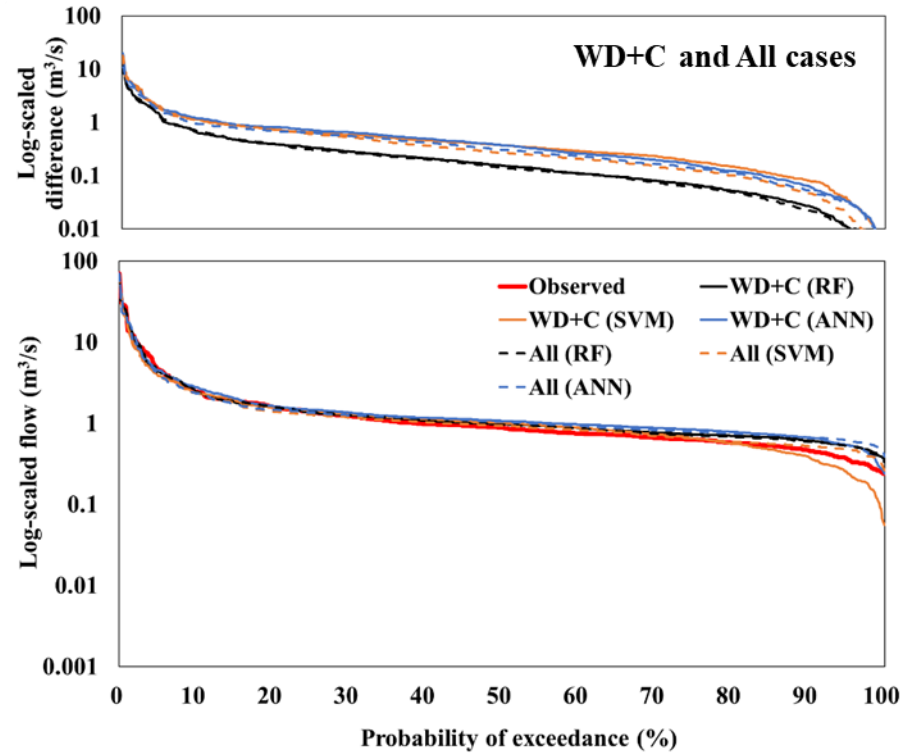


(a) Log-scaled flow duration curves at the outlet of the HN watershed

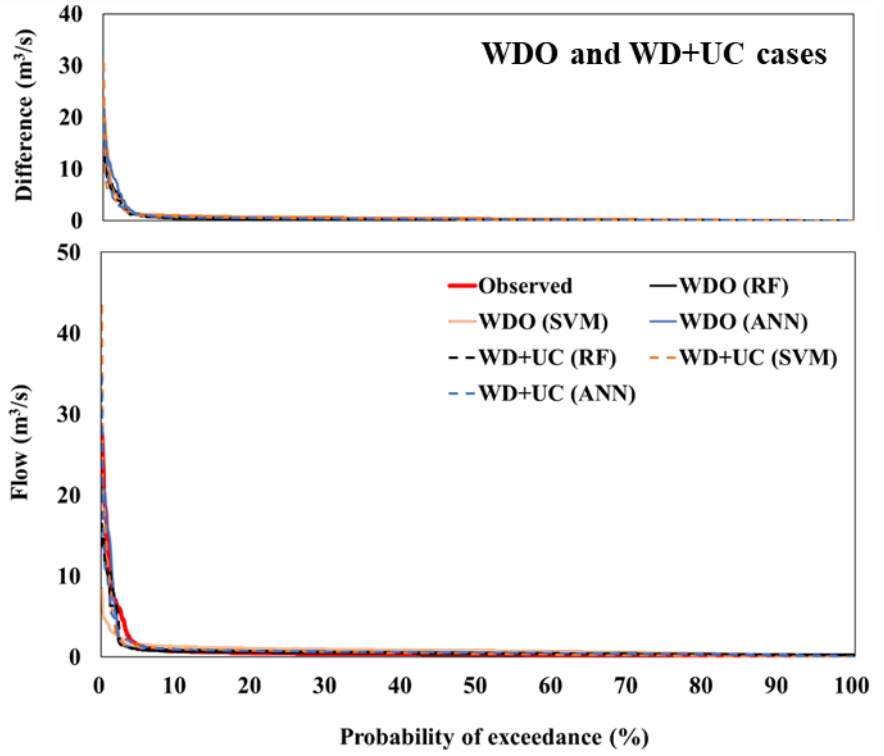


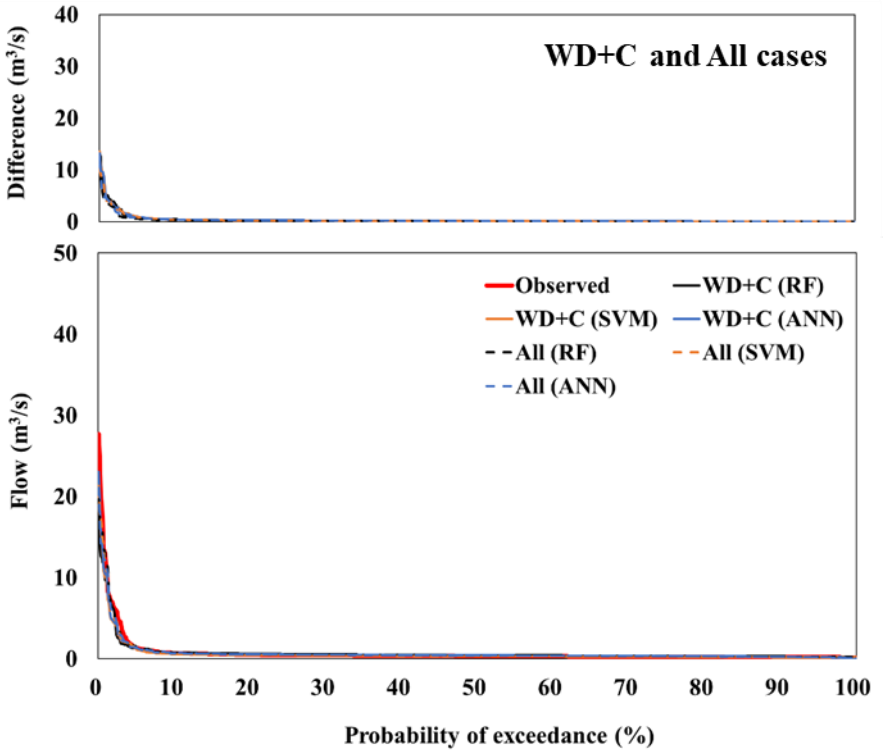
(b) Log-scaled flow duration curves at the outlet of the JS watershed



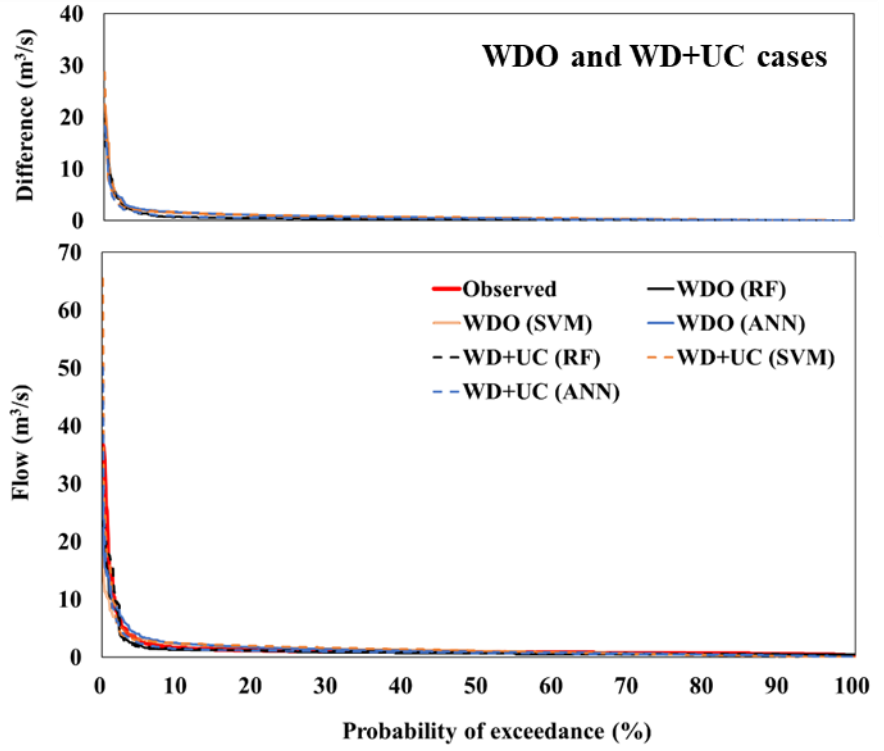


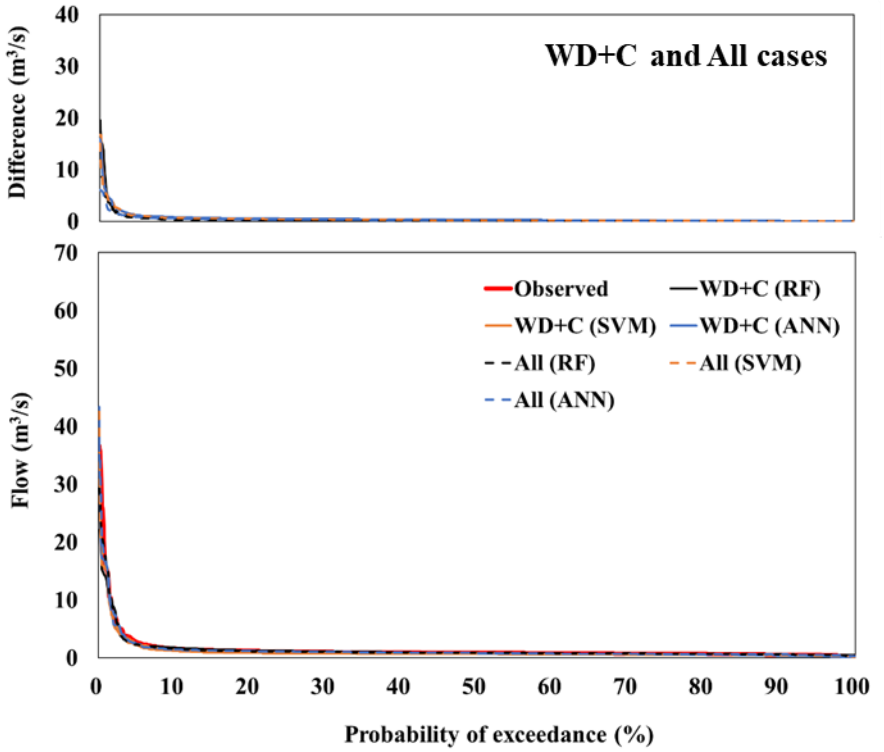
(c) Log-scaled flow duration curves at the outlet of the PYJ watershed



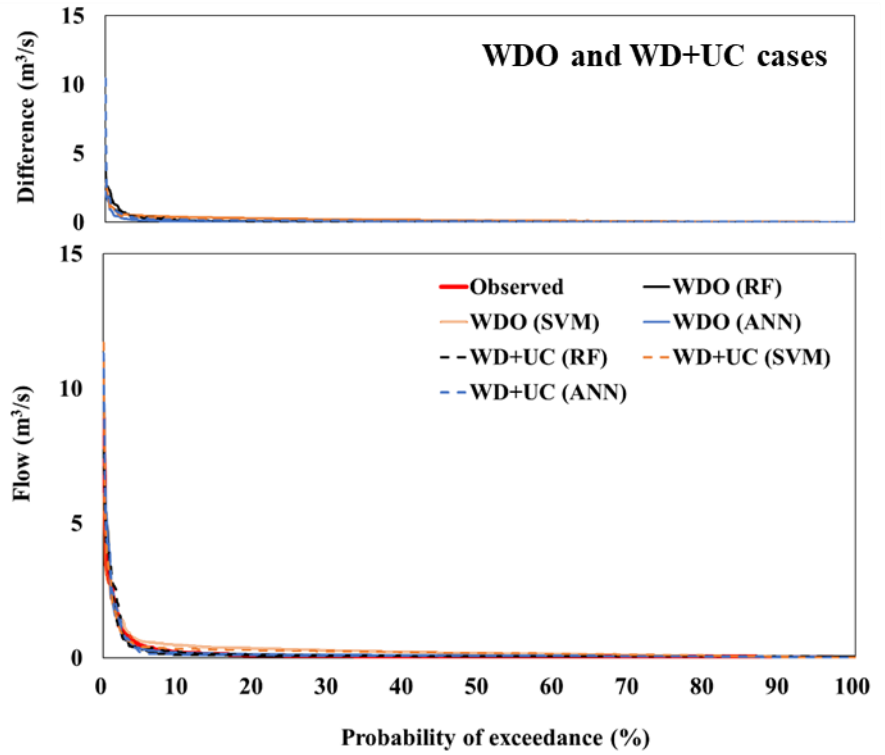


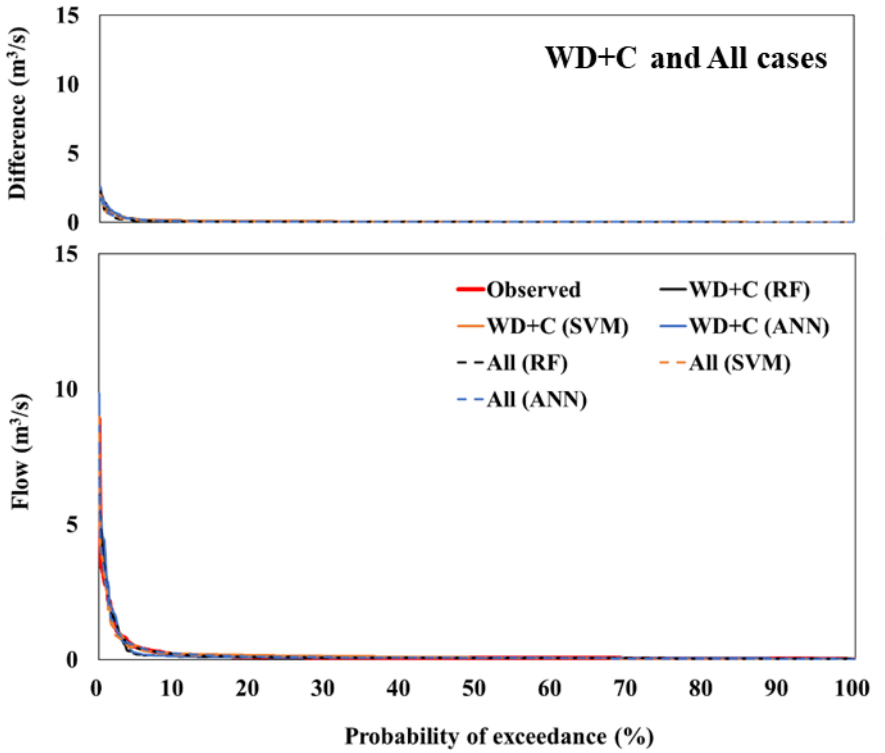
(d) Normal-scaled flow duration curves at the outlet of WJ watershed



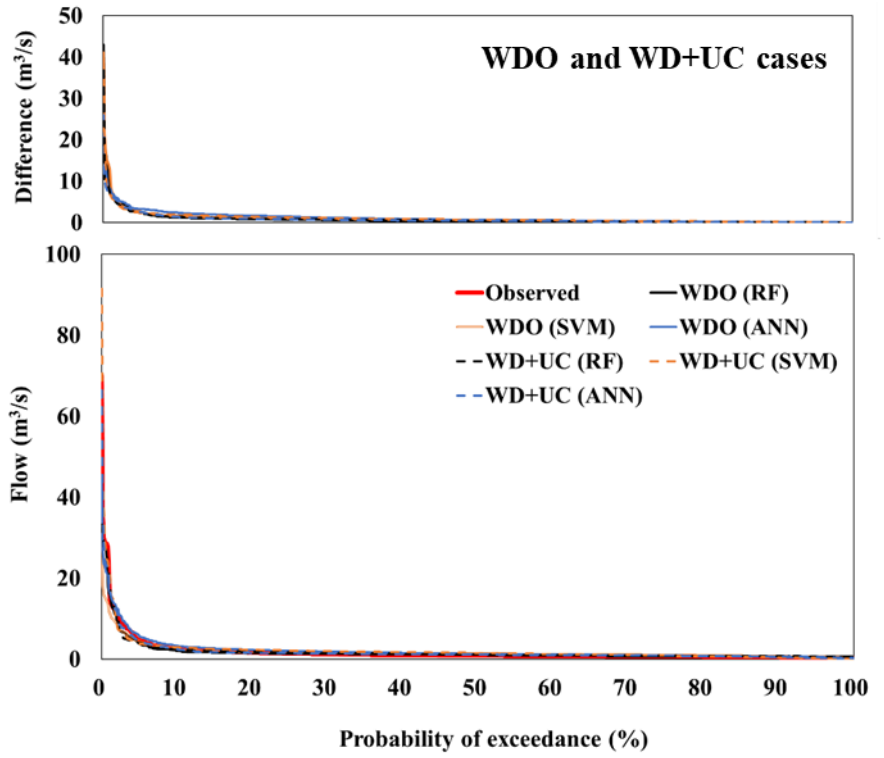


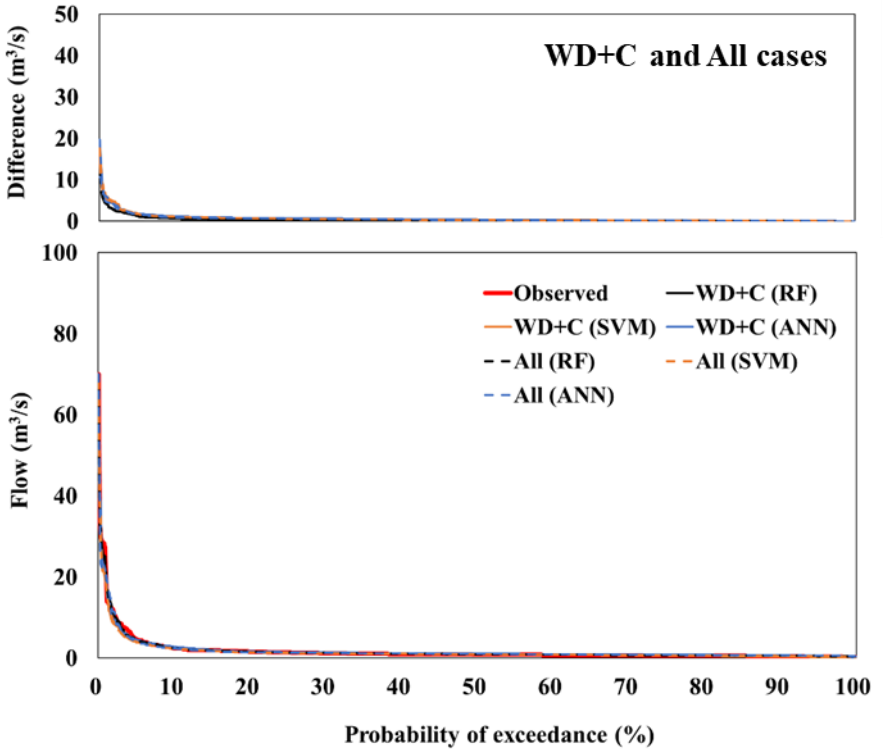
(e) Normal-scaled flow duration curves at the outlet of HN watershed





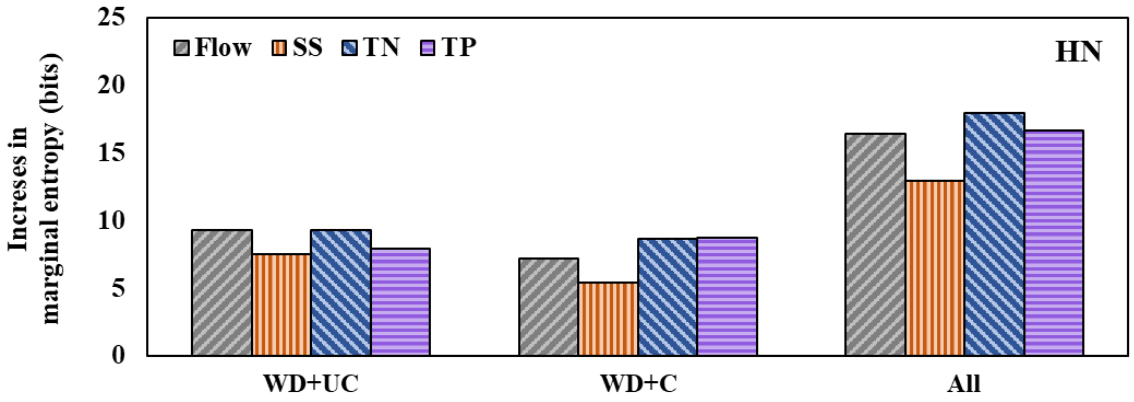
(f) Normal-scaled flow duration curves at the outlet of JS watershed



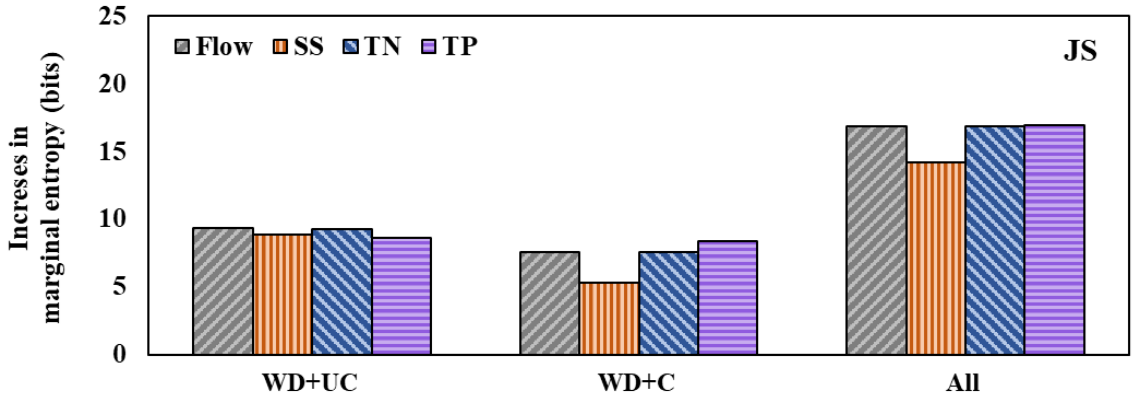


(g) Normal-scaled flow duration curves at the outlet of PYJ watershed

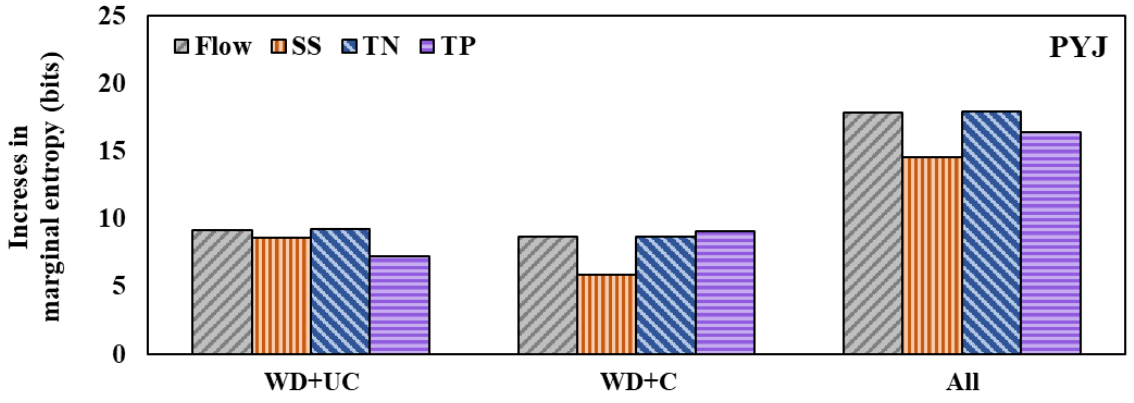
Figure S6. Comparison between observed and ML-predicted flow duration curves (FDC). (a, b, and c) represent the FDC in log scale at the outlets of the HN, JS, and PYJ watersheds. The FDC at the WJ watershed is presented in Figure 5 in the manuscript. (d, e, f, and g) represent the FDC in normal scale at the outlets of WJ, HN ,JS, and PYJ watersheds.



(a) HN watershed

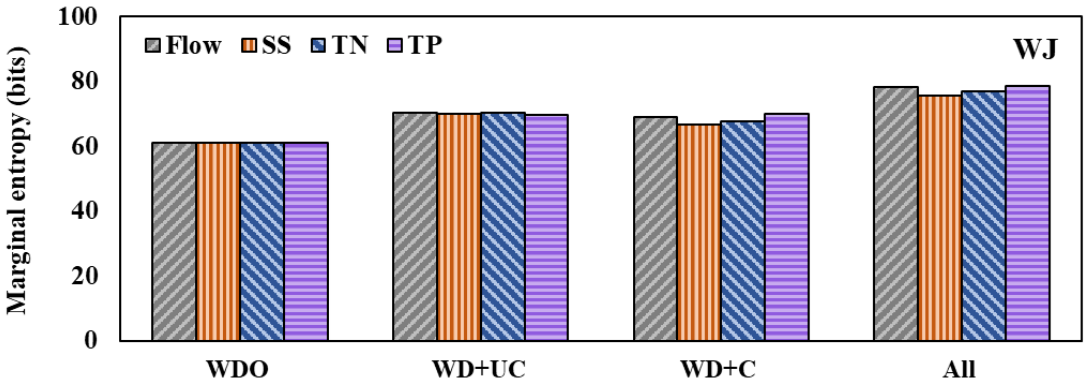


(b) JS watershed

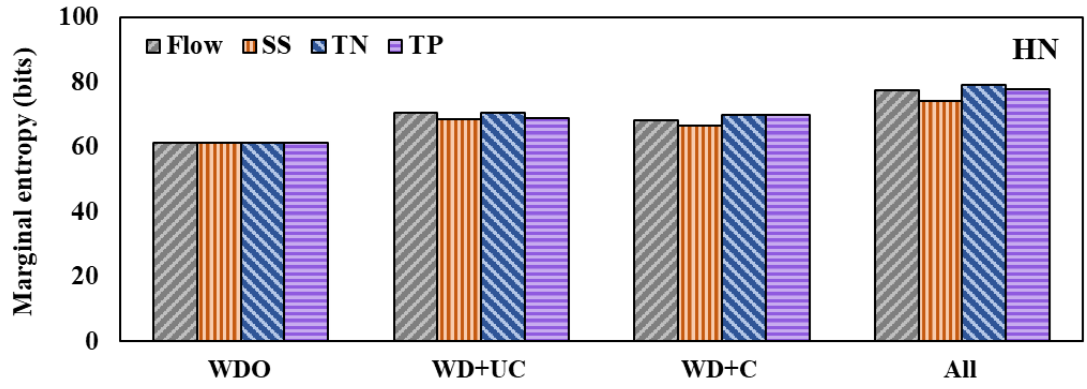


(C) PYJ watershed

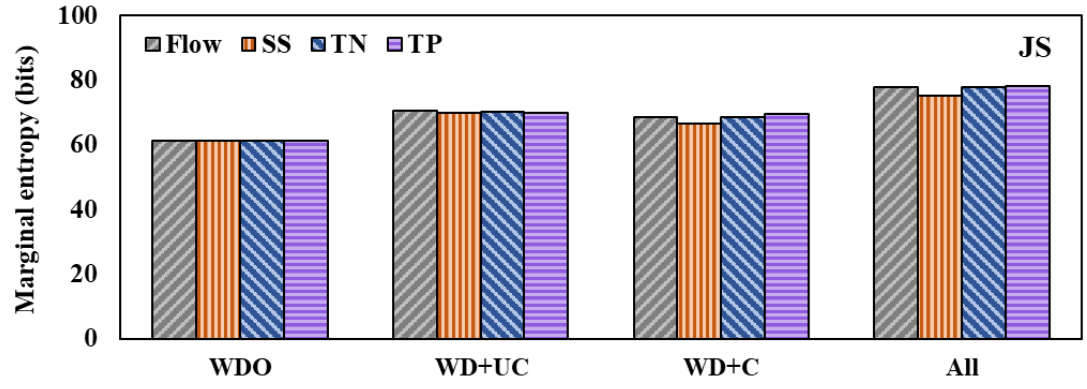
Figure S7. Increases in marginal entropy due to the addition of training data sets from different watersheds (HN, JS, and PYJ). The WDO training data set serves as the baseline for this comparison.



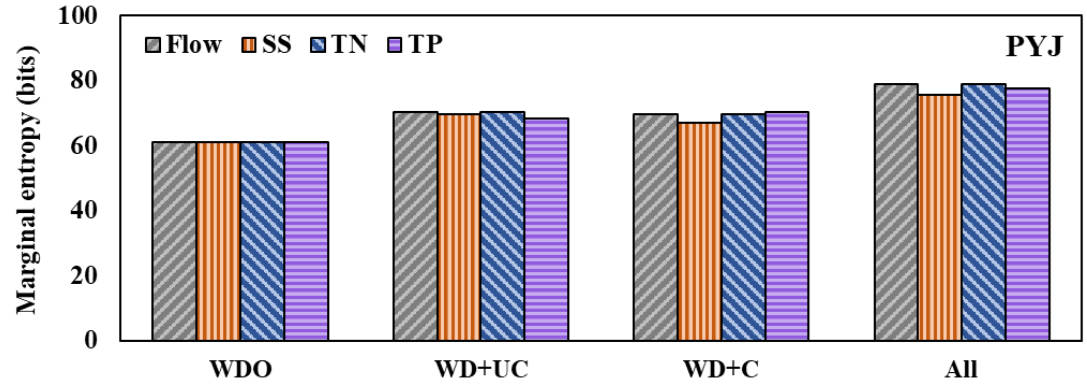
(a) WJ watershed



(b) HN watershed

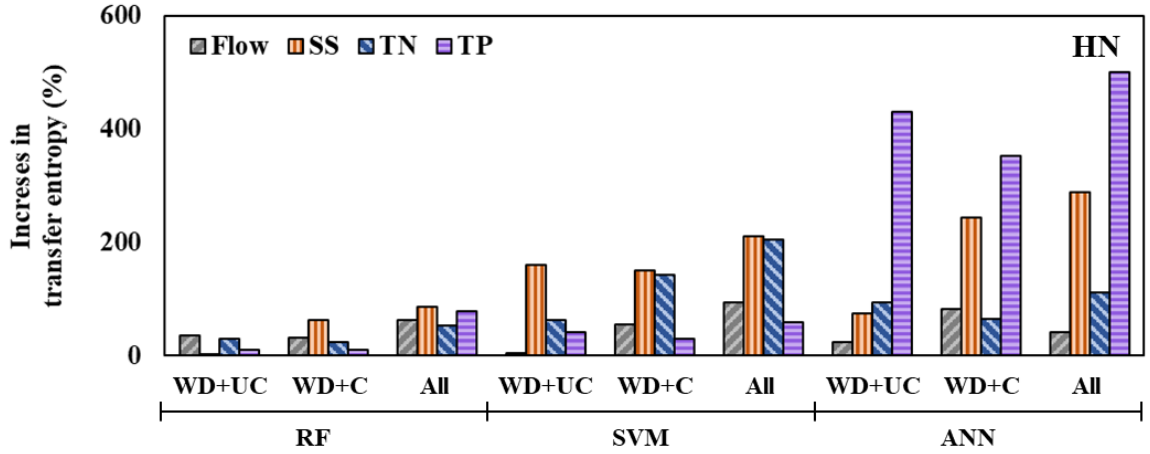


(c) JS watershed

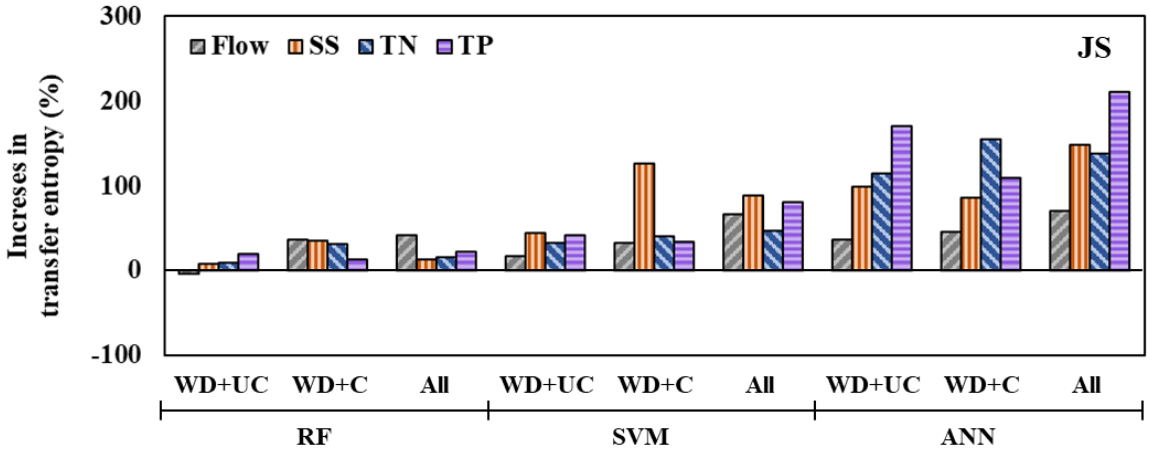


(d) PYJ watershed

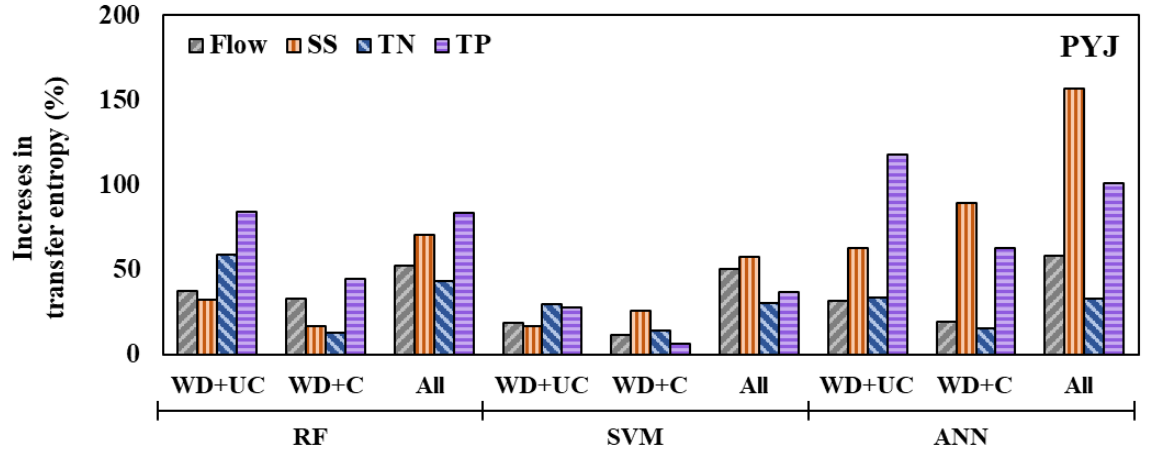
Figure S8. Marginal entropy of the training data sets by the watersheds with different target variables.



(a) HN watershed

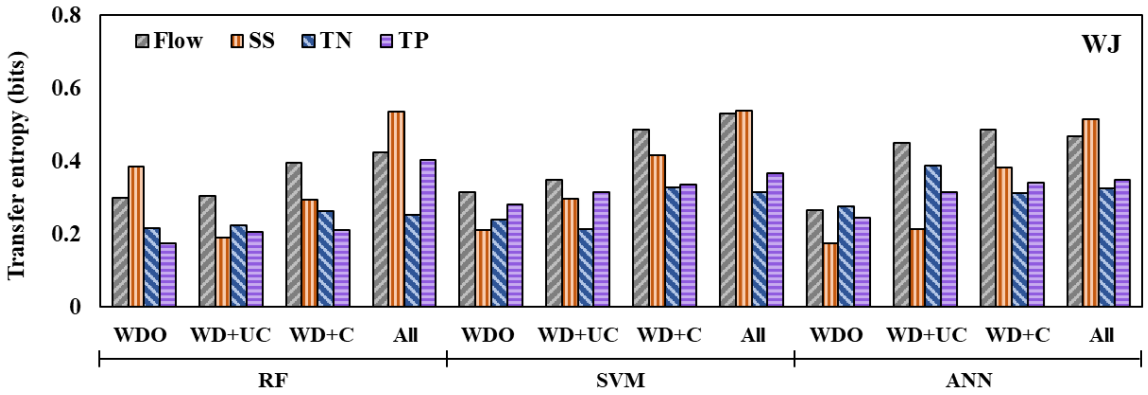


(b) JS watershed

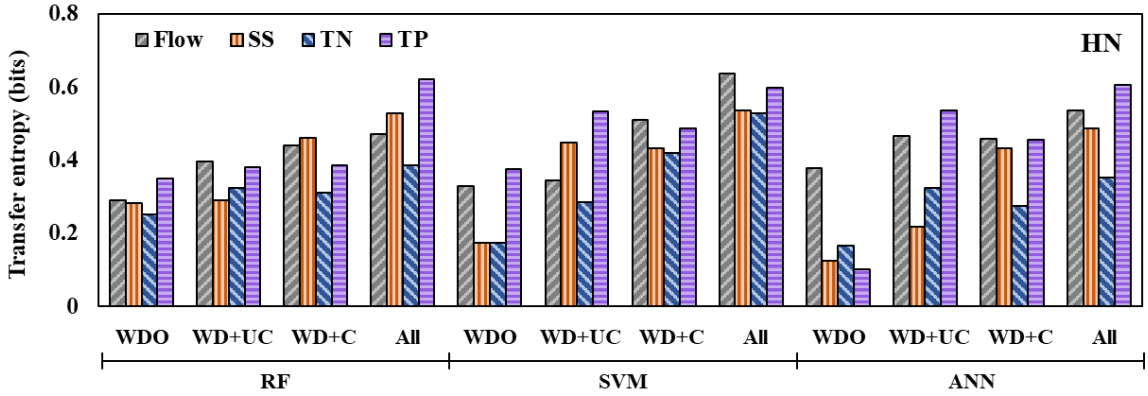


(C) PYJ watershed

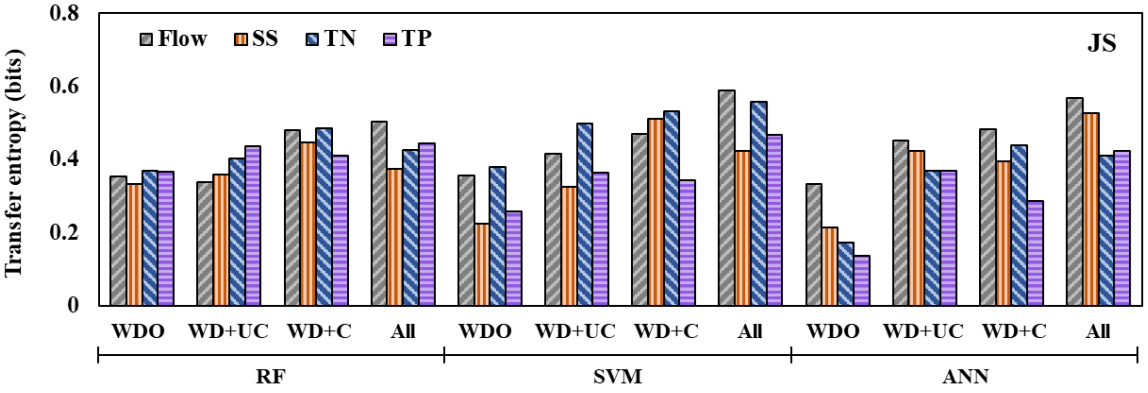
Figure S9. Increases in transfer entropy due to the addition of training data sets from different watersheds (HN, JS, and PYJ). The WDO training data set serves as the baseline for this comparison.



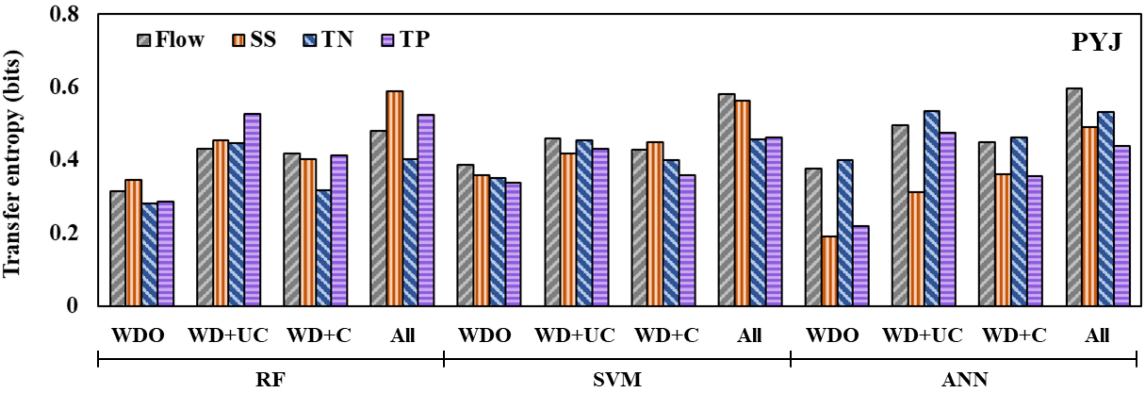
(a) WJ watershed



(b) HN watershed

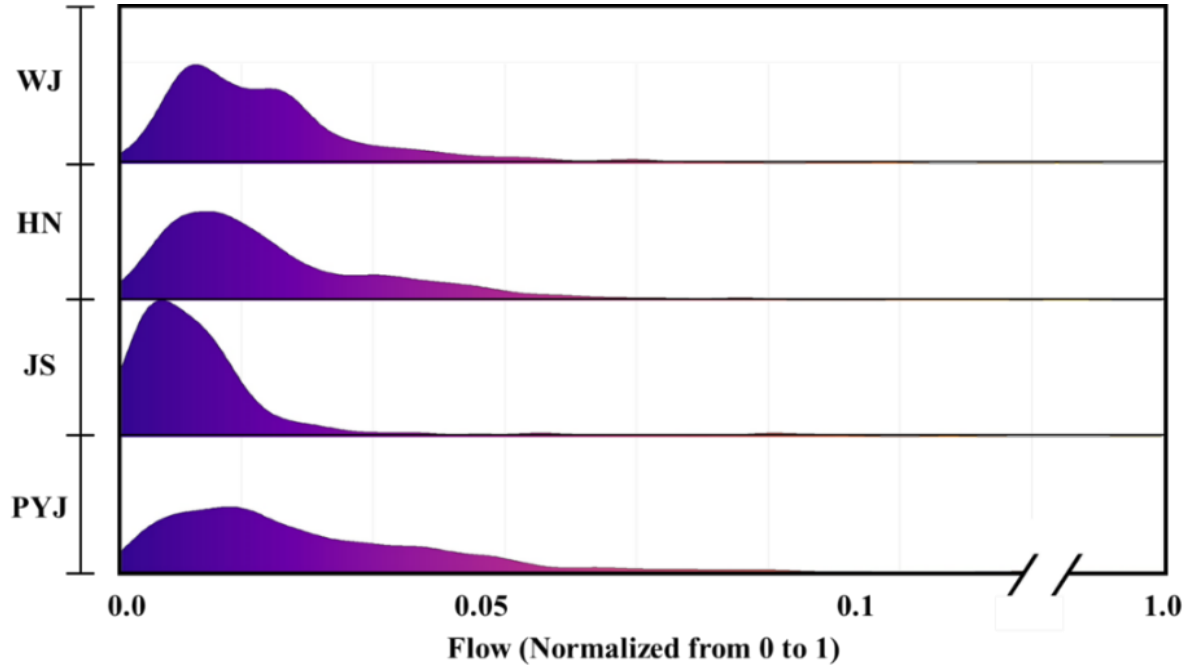


(c) JS watershed

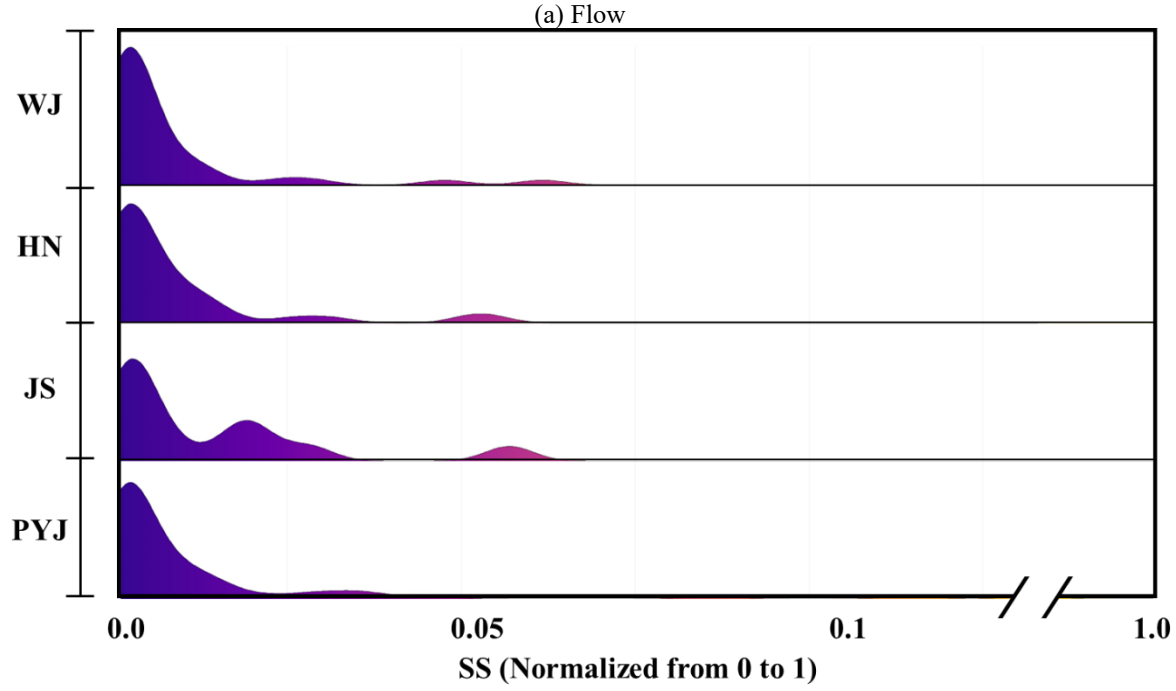


(d) PYJ watershed

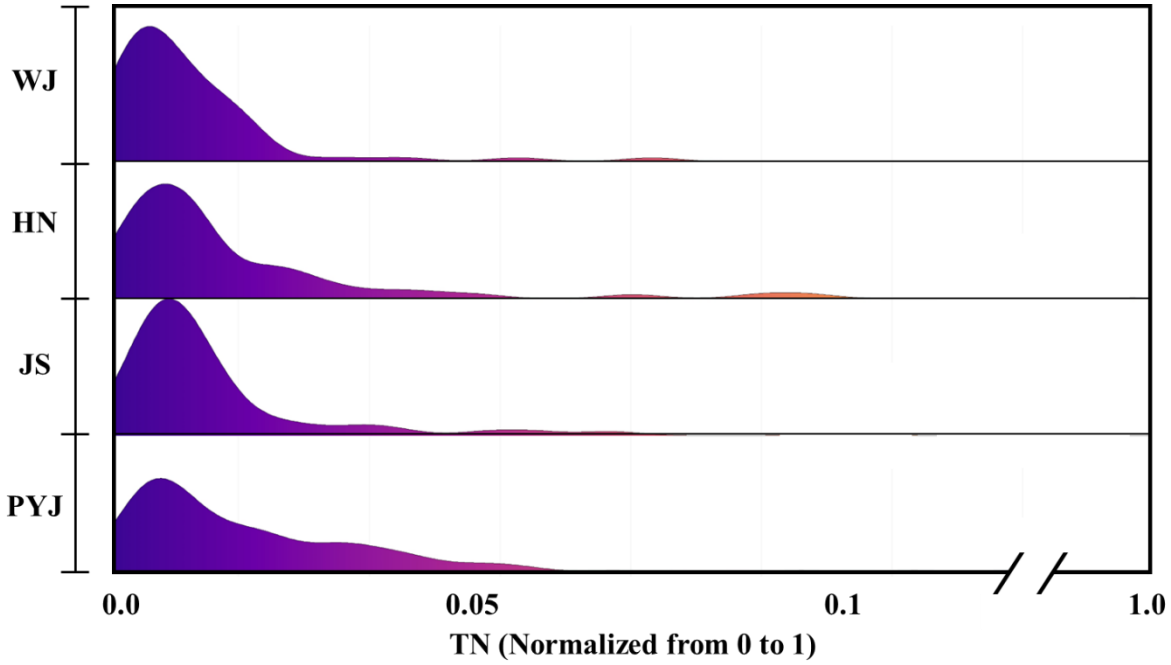
Figure S10. Transfer entropy of the training data sets by the machine learning algorithms with different target variables.


WJ
HN
JS
PYJ
0.0
0.05
0.1
1.0
Flow (Normalized from 0 to 1)

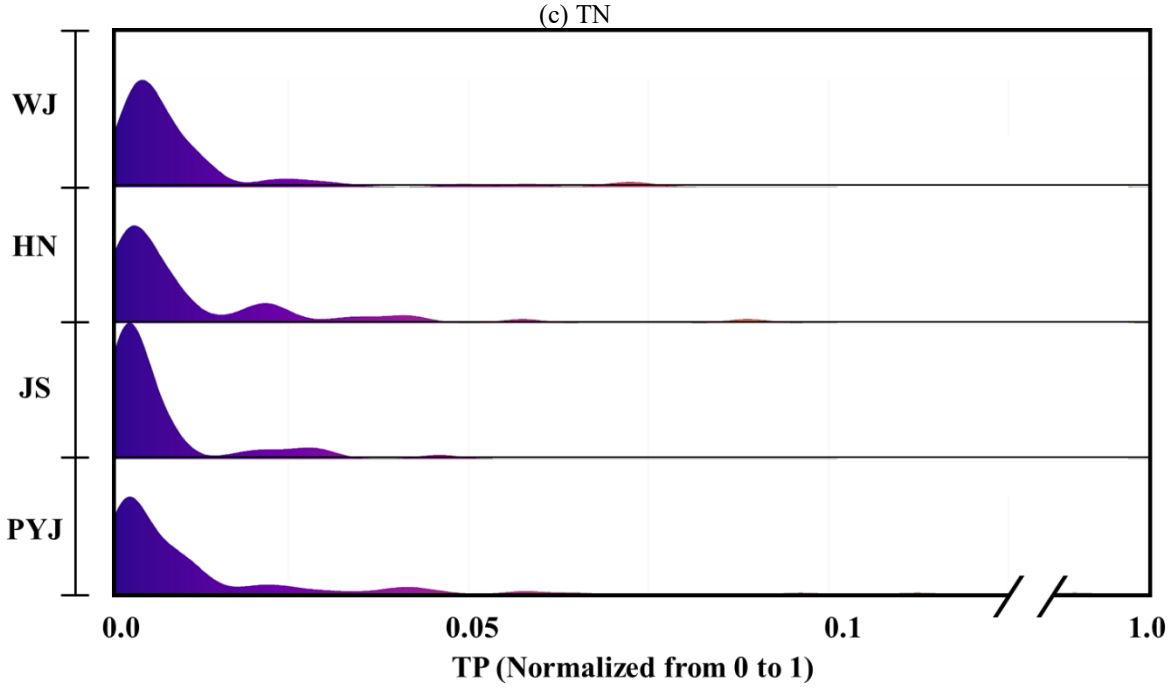
(a) Flow


WJ
HN
JS
PYJ
0.0
0.05
0.1
1.0
SS (Normalized from 0 to 1)

(b) SS



(c) TN



(d) TP

Figure S11. Density (or frequency) distributions of observed flow, SS, TN, and TP concentrations during the training period. The concentrations were normalized from 0 to 1 for each watershed.

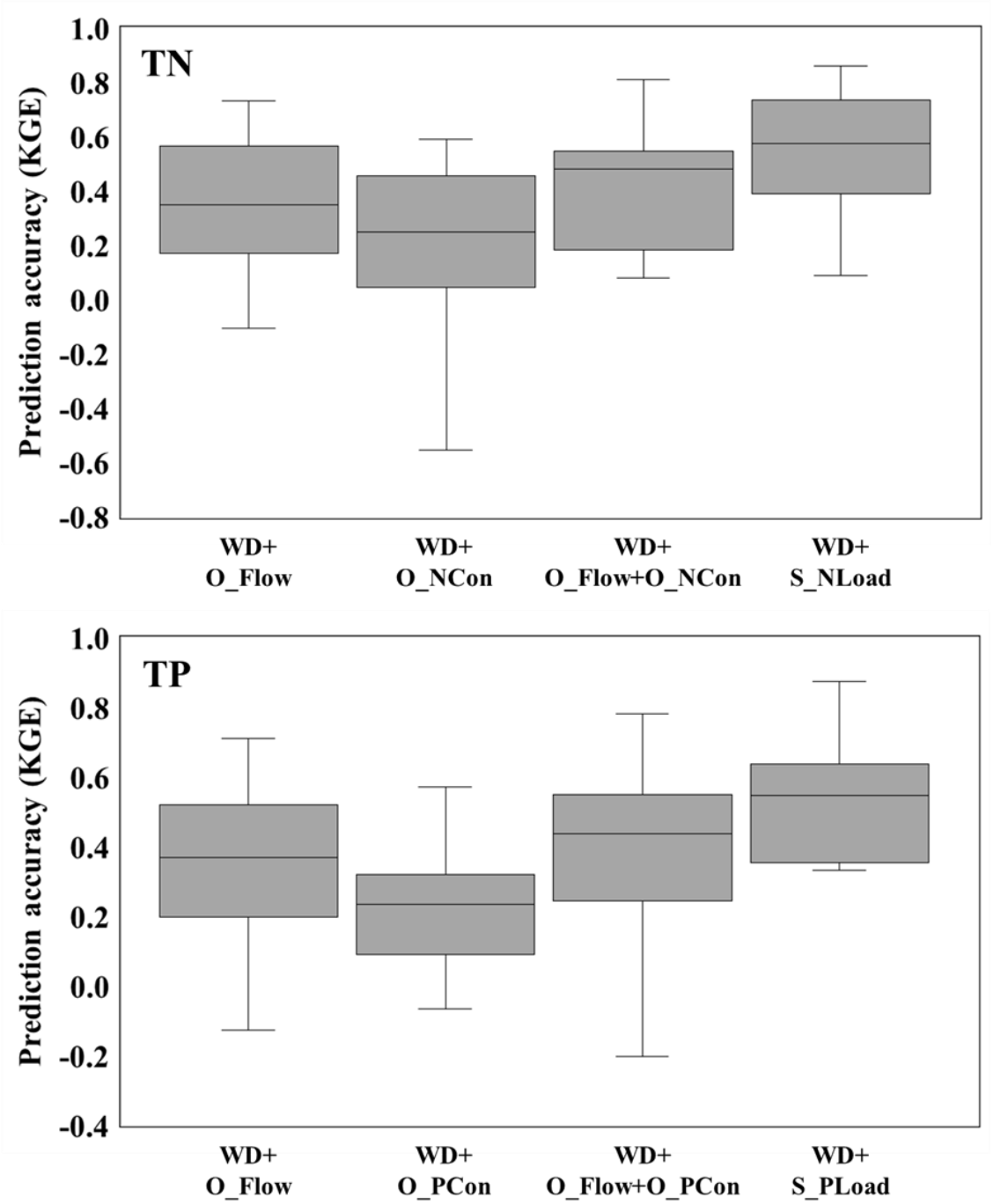


Figure S12. Comparison of TN and TP prediction accuracy (KGE) for ML models trained with different input data combinations across the four study watersheds. Each box and whisker plot represents the variation in prediction accuracy by watershed (WJ, HN, JS, and PYJ) and ML algorithm (RF, SVM, and ANN).

Table S1. Land use and cover statistics of the study watersheds.

Study Watersheds	Average Slope (%)	Land uses and Covers (km^2)					
		Urban	**Field**	**Paddy field**	**Forest**	**Other**	**Total**
WJ	6.43	1.88 (5%)	6.63 (19%)	14.77 (43%)	8.45 (24%)	3.02 (9%)	34.75 (100%)
HN	5.49	5.68 (13%)	9.28 (21%)	17.19 (38%)	9.33 (21%)	3.62 (8%)	45.09 (100%)
JS	4.39	2.40 (31%)	1.21 (15%)	2.07 (26%)	1.73 (22%)	0.42 (5%)	7.83 (100%)
PYJ	4.85	11.46 (19%)	22.36 (36%)	11.23 (18%)	11.45 (19%)	4.90 (8%)	61.40 (100%)

Table S2. Management practice application schedules and rates considered in the SWAT modelling.

Month	Day	Operation	Value	Crop
5	1	Fertilizer application	67.5 kg ha-1 (N) 45.0 kg ha-1 (P)	Rice
6	1	Plant/begin growing season		
7	10	Fertilizer application	22.5 kg ha-1 (N)	
10	1	Harvest and kill operation		

Table S3. Types of calibration parameters and their values in the SWAT modeling.

Variables	Parameters	Calibrated Values			
		WJ	HN	JS	PYJ
Flow	CN2.mgt	58.0	51.4	81.5	55.0
	ALPHA_BF.gw	0.02	0.58	0.12	0.23
	GW_DELAY.gw	424.6	470.4	414.6	34.2
	GW_REVAP.gw	0.12	0.15	0.09	0.12
	GWQMN.gw	3121	1671	404.2	125.0
	EPCO.bsn	0.20	0.55	0.33	0.19
	ESCO.bsn	0.57	0.98	0.01	0.39
	SUR_LAG.bsn	3.66	7.22	1.47	19.8
	OV_N.hru	0.23	0.23	0.28	0.14
	CH_N2.rte	0.15	0.18	0.17	0.27
	CH_K2.rte	85.3	5.42	21.3	0.83
	SOL_AWC.sol	0.04	0.09	0.03	0.51
SS	USLE_P.mgt	0.33	0.22	0.29	0.81
	SPCON.bsn	0.005	0.0004	0.008	0.003
	SPEXP.bsn	1.42	1.24	1.08	1.43
	ADJ_PKR.bsn	1.46	1.15	0.97	0.70
	EROS_SPL.bsn	1.08	3.00	1.40	2.14
	EROS_EXPO.bsn	2.43	1.94	1.78	2.08
	C_FACTOR.bsn	0.07	0.30	0.28	0.01
	RILL_MULT.bsn	0.57	0.98	0.54	1.97
	PRF.bsn	1.47	0.40	0.17	1.71
	CH_D50.bsn	84.5	60.3	45.0	54.1
	USLE_K.sol	0.64	0.65	0.47	0.65
	CH_COV1.rte	0.53	0.23	0.68	0.71
	CH_COV2.rte	0.64	0.31	0.65	0.36
	USLE_C.dat	0.49	0.40	0.34	0.42
TN	BIOMIX.mgt	0.13	0.68	0.15	0.67
	LAT_ORGN.gw	104.9	191.9	58.7	26.2
	RCN.bsn	3.68	4.85	11.0	9.68
	N_UPDIS.bsn	12.4	17.0	74.8	85.3
	NPERCO.bsn	0.59	0.17	0.86	0.29
	CMN.bsn	0.003	0.003	0.002	0.002
	CDN.bsn	2.05	0.60	2.66	0.60
	SDNCO.bsn	0.35	0.94	0.76	0.16
	ERORGN.hru	0.41	4.91	0.13	0.25
	RS4.swq	0.04	0.01	0.09	0.05
	BC3.swq	0.35	0.25	0.22	0.35
TP	LAT_ORGP.gw	33.1	8.50	2.10	1.00
	P_UPDIS.bsn	30.1	30.0	6.55	0.90
	PPERCO.bsn	15.1	17.1	16.3	12.5
	PSP.bsn	0.56	0.61	0.36	0.69
	ERORGP.hru	0.03	0.89	0.05	0.48
	RS5.swq	0.04	0.04	0.09	0.03
	BC4.swq	0.46	0.49	0.30	0.11

Table S4. Descriptive statistics of observations and training data.

Variables	WS*	Unit	Min	Mean	Max	Std. Dev. **	CoV (%)***	Number of observations
P	-	mm	0.00	2.98	135.0	9.66	324.2	1,634
AT	-	°C	-9.50	15.0	31.5	9.34	62.3	1,634
WS	-	m/s	0.40	1.71	5.30	0.74	43.3	1,634
RH	-	%	23.9	68.7	99.0	14.6	21.3	1,634
SR	-	MJ/m^2	0.0	14.1	32.1	7.06	50.1	1,634
E	-	mm	0.29	3.66	9.60	2.02	55.2	1,634
Flow	WJ	m^3/s	0.09	0.68	27.7	1.89	276.9	1,634
	HN		0.22	1.21	36.7	2.65	219.6	1,634
	JS		0.00	0.16	8.91	0.53	336.9	1,634
	PYJ		0.24	1.69	70.0	3.75	221.6	1,634
SS	WJ	mg/L	1.73	25.1	244.0	33.6	133.6	121
	HN		3.21	28.8	236.4	39.0	135.5	109
	JS		3.35	100.2	1,110.0	209.1	208.8	109
	PYJ		1.70	25.1	384.6	43.7	174.1	229
TN	WJ	mg/L	0.08	2.29	6.52	0.83	36.1	121
	HN		0.98	2.39	7.86	0.87	36.3	109
	JS		1.17	3.01	6.72	0.84	27.8	109
	PYJ		0.70	2.19	5.74	0.69	31.4	229
TP	WJ	mg/L	0.01	0.17	1.70	0.16	94.0	121
	HN		0.04	0.18	1.13	0.13	72.7	109
	JS		0.02	0.20	0.82	0.12	61.2	109
	PYJ		0.04	0.14	0.72	0.10	69.1	229

* WS: Study Watershed, ** Std. Dev.: Standard Deviation, *** CoV: Coefficient of Variation.

Table S5. Information use efficiency achieved by ML models trained with the different combinations of training data sets (unit: none or fraction). The highest efficiency statistics are in bold.

ML Models	Training Data Sets	Marginal entropy				Transfer entropy			
		Flow	**SS**	**TN**	**TP**	**Flow**	**SS**	**TN**	**TP**
RF	WD+UC	0.009	0.003	0.011	**0.023**	**5.583**	1.645	**4.674**	2.473
	WD+C	**0.017**	0.002	**0.025**	0.021	1.240	1.046	2.994	**4.161**
	All	0.014	**0.015**	0.012	0.017	1.571	**2.137**	3.368	1.562
SVM	WD+UC	0.031	0.062	0.012	0.019	**9.222**	**3.913**	1.807	2.936
	WD+C	**0.048**	**0.140**	**0.031**	**0.036**	4.033	3.527	**2.825**	**7.264**
	All	0.027	0.061	0.016	0.030	2.046	3.654	2.174	3.435
ANN	WD+UC	0.010	0.021	0.041	0.021	0.777	2.004	2.215	0.564
	WD+C	**0.018**	**0.074**	**0.070**	**0.053**	**1.470**	**2.033**	**4.453**	**3.007**
	All	0.011	0.036	0.033	0.031	0.968	1.513	4.063	2.417

Table S6. Summary statistics of transfer entropy (unit: bits) of the training data sets by the watersheds with different machine learning algorithms.

Watershed	Training Data Sets	Flow			SS			TN			TP		
		RF	SVM	ANN	RF	SVM	ANN	RF	SVM	ANN	RF	SVM	ANN
WJ	WDO	0.299	0.315	0.265	0.385	0.210	0.174	0.215	0.240	0.276	0.174	0.281	0.246
	WD+UC	0.304	0.348	0.451	0.190	0.297	0.214	0.223	0.214	0.388	0.207	0.314	0.315
	WD+C	0.394	0.486	0.485	0.294	0.415	0.381	0.263	0.327	0.312	0.210	0.335	0.340
	All	0.424	0.529	0.469	0.535	0.539	0.514	0.253	0.315	0.324	0.404	0.367	0.349
HN	WDO	0.290	0.328	0.378	0.282	0.173	0.125	0.251	0.173	0.167	0.349	0.375	0.101
	WD+UC	0.395	0.345	0.467	0.291	0.449	0.218	0.325	0.284	0.323	0.382	0.533	0.535
	WD+C	0.439	0.509	0.457	0.461	0.433	0.432	0.312	0.420	0.274	0.387	0.488	0.456
	All	0.472	0.637	0.535	0.527	0.536	0.487	0.387	0.528	0.353	0.620	0.598	0.606
JS	WDO	0.354	0.355	0.333	0.333	0.225	0.213	0.368	0.379	0.173	0.365	0.258	0.137
	WD+UC	0.337	0.415	0.451	0.358	0.325	0.423	0.403	0.499	0.370	0.435	0.364	0.370
	WD+C	0.481	0.470	0.482	0.447	0.511	0.394	0.485	0.531	0.439	0.409	0.343	0.285
	All	0.502	0.589	0.567	0.375	0.424	0.527	0.425	0.558	0.411	0.443	0.466	0.424
PYJ	WDO	0.314	0.386	0.376	0.345	0.358	0.191	0.281	0.351	0.400	0.285	0.337	0.218
	WD+UC	0.432	0.459	0.496	0.455	0.419	0.311	0.446	0.454	0.534	0.526	0.430	0.475
	WD+C	0.417	0.429	0.449	0.402	0.450	0.361	0.316	0.400	0.461	0.413	0.358	0.355
	All	0.479	0.581	0.596	0.588	0.563	0.490	0.403	0.457	0.531	0.524	0.461	0.439

References

Basu, N. B., Destouni, G., Jawitz, J. W., Thompson, S. E., Loukinova, N. V., Darracq, A., Zanardo, S., Yaeger, M., Sivapalan, M., Rinaldo, A., and Rao, P. S. C.: Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity, Geophys. Res. Lett., 37, L23404, https://doi.org/10.1029/2010GL045168, 2010.

Lee, C. J., Hirsch, R. M., Schwarz, G. E., Holtschlag, D. J., Preston, S. D., Crawford, C. G., and Vecchia, A. V.: An evaluation of methods for estimating decadal stream loads, J. Hydrol., 542, 185-203, https://doi.org/10.1016/j.jhydrol.2016.08.059, 2016.

Song, J.-H., Her, Y., and Guo, T.: Quantifying the contribution of direct runoff and baseflow to nitrogen loading in the Western Lake Erie Basins, Sci. Rep., 12, 9216, 10.1038/s41598-022-12740-1, 2022.

Song, J.-H., Her, Y., Park, Y. S., Yoon, K., and Kim, H.: Investigating the applicability and assumptions of the regression relationship between flow discharge and nitrogen concentrations for load estimation, Heliyon, 10, e23603, https://doi.org/10.1016/j.heliyon.2023.e23603, 2024.

Wainwright, J., Parsons, A. J., Cooper, J. R., Gao, P., Gillies, J. A., Mao, L., Orford, J. D., and Knight, P. G.: The concept of transport capacity in geomorphology, Rev. Geophys., 53, 1155-1202, https://doi.org/10.1002/2014RG000474, 2015.

Wu, D., Cao, M., Gao, W., Cheng, G., Duan, Z., Hou, X., and Zhang, Y.: Spatial-temporal source apportionment of nitrogen and phosphorus in a high-flow variable river, J. Hydrol.: Reg. Stud., 53, 101839, https://doi.org/10.1016/j.ejrh.2024.101839, 2024.