Supplementary Information

Identifying Trustworthiness Challenges in Deep Learning Models for Continental-Scale Water Quality Prediction

Xiaobo Xia[1,2] **Xiaofeng Liu**[3,4] **Jiale Liu**[5] **Kuai Fang**[6]
Lu Lu[2] **Samet Oymak**[7] **William S. Currie**[3,4] **Tongliang Liu**[1]

[1]School of Computer Science, University of Sydney, Sydney, NSW 2008, Australia

[2]Department of Statistics and Data Science, Yale University, New Haven, CT 06511, USA

[3]Michigan Institute for Data and AI in Society, University of Michigan, Ann Arbor, MI 48105, USA

[4]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109, USA

[5]College of Information Science and Technology, Penn State University, University Park, PA 16802, USA

[6]Department of Earth System Science, Stanford University, Stanford, CA 94305, USA

[7]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

{XIAOBOXIA.UNI@GMAIL.COM XIAOFENG.LIU0202@GMAIL.COM}

[†]Corresponding author.

Table of content

Supplementary figures

Supplementary tables

- **Table S1.** Summary of the studied water quality variables and the average number of observations per basin, based on 482 U.S. rivers between 01/01/1982 and 12/31/2018.
- **Table S2.** Model input features, consisting of 25 time series variables and 49 static basin attributes (sourced from the GAGES-II database).

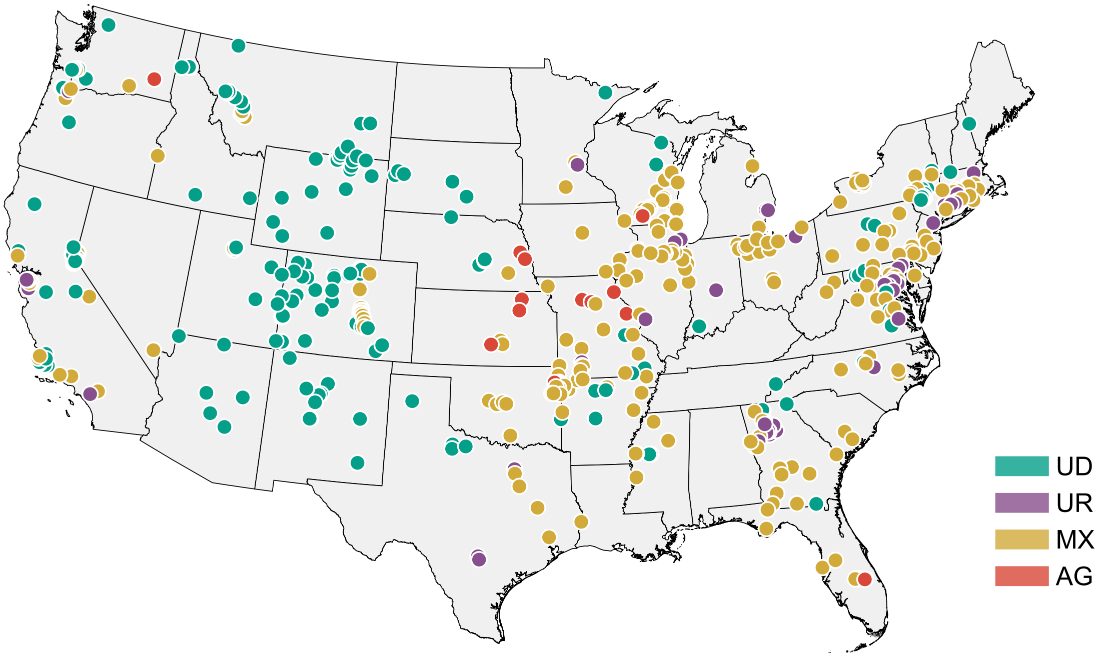

Fig. S1. Spatial distribution of studied basins classified by land uses. Following the USGS classification criteria [1], agricultural basins (AG, red) are defined as having more than 50% agricultural land (PLANTNLCD06 in the GAGES-II database) and at most 5% urban land (DEVNLCD06). Undeveloped basins (UD, green) have at most 5% urban land and at most 25% agricultural land. Urban basins (UR, purple) are defined as having more than 25% urban land and at most 25% agricultural land, while mixed basins (MX, yellow) include all other combinations of urban, agricultural, and undeveloped land. Among the selected basins, 3.1% were classified as AG, 11.2% as UR, 35.1% as UD, and 50.6% as MX.

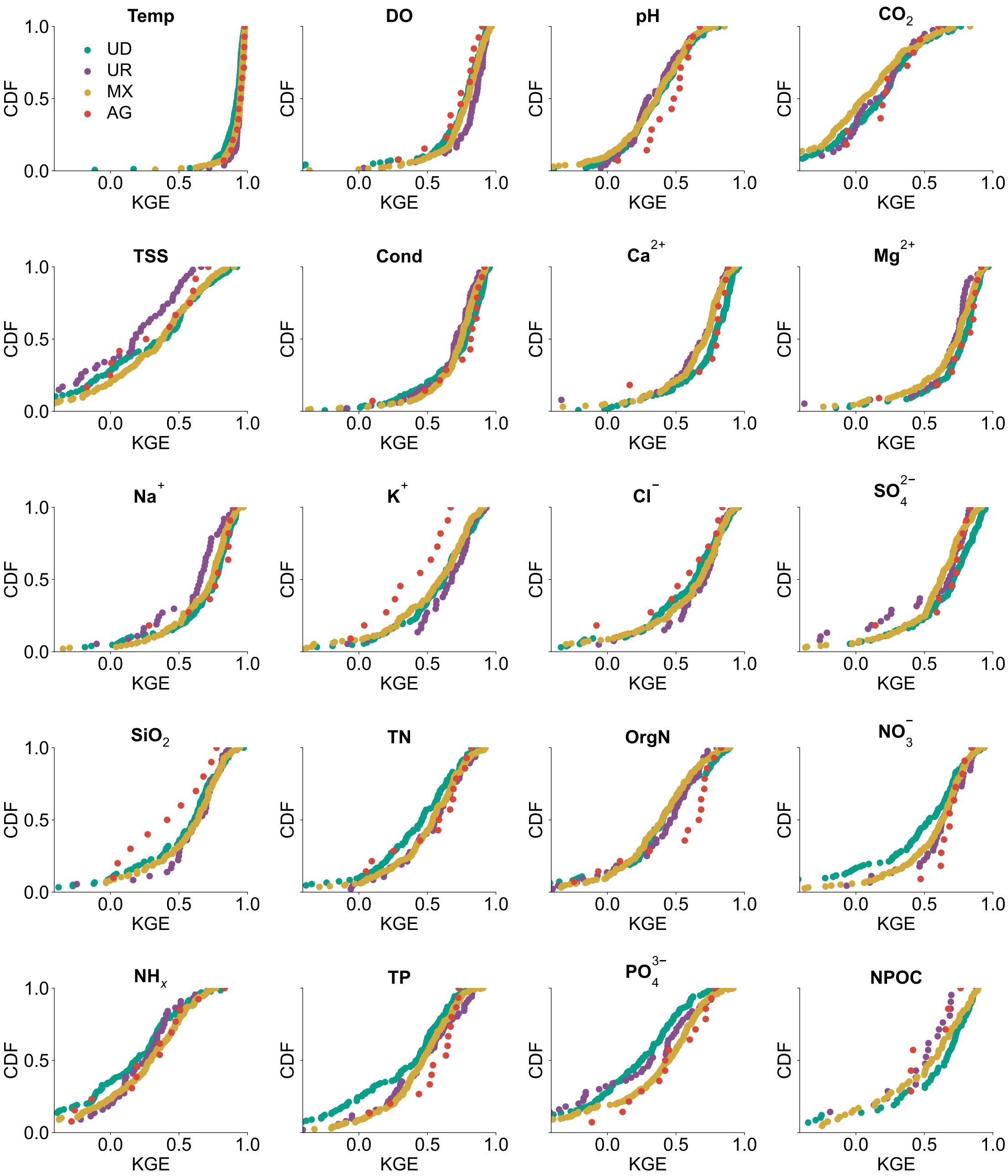

Fig. S2. Multi-task LSTM model performances across undeveloped basins (UD), urban basins (UR), mixed basins (MX), and agricultural basins (AG), are shown as the cumulative distribution function (CDF) of the KGE. Curves that remain lower demonstrate better performance.

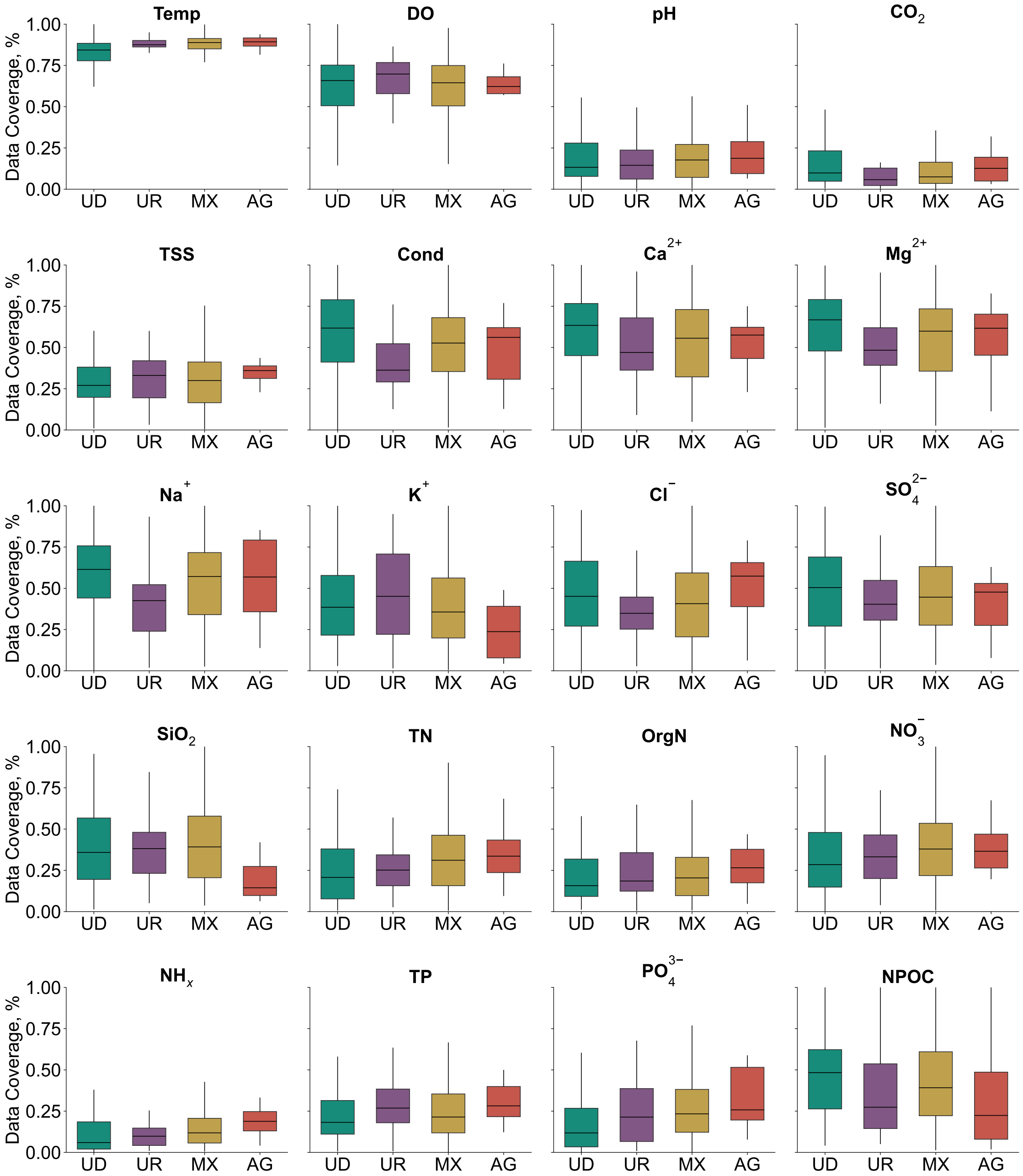

Fig. S3. Water quality data coverage (%) across basins of different land use types, computed as the ratio of days monitored to the total number of days between 01/01/1982 and 12/31/2018. A coverage of 100% indicates that water quality measurements were available for the entire study period and 0% indicates no measurements were available. The boxplots display the median (central line), interquartile range (IQR, represented by the boxes spanning the first (Q1) to the third quartile (Q3)), and whiskers extending to Q1 $-$ 1.5 × IQR and Q3 $+$ 1.5 × IQR.

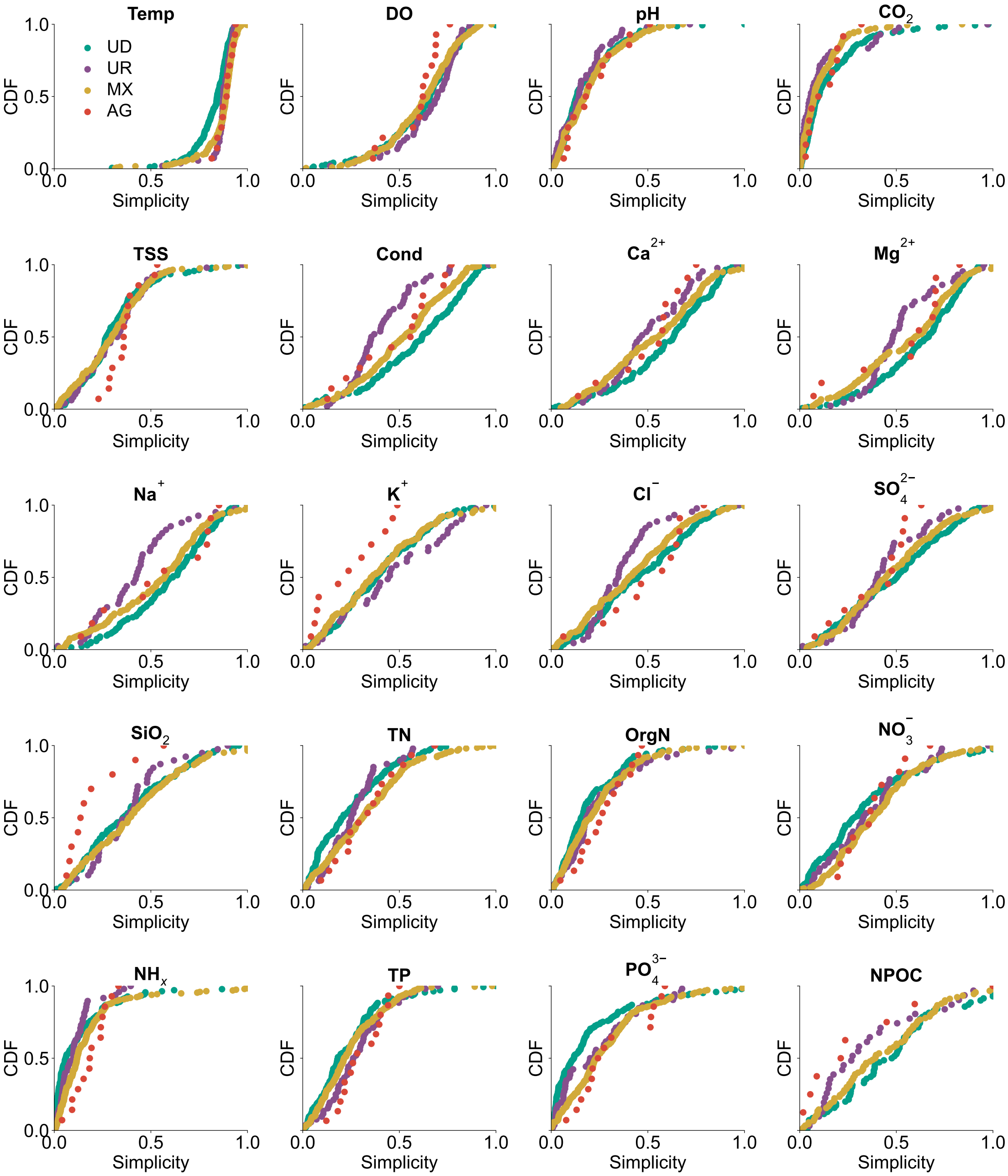

Fig. S4. Simplicity index distributions across undeveloped basins (UD), urban basins (UR), mixed basins (MX), and agricultural basins (AG). The simplicity index (adapted from [2]) quantifies the proportion of variance in water quality dynamics explained by linear relationships with runoff and annual cycles. Lower CDF (cumulative distribution function) curves indicate higher simplicity.

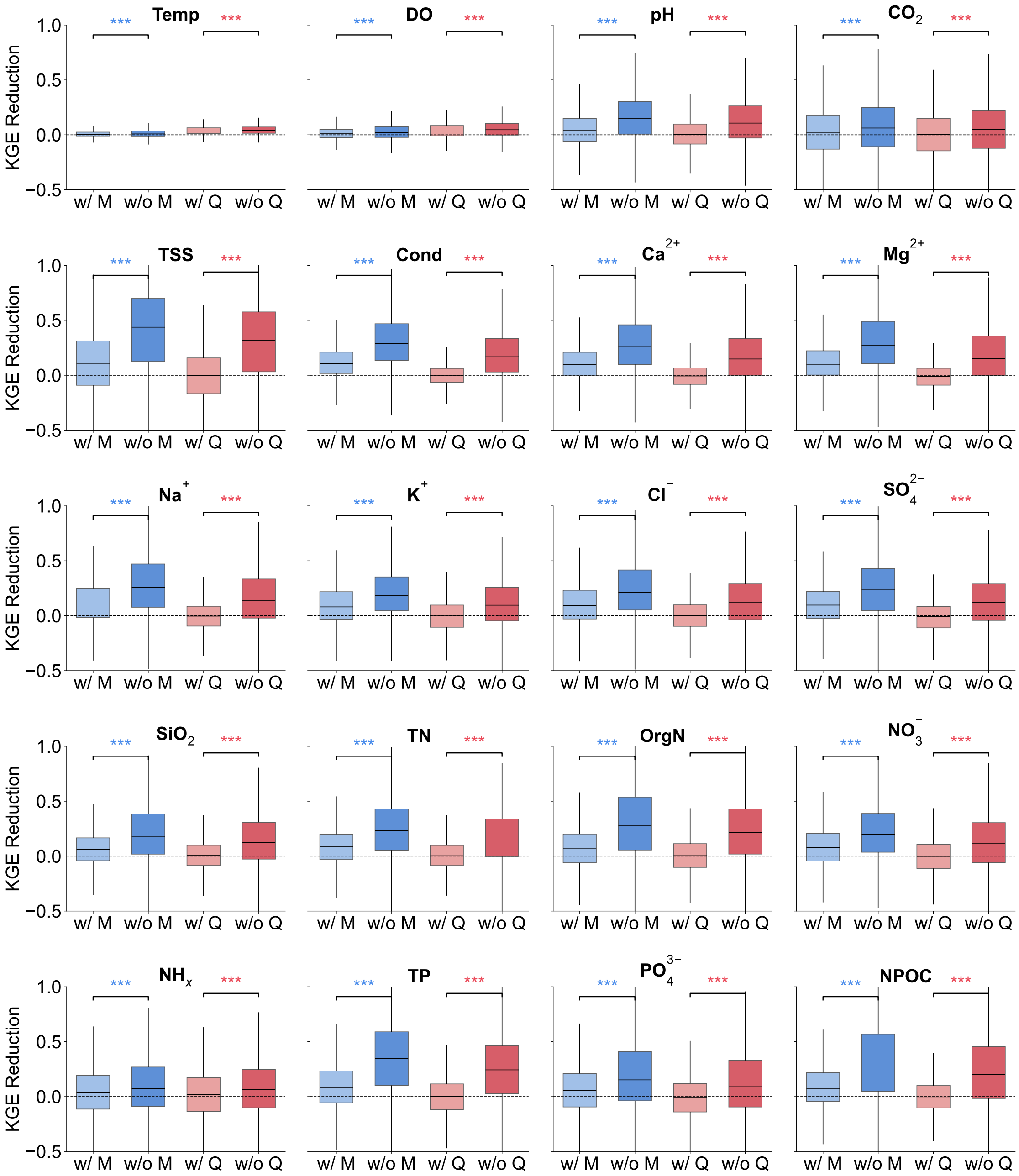

Fig. S5. Context-dependent feature importance (KGE reduction) of meteorological variables (M) and runoff (Q) derived via the Traverse method. Dark blue boxplots represent KGE reduction from excluding Q when M is already excluded, whereas light blue boxplots represent excluding Q when M is included. Similarly, dark red boxplots show the KGE reduction from excluding M when Q is absent, whereas light red boxplots represent excluding M when Q is included. Wilcoxon signed-rank tests were conducted to assess whether median KGE reductions from subsets lacking Q or M were significantly greater than those from subsets where Q or M were present ($^{***}p \leq 0.001$). The results indicate that meteorological variables become largely redundant when runoff is included.

Table S1. Summary of the studied water quality variables and the average number of observations per basin, based on 482 U.S. rivers between 01/01/1982 and 12/31/2018.

USGS code	Description	Abbreviation	Unit	# Observations per basin
00010	Water temperature	Temp	°C	330.5
00095	Specific conductance	Cond	uS/cm at 25°C	285.6
00300	Oxygen	DO	mg/L	197.8
00040	pH	pH	-	224.9
00405	Carbon dioxide	CO_2	mg/L	129.2
00600	Total nitrogen	TN	mg/L	193.3
00605	Organic nitrogen	OrgN	mg/L	171.7
00618	Nitrate	NO_3^-	mg/L as N	138.3
00660	Orthophosphate	PO_4^{3-}	mg/L as PO_4^{3-}	204.9
00665	Total phosphorus	TP	mg/L as P	266.9
00681	Organic carbon	NPOC	mg/L	60.3
00915	Calcium	Ca^{2+}	mg/L	131.7
00925	Magnesium	Mg^{2+}	mg/L	131.8
00930	Sodium	Na^+	mg/L	117.3
00935	Potassium	K^+	mg/L	114.8
00940	Chloride	Cl^-	mg/L	184.1
00945	Sulfate	SO_4^{2-}	mg/L	154.3
00955	Silica	SiO_2	mg/L	116.1
71846	Ammonia and ammonium	NH_x (NH_3 and NH_4^+)	mg/L as NH_4^+	184.1
80154	Suspended sediment concentration	TSS	mg/L	305.4

Table S2. Model input features, consisting of 25 time series variables and 49 static basin attributes (sourced from the GAGES-II database).

Group	Name	Type	Description	Unit
Runoff	runoff	time-varying	Area normalized streamflow from USGS	m/y
Meteorological forcings	pr	time-varying	Daily total precipitation	mm/day
	sph	time-varying	Specific humidity	
	srad	time-varying	Surface downwelling solar radiation	W/m^2
	tmmn	time-varying	Daily minimum 2-meter air temperature	F
	tmmx	time-varying	Daily maximum 2-meter air temperature	F
	pet	time-varying	Reference grass evapotranspiration	mm/day
	etr	time-varying	Reference alfalfa evapotranspiration	mm/day
Rainfall chemistry	pH	time-varying	Logarithm of the H ion activity	unitless
	Cond	time-varying	Electrical conductivity of water	μS/cm
	Ca^{2+}	time-varying	Ca ion concentration	mg/L
	Mg^{2+}	time-varying	Mg ion concentration	mg/L
	K^{+}	time-varying	K ion concentration	mg/L
	Na^{+}	time-varying	Na ion concentration	mg/L
	NH$_4$	time-varying	NH$_4$ concentration	mg/L
	NO$_3$	time-varying	NO$_3$ concentration	mg/L
	Cl^{-}	time-varying	Cl ion concentration	mg/L
	SO$_4$	time-varying	SO$_4$ concentration	mg/L
	distNTN	time-varying	The distance to the nearest NTN sampling site	km
Vegetation indices	LAI	time-varying	Leaf area index of vegetation	$\mathrm{m}^2/\mathrm{m}^2$
	FAPAR	time-varying	Fraction of absorbed photosynthetically active radiation	unitless
	NPP	time-varying	Net primary production	$\mathrm{gC}/\mathrm{m}^2/\mathrm{day}$
Time variables	datenum	time-varying	The number of days relative to January 1, 2000	unitless
	sinT	time-varying	Sine of datenum	unitless
	cosT	time-varying	Cosine of datenum	unitless
Basic characteristics	HYDRO_DISTURB_INDX	static	Hydrologic "disturbance index" score, based on 7 variables: 1) MAJ_DDENS_2009, 2) WATER_WITHDR, 3) change in dam storage 1950-2009, 4) CANALS_PCT, 5) RAW_DIS_NEAREST_MAJ_NPDES, 6) ROADS_KM_SQ_KM, and 7) FRAGUN_BASIN	unitless
	BAS_COMPACTNESS	static	Watershed compactness ratio, = area/perimeter^2 * 100; higher number = more compact shape	unitless
	DRAIN_SQKM	static	Watershed drainage area, sq km, as delineated in our basin boundary	km^2
Geology	GEOL_REEDBUSH_DOM	static	Dominant (highest percent of area) geology, derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States	unitless
	GEOL_REEDBUSH_DOM_PCT	static	Percentage of the watershed covered by the dominant geology type	percentage
Hydrologic characteristics	STREAMS_K_S_KM	static	Stream density, km of streams per watershed sq km, from NHD 100k streams	$\mathrm{km}/\mathrm{km}^2$
	STRAHLER_MAX	static	Maximum Strahler stream order in the watershed, from NHDPlus	unitless
	MAINSTEM_SINUOUSITY	static	Sinuosity of mainstem stream line, from our delineation of mainstem stream lines. Defined as curvilinear length of the mainstem stream line dividedby the straight-line distance between the end points of the line.	unitless
	BFI_AVE	static	Base Flow Index (BFI). The BFI is a ratio of base flow to total streamflow, expressed as a percentage and ranging from 0 to 100. Base flow is the sustained, slowly varying component of streamflow, usually attributed to ground-water discharge to a stream.	percentage
	CONTACT	static	Subsurface flow contact time index	days
	PCT_1ST_ORDER	static	Percent of stream lengths in the watershed which are first-order streams (Strahler order); from NHDPlus & percentage	percentage
	PCT_2ND_ORDER	static	Percent of stream lengths in the watershed which are second-order streams (Strahler order); from NHDPlus & percentage	percentage
	PCT_3RD_ORDER	static	Percent of stream lengths in the watershed which are third-order streams (Strahler order); from NHDPlus & percentage	percentage
	PCT_4TH_ORDER	static	Percent of stream lengths in the watershed which are fourth-order streams (Strahler order); from NHDPlus & percentage	percentage
	PCT_5TH_ORDER	static	Percent of stream lengths in the watershed which are fifth-order streams (Strahler order); from NHDPlus & percentage	percentage
	PCT_6TH_ORDER_OR_MORE	static	Percent of stream lengths in the watershed which are sixth or greater-order streams (Strahler order); from NHDPlus & percentage	percentage

Group	Name	Type	Description	Unit
Historical and current dams information	DDENS_2009	static	Dam density; number per 100 km sq	number of dams/100 km^2
	STOR_NOR_2009	static	Dam storage in watershed (“NORMAL_STORAGE”); megaliters total storage per sq km (1 megalitres = 1,000,000 liters = 1,000 cubic meters)	megaliters/km^2
NPDES	NPDES_MAJ_DENS	static	Density of NPDES (National Pollutant Discharge Elimination System) “major” point locations in the watershed; number per 100 km sq. Major locations are defined by an EPA-assigned major flag. From the download of NPDES national database summer 2006.	number of sites/100km^2
Percentages of land cover 2006 in the watershed and lanscape	DEVNLCD06	static	Watershed percent “developed” (urban), 2006 era (2001 for AK-HI-PR). Sum of classes 21, 22, 23, and 24.	percentage
	FORESTNLCD06	static	Watershed percent “forest”, 2006 era (2001 for AK-HI-PR). Sum of classes 41, 42, and 43.	percentage
	PLANTNLCD06	static	Watershed percent “planted/cultivated” (agriculture), 2006 era (2001 for AK-HI-PR). Sum of classes 81 and 82.	percentage
	WATERNLCD06	static	Watershed percent Open Water (class 11)	percentage
	WOODYWETNLCD06	static	Watershed percent Woody Wetlands (class 90)	percentage
	EMERGWETNLCD06	static	Watershed percent Emergent Herbaceous Wetlands (class 95)	percentage
Nitrogen and phosphorus application rate in the watershed	NITR_APP_KG_SQKM	static	Estimate of nitrogen from fertilizer and manure, from Census of Ag 1997, based on county-wide sales and percent agricultural land cover in the watershed.	kg/km^2
	PHOS_APP_KG_SQKM	static	Estimate of nitrogen from fertilizer and manure, from Census of Ag 1997, based on county-wide sales and percent agricultural land cover in the watershed.	kg/km^2
Pesticide	PESTAPP_KG_SQKM	static	Estimate of agricultural pesticide application (219 types), kg/sq km, from Census of Ag 1997, based on county-wide sales and percent agricultural land cover in the watershed	kg/km^2
Regions	ECO2_BAS_DOM	static	Dominant (highest % of the area) Level II ecoregion within the watershed. See X_Region_Names sheet for crosswalk to name.	unitless
	ECO3_BAS_DOM	static	Dominant (highest % of the area) Level III ecoregion within the watershed. See X_Region_Names sheet for crosswalk to name.	Level III ecoregion (1-84)
	NUTR_BAS_DOM	static	Dominant (highest % of the area) nutrient ecoregion within the watershed. See X_Region_Names sheet for crosswalk to name.	Nutrient ecoregion (1-14)
	HLR_BAS_DOM_100M	static	Dominant (highest % of the area) Hydrologic Landscape Region within the watershed. See X_Region_Names sheet for crosswalk to name.	HLR region (1-20)
	PNV_BAS_DOM	static	Dominant (highest % of the area) Potential Natural Vegetation (PNV) within the watershed. See X_Region_Names sheet for crosswalk to name.	PNV type (1-63)
Soil	AWCAVE	static	Average value for the range of available water capacity for the soil layer or horizon (inches of water per inch of soil depth)	unitless
	PERMAVE	static	Average permeability (inches/hour)	inches/hour
	BDAVE	static	Average value of bulk density (grams per cubic centimeter)	grams per cubic centimeter
	OMAVE	static	Average value of organic matter content (percent by weight)	percentage
	WTDEPAVE	static	Average value of depth to seasonally high water table (feet)	feet
	ROCKDEPAVE	static	Average value of total soil thickness examined (inches)	inches
	CLAYAVE	static	Average value of clay content (percentage)	percentage
	SILTAVE	static	Average value of silt content (percentage)	percentage
	KFACT_UP	static	Average K-factor value for the uppermost soil horizon in each soil component. K-factor is an erodibility factor which quantifies the susceptibility of soil particles to detachment and movement by water. The K-factor is used in the Universal Soil Loss Equation (USLE) to estimate soil loss by water. Higher values of the K-factor indicate greater potential for erosion	unitless
	RFACT	static	Rainfall and Runoff factor (“R factor” of Universal Soil Loss Equation); average annual value for the period 1971-2000.	100s ft-tonf in/h/ac/yr
Topographic characteristics	ELEV_MEAN_M_BASIN	static	Mean watershed elevation (meters) from 100m National Elevation Dataset	m
	SLOPE_PCT	static	Mean watershed slope, percent. Derived from 100m resolution National Elevation Dataset, so slope values may differ from those calculated from data of other resolutions.	percentage
	ASPECT_DEGREES	static	Mean watershed aspect, degrees (degrees of the compass, 0-360). Derived from 100m resolution National Elevation Data. 0 and 360 point to north, because of the national Albers projection actual aspect may vary.	degrees (0-360)
Latitude and Longitude	LAT_GAGE	static	Latitude at gage, decimal degrees	decimal degrees, datum NAD83
	LNG_GAGE	static	Longitude at gage, decimal degrees	decimal degrees, datum NAD83
Snow	SNOW_PCT_PRECIP	static	Snow percent of total precipitation estimate, mean for period 1901-2000. From McCabe and Wolock (submitted, 2008), 1km grid.	percentage

References

[1] Spahr, N. E., Dubrovsky, N. M., Gronberg, J. M., Franke, O. L. & Wolock, D. M. Nitrate loads and concentrations in surface-water base flow and shallow groundwater for selected basins in the united states, water years 1990-2006. Tech. Rep., US Geological Survey (2010).

[2] Fang, K., Caers, J. & Maher, K. Modeling continental us stream water quality using long-short term memory and weighted regressions on time, discharge, and season. *Frontiers in Water* **6**, 1456647 (2024).