Supplement of Earth Syst. Sci. Data, 17, 2713–2733, 2025
https://doi.org/10.5194/essd-17-2713-2025-supplement
© Author(s) 2025. CC BY 4.0 License.

Open Access Earth System Science Data

Supplement of

Transformation rate maps of dissolved organic carbon in the contiguous US

Lingbo Li et al.

Correspondence to: Hong-Yi Li (hongyili.jadison@gmail.com)

The copyright of individual parts of the supplement might differ from the article licence.

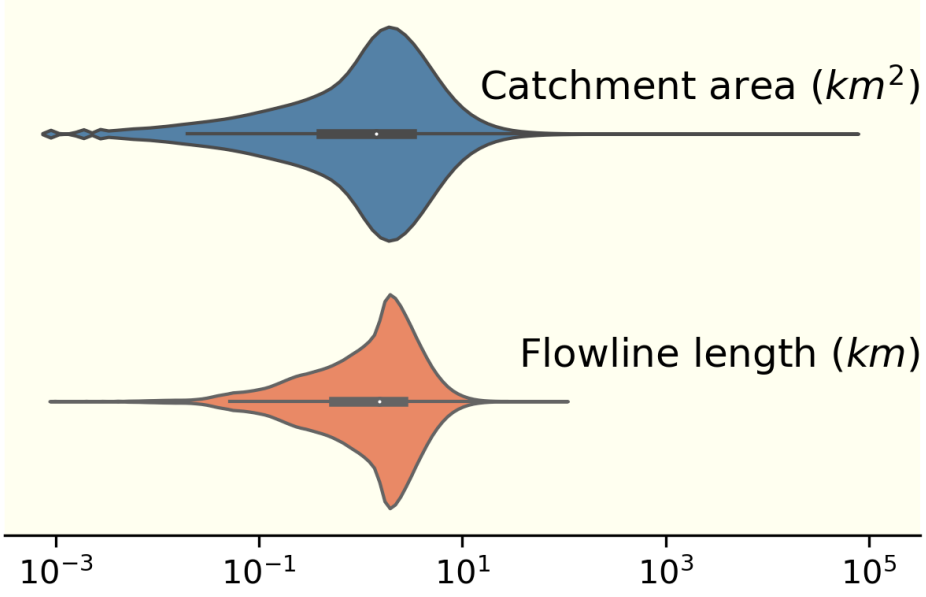

Figure S1. Distribution of the 2.6 million NHDPlus local catchment areas and flowline lengths.

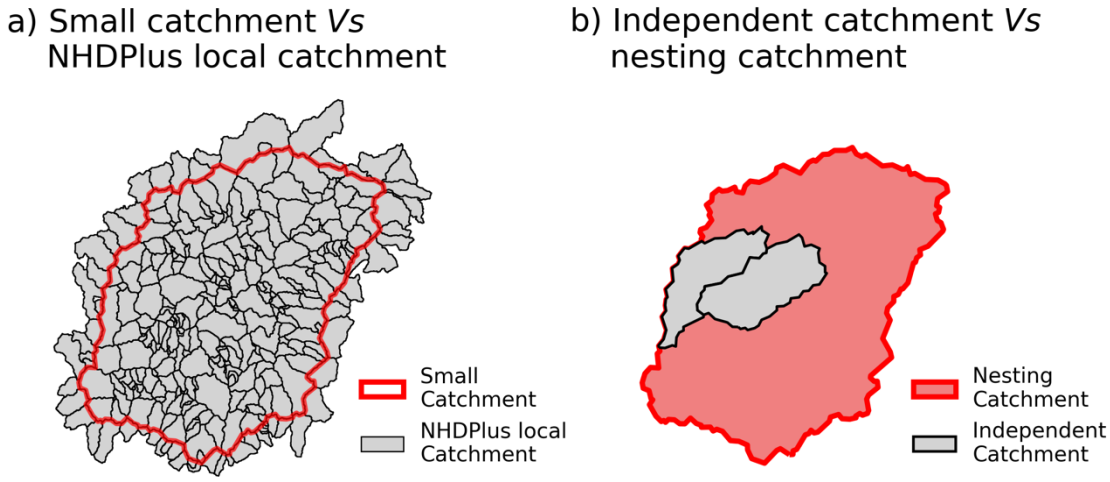

Figure S2. Comparison of catchment relationships: a) a small catchment and its containing NHDPlus local catchments, and b) independent catchment and nesting catchment. Note: The catchments outlined in red represent the same area, but only the boundary is shown in subplot a) for better visual clarity.

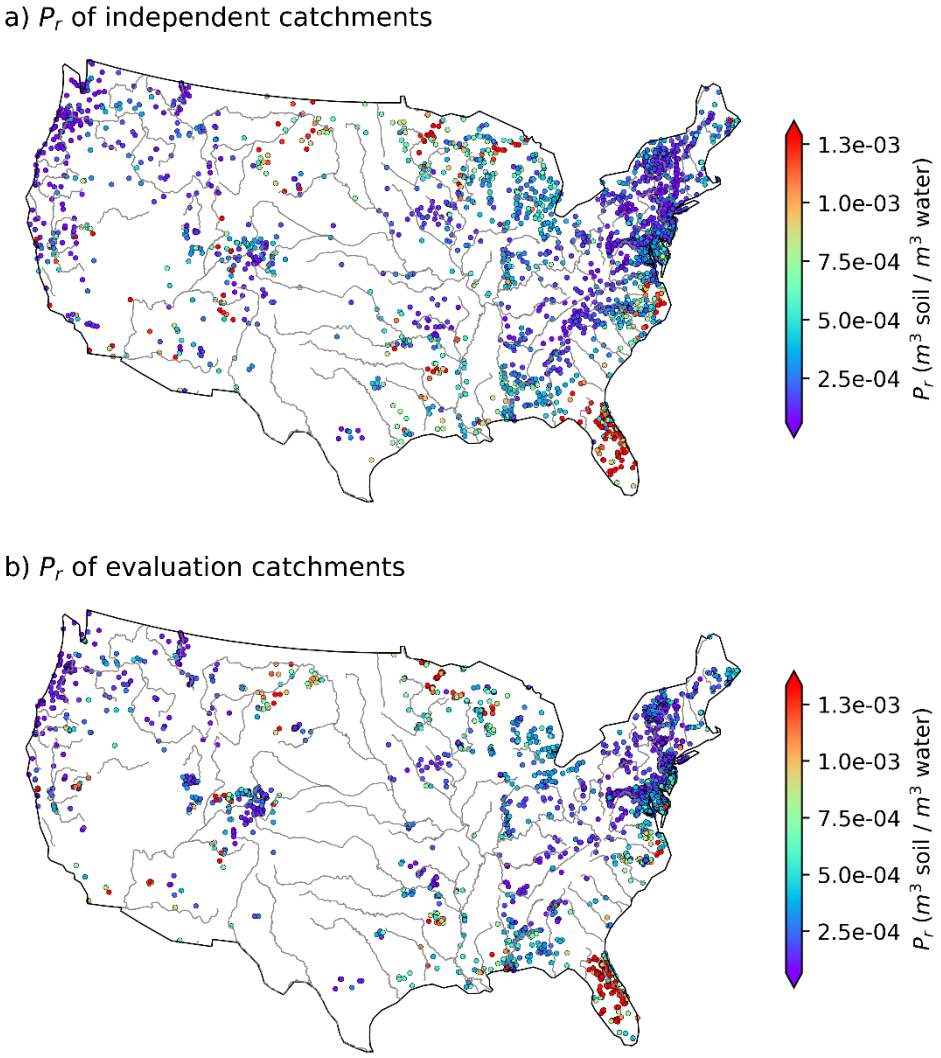

Figure S3. Variability in estimated P_r (using SoilGrids 2.0) across CONUS: a) For independent catchments (n=2583), and b) For evaluation catchments (n=3210). The points indicate the locations of the WQP stations, which are also the outlets of the corresponding small catchments. The CONUS boundary and river shapefiles are directly from open-source datasets GeoPandas (geopandas.org) and Natural Earth (Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com), respectively. The color bars have been adjusted to enhance visual display by showing only the main body of values (from the 5th percentile to the 95th percentile).

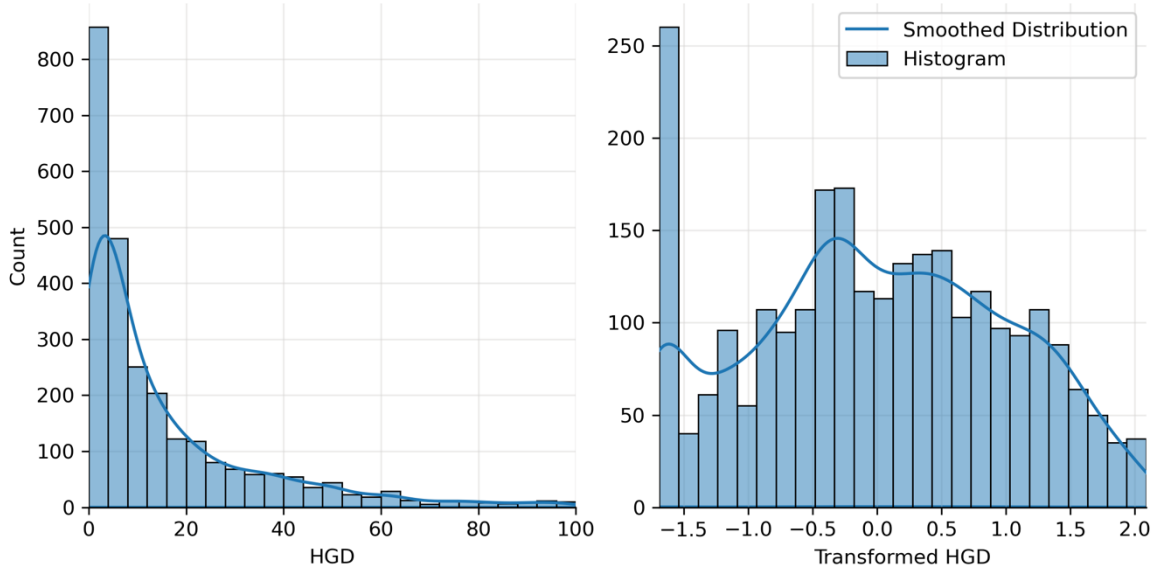

Figure S4. Histogram of the percentage of hydrologic group D soil (HGD) predictor before (left panel) and after (right panel) power transformation.

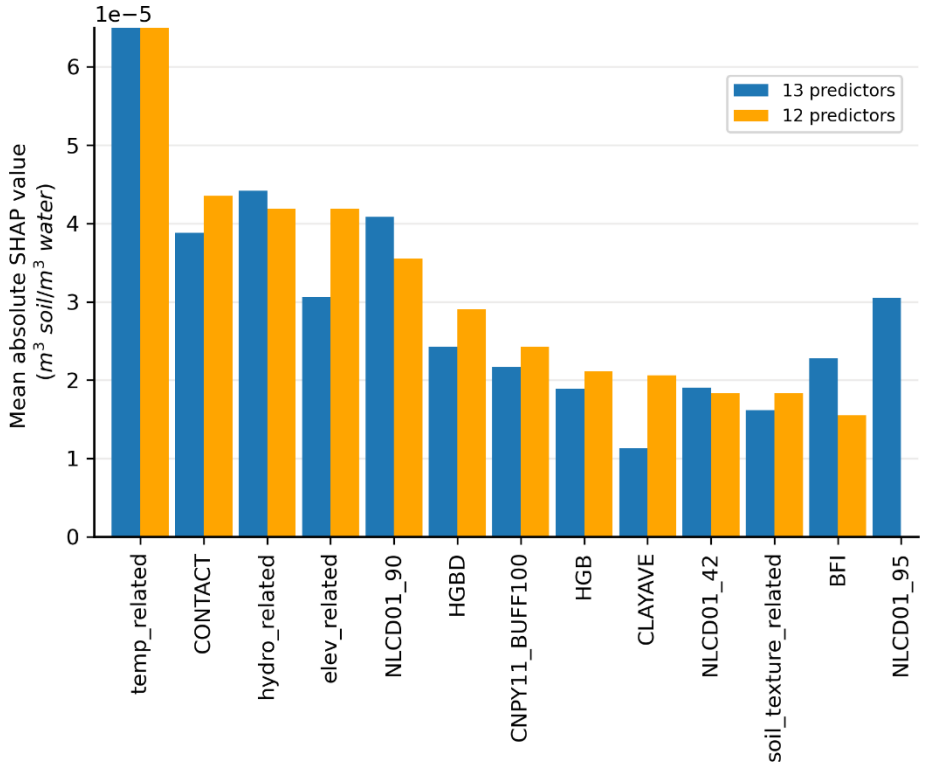

Figure S5. Mean absolute SHAP values of predictors in models (using SoilGrids 2.0) with 13 predictors (blue) and 12 predictors (orange). Note that the SHAP values have the same units as the target variable, P_r. Abbreviations: temp_related (merged predictor encompassing potential evapotranspiration, first/last freeze timing, snow fraction, actual evapotranspiration, and mean/min/max temperature); CONTACT (subsurface contact time); hydro_related (merged predictor representing recharge, runoff, and precipitation); elev_related (merged predictor for mean/min/max elevation); NLCD01_90 (areal percentage of woody wetlands); HGBD (areal percentage of Hydrologic Group BD soil); CNPY11_BUFF100 (areal percentage of canopy in the riparian buffer); HGB (areal percentage of Hydrologic Group B soil); CLAYAVE (clay content percentage); NLCD01_42 (areal percentage of evergreen forest); soil_texture_related (merged predictor for silt and sand content); BFI (base flow index); NLCD01_95 (areal percentage of herbaceous wetlands). For detailed descriptions, refer to Supplementary Tables S2 and S3.

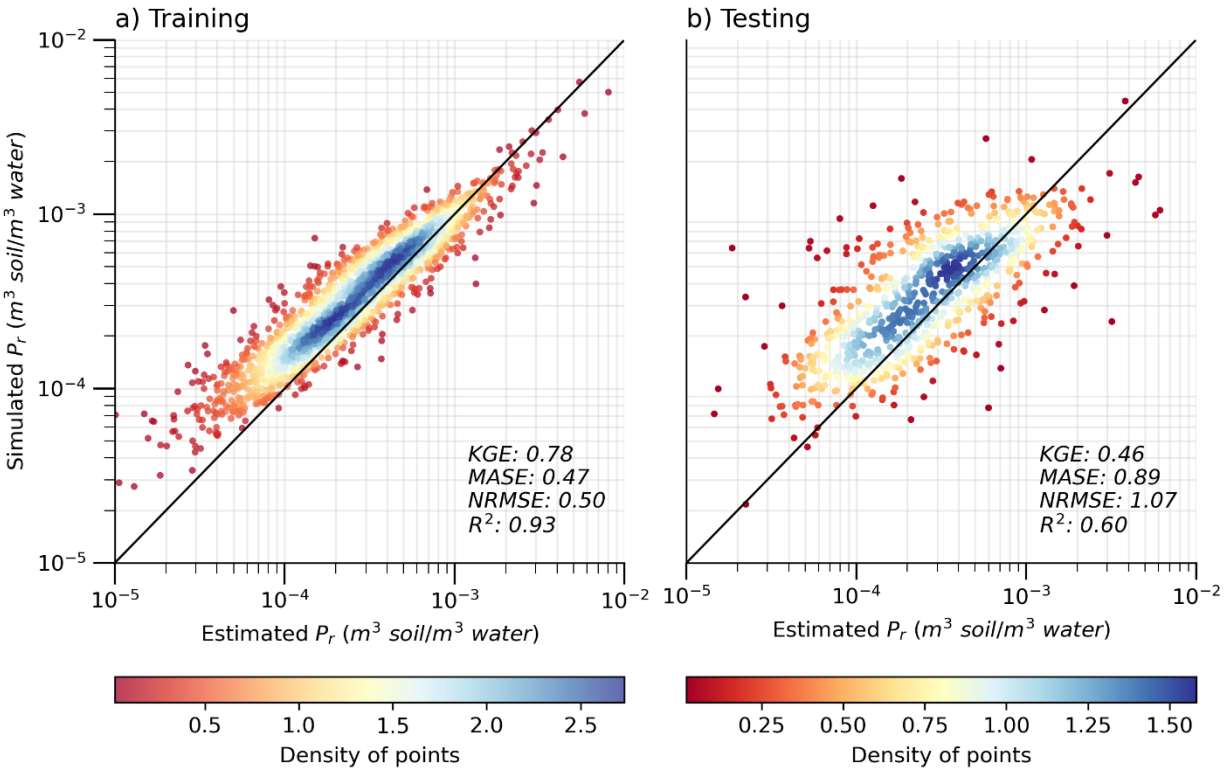

Figure S6. Performance of the XGBoost model (using HWSD) with 12 predictors trained by using KGE during a) the training phase (n=1816) and b) the testing phase (n=779). The solid black line indicates a 1:1 ratio. Note that the axes are in a log-log scale.

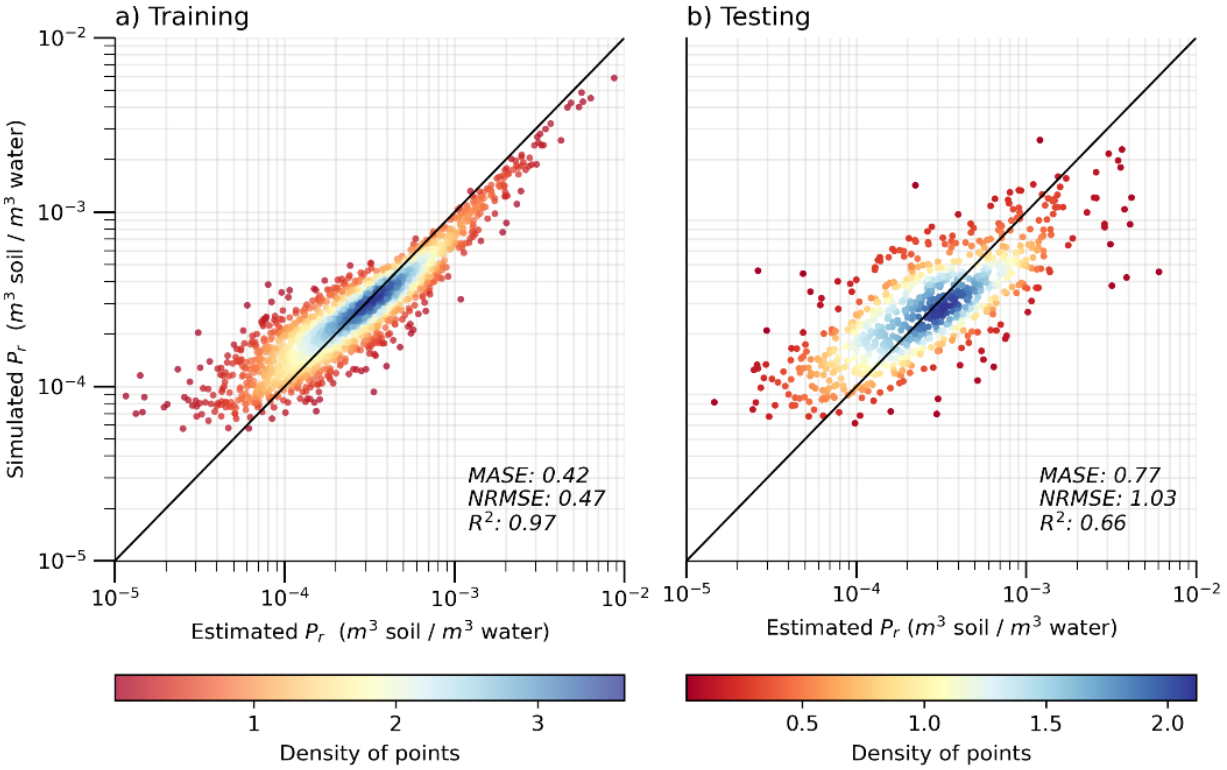

Figure S7. Performance of the XGBoost model (using SoilGrids 2.0) with 12 predictors during a) the training phase (n=1808) and b) the testing phase (n=775). The solid black line indicates a 1:1 ratio. The varying colours indicate the density of points in the scatter plot.

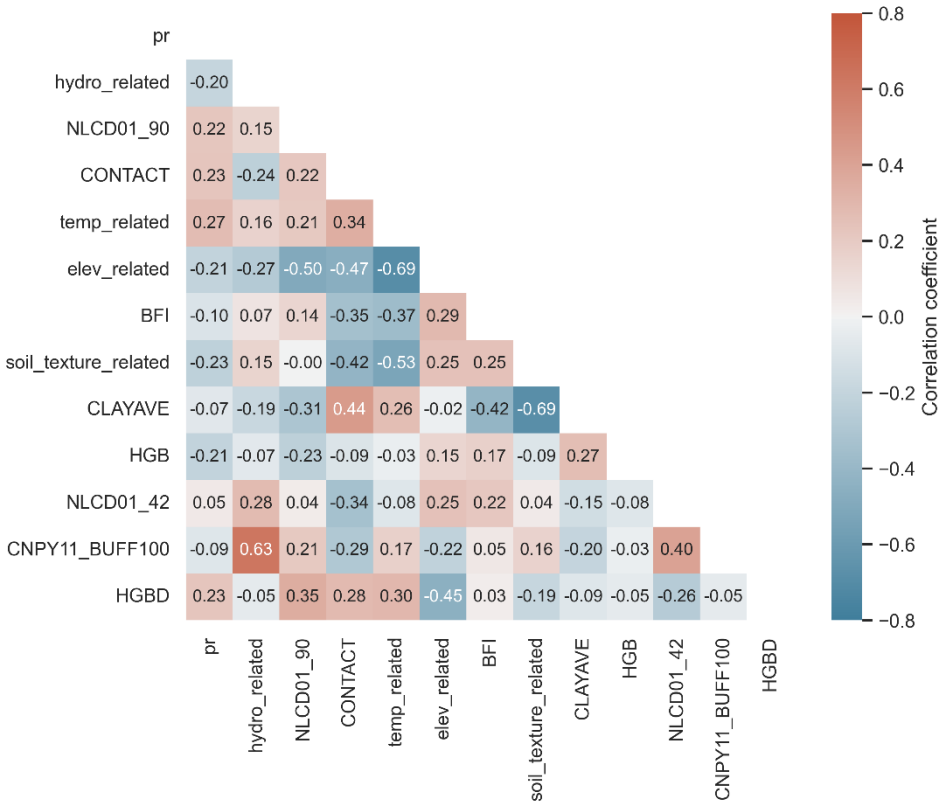

Figure S8. Covariance heatmap of P_r (using SoilGrids 2.0) and the 12 selected NHDPlus predictors. The Pearson correlation coefficient is used. Abbreviations: hydro_related (merged predictor representing recharge, runoff, and precipitation); NLCD01_90 (areal percentage of woody wetlands); CONTACT (subsurface contact time); temp_related (merged predictor encompassing potential evapotranspiration, first/last freeze timing, snow fraction, actual evapotranspiration, and mean/min/max temperature); elev_related (merged predictor for mean/min/max elevation); BFI (base flow index); soil_texture_related (merged predictor for silt and sand content); CLAYAVE (clay content percentage); HGB (areal percentage of Hydrologic Group B soil); NLCD01_42 (areal percentage of evergreen forest); CNPY11_BUFF100 (areal percentage of canopy in the riparian buffer); HGBD (areal percentage of Hydrologic Group BD soil). For detailed descriptions, refer to Supplementary Tables S2 and S3.

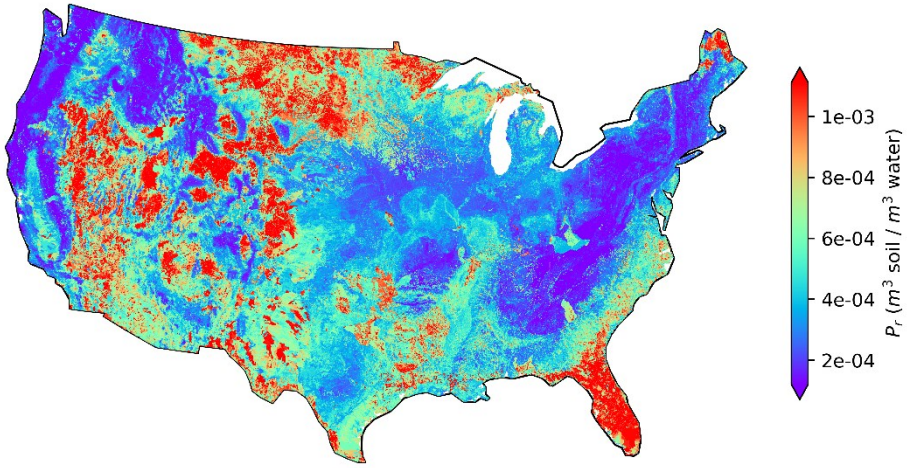

Figure S9. ML model (using SoilGrids 2.0) simulated P_r at over 2.6 million NHDPlus local catchments.

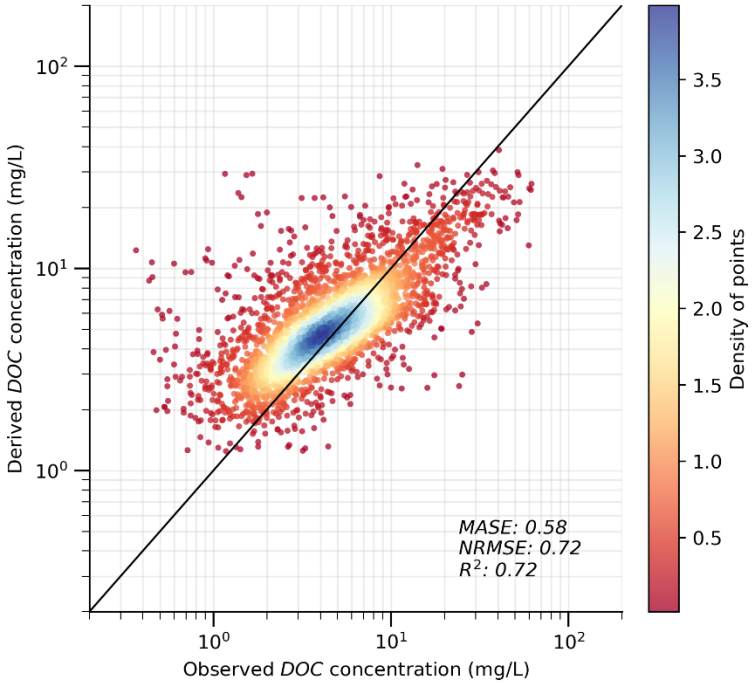

Figure S10. Evaluation of derived DOC (using SoilGrids 2.0) concentration at the catchment scale (n=3208). The solid black line indicates a 1:1 ratio. The varying colours indicate the density of points in the scatter plot.

65

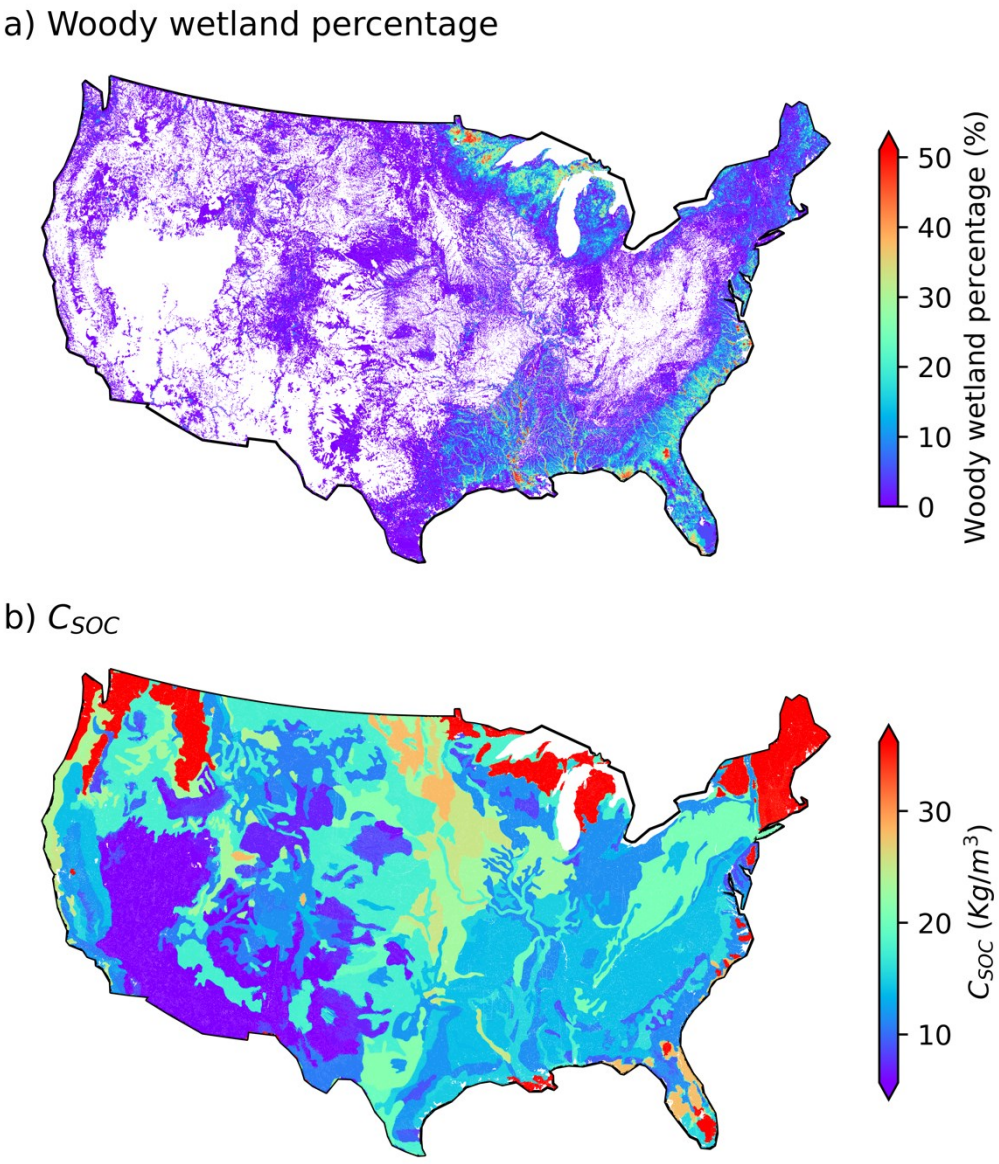

Figure S11. CONUS maps of: a) HWSD top-layer soil organic carbon (SOC) concentration, and b) woody wetland fraction across over 2 million NHDPlus catchments. Regions displayed in white may indicate missing data or a zero fraction.

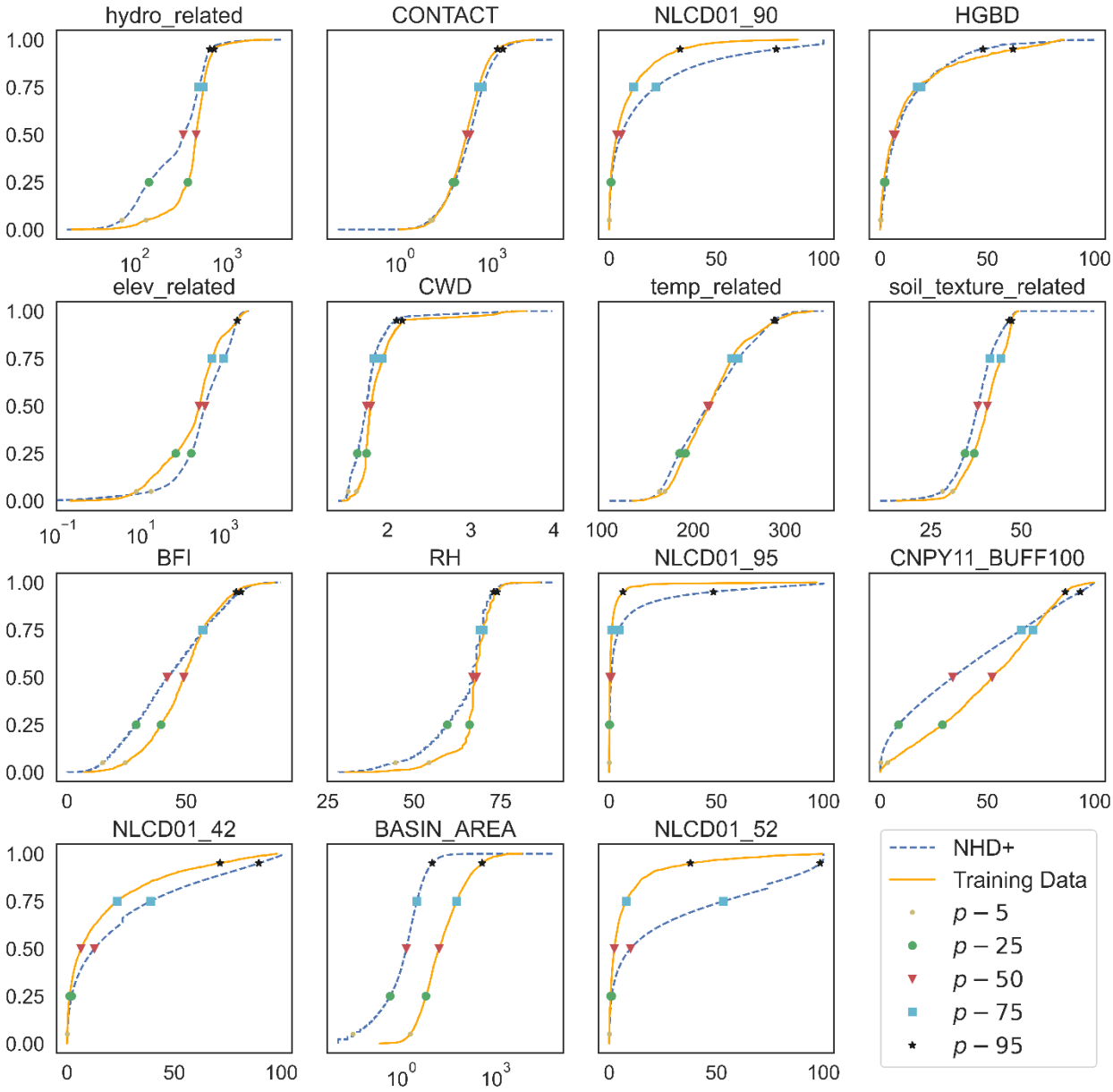

Figure S12. Comparison of the cumulative distribution function (CDF) of 15 selected predictors between training data and all flowlines (i.e., NHDPlus). Abbreviations: hydro_related (merged predictor representing recharge, runoff, and precipitation); CONTACT (subsurface contact time); NLCD01_90 (areal percentage of woody wetlands); HGBD (areal percentage of Hydrologic Group BD soil); elev_related (merged predictor for mean/min/max elevation); CWD (consecutive wet days); temp_related (merged predictor encompassing potential evapotranspiration, first/last freeze timing, snow fraction, actual evapotranspiration, and mean/min/max temperature); soil_texture_related (merged predictor for silt and sand content); BFI (base flow index); RH (relative humidity); NLCD01_95 (areal percentage of herbaceous wetlands); CNPY11_BUFF100 (areal percentage of canopy in the riparian buffer); NLCD01_42 (areal percentage of evergreen forest); BASIN_AREA (catchment area); NLCD01_52 (areal percentage of shrub). For detailed descriptions, refer to Supplementary Tables S1 and S2.

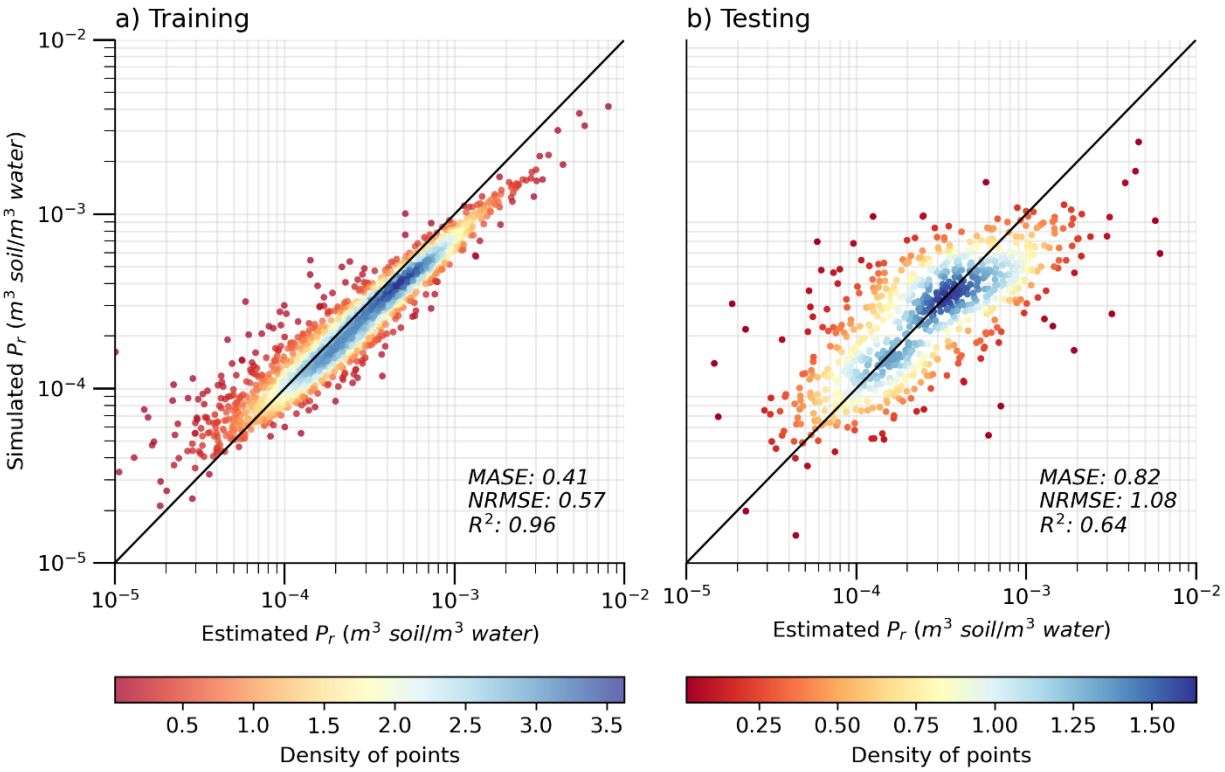

Figure S13. Performance of the XGBoost model (using HWSD) with 15 predictors during a) the training phase (n=1816) and b) the testing phase (n=779). The solid black line indicates a 1:1 ratio. Note that the axes are in a log-log scale.

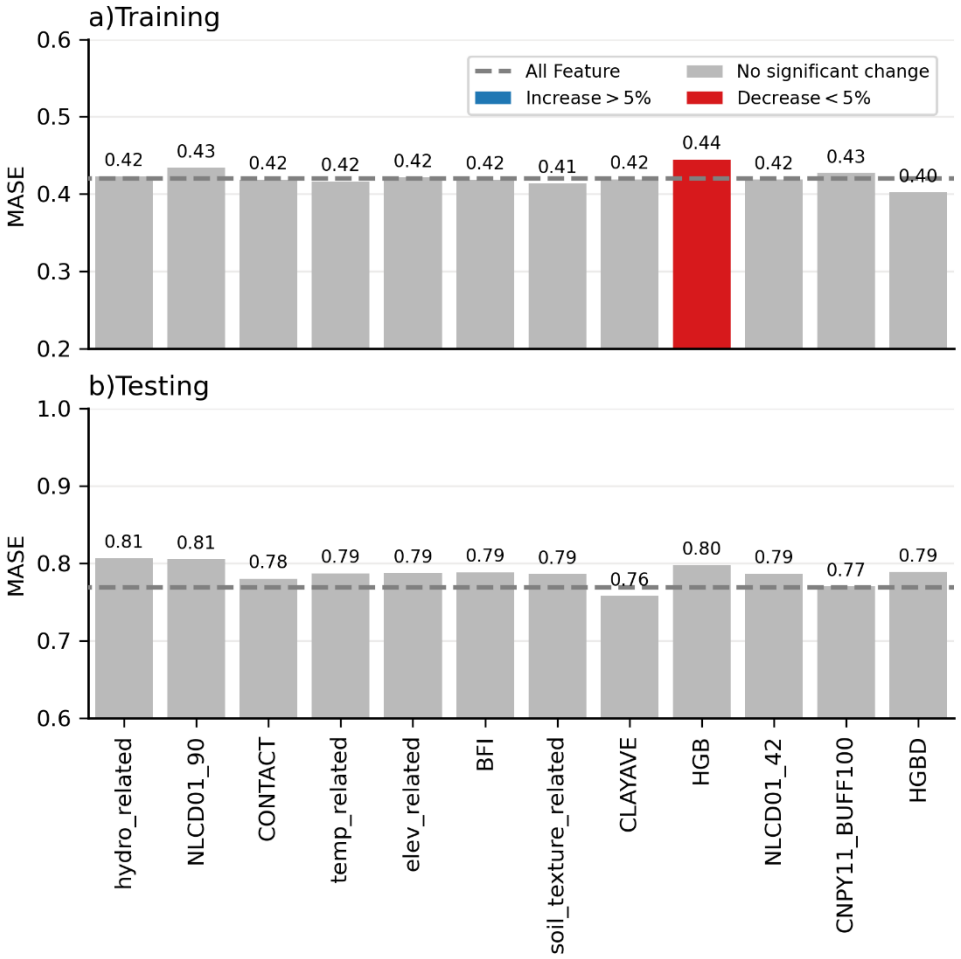

Figure S14. Sensitivity of XGBoost model (using SoilGrids 2.0) to predictors in the training and testing phases. The MASE value is represented by the blue, red, and grey bars, indicating whether the model performance increases, decreases, or remains relatively unchanged after dropping the corresponding predictor. The dashed grey line indicates the model performance with all variables included. Abbreviations: hydro_related (merged predictor representing recharge, runoff, and precipitation); NLCD01_90 (areal percentage of woody wetlands); CONTACT (subsurface contact time); temp_related (merged predictor encompassing potential evapotranspiration, first/last freeze timing, snow fraction, actual evapotranspiration, and mean/min/max temperature); elev_related (merged predictor for mean/min/max elevation); BFI (base flow index); soil_texture_related (merged predictor for silt and sand content); CLAYAVE (clay content percentage); HGB (areal percentage of Hydrologic Group B soil); NLCD01_42 (areal percentage of evergreen forest); CNPY11_BUFF100 (areal percentage of canopy in the riparian buffer); HGBD (areal percentage of Hydrologic Group BD soil). For detailed descriptions, refer to Supplementary Tables S2 and S3.

Table S1. In-stream DOC degradation rate (k) from previous modeling and experimental studies

Study Type	First-Order Decay rate (k d^{-1})	Study Domain	Reference
Modeling	0.01	Eastern North America	Tian et al., 2015
	0.01	Global	Li et al., 2019
	0.0163/0.0223[a]	Upland and forested catchments in Canada	Futter et al., 2007
Experimental	0.011[b]	Upland and forested catchment (Southern Appalachian Mountains, USA)	Qualls and Haines, 1992
	0.009[b]	Upland and forested catchment (Catskill Mountains, USA)	Sobczak et al., 2003
	0.013[c]	Forested headwater catchment (Haean Basin, South Korea)	Jung et al., 2014
	0.09[c]	Agro-urban headwater catchments (Taihu Lake Watershed, China)	Wu et al., 2019

a. calibrated for the two catchments separately.

b. adopted from Table 2 in Mineau et al., 2016

c. calculated by fitting a first-order decay model using the published data.

Table S2. List of 46 independent predictive attributes

Group	Acronym	Description
Chemical	NEMATICIDE	Nematicide use on agricultural land (kg/km^2)
Climate	CWD	Average number of consecutive days with measurable precipitation
	MINP6190	Watershed minimum average annual precipitation (mm)
	MINWD6190	Watershed average of minimum monthly number of days of measurable precipitation
	RH	Percent of the watershed average relative humidity
Geology	BEDPERM_1	Percent of NHDPlus version 2 flowline catchment whose bedrock permeability class is not a principal aquifer
	BEDPERM_6	Percent of NHDPlus version 2 flowline catchment whose bedrock permeability class is unconsolidated sand and gravel
	OLSON_PERM	Rock hydraulic conductivity (10^{-6} m/s)
	ROCKTYPE_200	Estimated percent of catchment that is underlain by the principal aquifer rock type, semi-consolidated sand aquifers
	ROCKTYPE_999	Estimated percent of catchment that is underlain by the principal aquifer rock type, other rocks
Hydrologic	BFI	Base Flow Index, a ratio of base flow to total streamflow, expressed as a percentage and ranging from 0 to 100
	CONTACT	The Subsurface flow contact time index estimates the duration infiltrated water resides in the saturated subsurface zone of the basin before discharging into the stream, measured in days
	IEOF	Percent of Horton overland flow
	SATOF	Percent of Dunne overland flow
Hydrologic Modifications	DITCHES92	Percent of watershed subjected to ditches for the year 1992
	MIRAD_2002	Percent of watershed in irrigated agriculture, from USGS 2002 250-m MODIS data

	MIRAD_2007	Percent of watershed in irrigated agriculture, from USGS 2007 250-m MODIS data
	MIRAD_2012	Percent of watershed in irrigated agriculture, from USGS 2012 250-m MODIS data
	TILES92	Percent of watershed subjected to tile drains for the year 1992
Land Cover	CNPY11_BUFF100	Percent of tree canopy in 100meter riparian buffer
	LAKEPOND	Percent of lakes or ponds
	NLCD01_11	Areal percent of 2001 land-use and land-cover type Open Water: All areas of open water, generally with less than 25 percent cover of vegetation or soil.
	NLCD01_31	Areal percent of 2001 land-use and land-cover type Barren Land: Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits, and other accumulations of earthen material. Generally, vegetation accounts for less than 15 percent of total cover.
	NLCD01_41	Areal percent of 2001 land-use and land-cover type Deciduous Forest: Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.
	NLCD01_42	Areal percent of 2001 land-use and land-cover type Evergreen Forest: Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent of total vegetation cover. More than 75 percent of the trees maintain their leaves all year. Canopy is never without green foliage.
	NLCD01_43	Areal percent of 2001 land-use and land-cover type Mixed Forest: Areas dominated by trees generally greater than 5 meters tall, and greater than 20 percent of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.
	NLCD01_52	Areal percent of 2001 land-use and land-cover type Shrub/Scrub: Areas dominated by shrubs less than 5 meters tall. Shrub canopy is typically greater than 20 percent of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
	NLCD01_71	Areal percent of 2001 land-use and land-cover type Grassland/Herbaceous: Areas dominated by graminoid or herbaceous vegetation, generally greater than 80 percent of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.

	NLCD01_81	Areal percent of 2001 land-use and land-cover type Pasture/Hay: Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.
	NLCD01_90	Areal percent of 2001 land-use and land-cover type Woody Wetlands: Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
	NLCD01_95	Areal percent of 2001 land-use and land-cover type Emergent Herbaceous Wetlands: Areas where perennial herbaceous vegetation accounts for greater than 80 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
	SWAMPMARSH	Percent of swamps or marshes
Soils	CLAYAVE	Percent of average clay content
	HGA	Areal percent of Hydrologic Group A soil. Hydrologic group A soils have high infiltration rates. Soils are deep and well drained and, typically, have high sand and gravel content
	HGAD	Areal percent of Hydrologic Group AD soil. Hydrologic group AD soils have group A characteristics (high infiltration rates) when artificially drained and have group D characteristics (very slow infiltration rates) when not drained
	HGB	Areal percent of Hydrologic Group B soil. Hydrologic group B soils have moderate infiltration rates. Soils are moderately deep, moderately well drained, and moderately coarse in texture
	HGBD	Areal percent of Percentage of Hydrologic Group BD soil. Hydrologic group BD soils have group B characteristics (moderate infiltration rates) when artificially drained and have group D characteristics (very slow infiltration rates) when not drained
	HGC	Areal percent of Hydrologic Group C soil. Hydrologic group C soils have slow soil infiltration rates. The soil profiles include layers impeding downward movement of water and, typically, have moderately fine or fine texture
	HGCD	Areal percent of Hydrologic Group CD soil. Hydrologic group CD soils have group C characteristics (slow infiltration rates) when artificially drained and have group D characteristics (very slow infiltration rates) when not drained
	HGD	Areal percent of Percentage of Hydrologic Group D soil. Hydrologic group D soils have very slow infiltration rates. Soils are clayey, have a high water table, or have a shallow impervious layer

	SALINAVE	Salinity measured as average millimhos per centimeter
Topographic	ARTIFICIAL	Flowlines coded as Artificial paths
	BASIN_AREA	NHDPlusV2 flowline catchment area (km^2)
	RDX	Number of road and stream intersections
	STREAMRIVER	Flowlines coded as Streams or Rivers
Water Use	FRESHWATER_WD	Freshwater withdrawals from 1995-2000 county-level estimates

105 **Table S3. List of 49 correlated predictive attributes**

Correlated Group	Acronym	Description
hydro_related	RECHG	Mean annual natural groundwater recharge (mm/yr).
	WB5100_ANN	Annual averaged runoff from McCabe and Wolock's Runoff Model 1951-2000 (mm).
	MAXP6190	Watershed maximum average annual precipitation (mm).
	PPT7100_ANN	Mean annual precipitation for the watershed, from 800m PRISM data (mm). 30 years period of record 1971-2000.
	RUN7100	Estimated 30-year (1971-2000) average annual runoff (mm/yr).
temp_related	PET	Mean annual potential evapotranspiration (PET), estimated using the Hamon equation.
	FSTFZ6190	Watershed average of mean day of the year of first freeze, derived from 30 years of record (1961-1990), 2km PRISM.
	LSTFZ6190	Watershed average of mean day of the year of last freeze, derived from 30 years of record (1961-1990), 2km PRISM.
	PRSNOW	Snow percent of total precipitation estimate, mean for period 1901-2000.
	ET	Mean-annual actual evapotranspiration (ET), estimated using regression equation of Sanford and Selnick (2013).
	TMAX7100	Watershed average of maximum monthly air temperature from 800m PRISM, derived from 30 years (1971-2000) of record (ºC).
	TAV7100_ANN	Watershed average of monthly air temperature from 800m PRISM, derived from 30 years (1971-2000) of record (ºC).
	TMIN7100	Watershed average of minimum monthly air temperature from 800m PRISM, derived from 30 years (1971-2000) of record (ºC).
agri_chem_related	FUNGICIDE	Fungicide use on agricultural land (kg/km^2).
	HERBICIDE	Herbicide use on agricultural land (kg/km^2).

	INSECTICIDE	Insecticide use on agricultural land (kg/km^2).
	N97	Estimated nitrogen from fertilizer and manure.
	P97	Estimated phosphorous from fertilizer and manure.
	NLCD01_82	Areal percent of 2001 land-use and land-cover type Cultivated Crops: Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20 percent of total vegetation. This class also includes all land being actively tilled.
	PEST219	Estimate of agricultural pesticide application (kg/km^2).
	KGBI	Toxicity Weighted Use for benthic invertebrates on agricultural land, 2013 (kg/km^2).
	KGCLADO	Toxicity Weighted Use for cladocerans on agricultural land, 2013 (kg/km^2).
	KGFISH	Toxicity Weighted Use for fish on agricultural land, 2013 (kg/km^2).
urban_related	POPDENS90	Population density from 1990 Census block level data, persons per km^2 per NHDPlus version 2 catchment.
	IMPV01_BUFF100	NLCD 2001 percent of imperviousness in the 100meter riparian buffer zones.
	IMPV06	NLCD 2006 percent of imperviousness per NHDPlus version 2 catchment.
	IMPV06_BUFF100	NLCD 2006 percent of imperviousness in the 100meter riparian buffer zones.
	POPDENS00	Population density from 2000 Census block level data, persons per km^2 per NHDPlus version 2 catchment.
	POPDENS10	Population density from 2010 Census block level data, persons per km^2 per NHDPlus version 2 catchment.

	NLCD01_21	Areal percent of 2001 land-use and land-cover type Developed, Open Space: Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
	NLCD01_22	Areal percent of 2001 land-use and land-cover type Developed, Low Intensity: Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family housing units.
	NLCD01_23	Areal percent of 2001 land-use and land-cover type Developed, Medium Intensity: Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 percent of the total cover. These areas most commonly include single-family housing units.
	NLCD01_24	Areal percent of 2001 land-use and land-cover type Developed, High Intensity: Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses, and commercial/industrial. Impervious surfaces account for 80-100 percent of the total cover.
	TOTAL_ROAD_DENS	Density of all road types per NHDPlusV2 catchment. Density is defined as the length of road divided by the catchment area.
	HDENS10	Historic housing densities for 2010.
soil_texture_related	SILTAVE	Percent of average value of silt content.
	SANDAVE	Percent of average value of sand content.
soil_restrictive_related	SRL25AG	Estimated percent of the soil restrictive layer in the upper 25cm of agricultural land.
	SRL35AG	Estimated percent of the soil restrictive layer in the upper 35cm of agricultural land.
	SRL45AG	Estimated percent of the soil restrictive layer in the upper 45cm of agricultural land.
	SRL55AG	Estimated percent of the soil restrictive layer in the upper 55cm of agricultural land.
wetd_related	MAXWD6190	Watershed average of maximum monthly number of days of measurable precipitation, derived from 30 years of record (1961-1990), 2.3km PRISM.

	WDANN	NHDPlusV2 catchment value for the 30year annual average (1961-1990) number of days of measurable precipitation.
topo_related	EWT	Average depth to water table relative to the land surface (m).
	TWI	Topographic wetness index, ln(a/S); where ln is the natural log, a is the upslope area per unit contour length and S is the slope at that point.
	BASIN_SLOPE	NHDPlusV2 flowline catchment's average slope in percent.
elev_related	ELEV_MEAN	NHDPlusV2 flowline catchment's mean elevation (m).
	ELEV_MIN	NHDPlusV2 flowline catchment's minimum elevation (m).
	ELEV_MAX	NHDPlusV2 flowline catchment's maximum elevation (m).

Table S4. The optimal hyperparameters values of the XGBoost model (using SoilGrids 2.0).

Hyperparameter	Optimal Value	Tuning Range	Default value	Description
lambda	8.497×10^{-1}	[0, ∞]	1	Control L1 and L2 regularization; the larger the value, the more conservative the model will be
alpha	2.198×10^{-2}	[0, ∞]	0	
gamma	9.045×10^{-2}	[0, ∞]	0	Govern the model learning process by changing the step size shrinkage and minimum loss reduction; the larger the value, the more conservative the model will be
eta	1.146×10^{-1}	(0, 1]	0.3	
colsample_bytree	5.004×10^{-1}	(0, 1]	1	Control the subsample ratio of columns and training instances; a proper set of those values will prevent the model from over-fitting
subsample	9.730×10^{-1}	(0, 1]	1	
min_child_weight	3.123×10^{-1}	[0, ∞]	1	Determine the growth of the tree
max_depth	8	[0, ∞]	6	

Table S5. Representativeness of XGBoost model (using SoilGrids 2.0) input predictors over CONUS.

Attributes	Relative difference in percentiles between *P_r-available* and *whole_conus* data					Average
	5th	25th	50th	75th	95th	
NLCD01_95	0.667	0.667	0.853	1.144	1.528	0.972
CNPY11_BUFF100	1.686	1.098	0.429	0.080	0.078	0.674
NLCD01_90	0.769	0.307	0.448	0.612	0.806	0.589
NLCD01_42	0.689	0.548	0.648	0.501	0.224	0.522
elev_related	0.736	0.783	0.312	0.618	0.008	0.491
hydro_related	0.586	0.899	0.317	0.109	0.106	0.403
HGBD	0.920	0.276	0.159	0.098	0.252	0.341
CONTACT	0.165	0.140	0.248	0.294	0.410	0.251
CLAYAVE	0.233	0.412	0.249	0.178	0.140	0.243
HGB	0.232	0.308	0.180	0.160	0.093	0.194
BFI	0.477	0.306	0.152	0.002	0.028	0.193
soil_texture_related	0.096	0.071	0.068	0.070	0.015	0.064
temp_related	0.036	0.034	0.008	0.029	0.005	0.022

Abbreviations: NLCD01_95 (areal percentage of herbaceous wetlands); CNPY11_BUFF100 (areal percentage of canopy in the riparian buffer); NLCD01_90 (areal percentage of woody wetlands); NLCD01_42 (areal percentage of evergreen forest); elev_related (merged predictor for mean/min/max elevation); hydro_related (merged predictor representing recharge, runoff, and precipitation); HGBD (areal percentage of Hydrologic Group BD soil); CONTACT (subsurface contact time); CLAYAVE (clay content percentage); HGB (areal percentage of Hydrologic Group B soil); BFI (base flow index); soil_texture_related (merged predictor for silt and sand content); temp_related (merged predictor encompassing potential evapotranspiration, first/last freeze timing, snow fraction, actual evapotranspiration, and mean/min/max temperature). For detailed descriptions, refer to Supplementary Tables S2 and S3.

Table S6. Sobol sensitivity analysis results for the 12 selected predictors (using SoilGrids 2.0).

Predictors	Total Indices (ST)	First Order Indices (S1)	Difference ((ST-S1)/ST)
temp_related	0.573	0.316	0.448
hydro_related	0.192	0.124	0.351
CONTACT	0.118	0.044	0.624
NLCD01_42	0.115	0.009	0.918
elev_related	0.102	0.036	0.643
CNPY11_BUFF100	0.098	0.016	0.833
HGB	0.075	0.007	0.910
NLCD01_90	0.057	0.028	0.507
CLAYAVE	0.049	0.001	0.978
BFI	0.047	0.003	0.938
soil_texture_related	0.032	0.000	1.000
HGBD	0.005	0.002	0.499

Abbreviations: temp_related (merged predictor encompassing potential evapotranspiration, first/last freeze timing, snow fraction, actual evapotranspiration, and mean/min/max temperature); hydro_related (merged predictor representing recharge, runoff, and precipitation); CONTACT (subsurface contact time); NLCD01_42 (areal percentage of evergreen forest); elev_related (merged predictor for mean/min/max elevation); CNPY11_BUFF100 (areal percentage of canopy in the riparian buffer); HGB (areal percentage of Hydrologic Group B soil); NLCD01_90 (areal percentage of woody wetlands); CLAYAVE (clay content percentage); BFI (base flow index); soil_texture_related (merged predictor for silt and sand content); HGBD (areal percentage of Hydrologic Group BD soil). For detailed descriptions, refer to Supplementary Tables S2 and S3.

Reference

Futter, M. N., Butterfield, D., Cosby, B. J., Dillon, P. J., Wade, A. J., and Whitehead, P. G.: Modeling the mechanisms that control in-stream dissolved organic carbon dynamics in upland and forested catchments, Water Resour Res, 43, https://doi.org/10.1029/2006WR004960, 2007.

Jung, B. J., Lee, J. K., Kim, H., and Park, J. H.: Export, biodegradation, and disinfection byproduct formation of dissolved and particulate organic carbon in a forested headwater stream during extreme rainfall events, Biogeosciences, 11, 6119–6129, https://doi.org/10.5194/bg-11-6119-2014, 2014.

Li, M., Peng, C., Zhou, X., Yang, Y., Guo, Y., Shi, G., and Zhu, Q.: Modeling Global Riverine DOC Flux Dynamics From 1951 to 2015, J Adv Model Earth Syst, 11, 514–530, https://doi.org/10.1029/2018MS001363, 2019.

Mineau, M. M., Wollheim, W. M., Buffam, I., Findlay, S. E. G., Hall, R. O., Hotchkiss, E. R., Koenig, L. E., McDowell, W. H., and Parr, T. B.: Dissolved organic carbon uptake in streams: A review and assessment of reach-scale measurements, https://doi.org/10.1002/2015JG003204, 1 August 2016.

Qualls, R. G. and Haines, B. L.: Biodegradability of Dissolved Organic Matter in Forest Throughfall, Soil Solution, and Stream Water, Soil Science Society of America Journal, 56, 578–586, https://doi.org/10.2136/sssaj1992.03615995005600020038x, 1992.

Sobczak, W. V, Findlay, S., and Dye, S.: Relationships between DOC bioavailability and nitrate removal in an upland stream: An experimental approach, 2003.

Tian, H., Ren, W., Yang, J., Tao, B., Cai, W. J., Lohrenz, S. E., Hopkinson, C. S., Liu, M., Yang, Q., Lu, C., Zhang, B., Banger, K., Pan, S., He, R., and Xue, Z.: Climate extremes dominating seasonal and interannual variations in carbon export from the Mississippi River Basin, Global Biogeochem Cycles, 29, 1333–1347, https://doi.org/10.1002/2014GB005068, 2015.

Wu, Z., Wu, W., Lin, C., Zhou, S., and Xiong, J.: Deciphering the origins, composition and microbial fate of dissolved organic matter in agro-urban headwater streams, Science of the Total Environment, 659, 1484–1495, https://doi.org/10.1016/j.scitotenv.2018.12.237, 2019.