nature geoscience

Supplementary information

https://doi.org/10.1038/s41561-026-02096-w

Curbing six key sources could halve pharmaceutical mixture exceedance in global rivers

In the format provided by the authors and unedited

Supplementary Information

Table of Contents

Supplementary Texts

Text S1. Organism cultivation
Text S2. Quality assurance and quality control
Text S3. The selection and processing of predictive variables

Supplementary Figures

Fig. S1. Flowchart for addressing the toxicity assessment of pharmaceutical (API) mixture in global rivers
Fig. S2. The number of APIs with different modes of action detected at the 1,017 riverine sites
Fig. S3. The ROC curves for the three XGBoost models based on 24 geospatial predictors
Fig. S4. The probability cutoffs in the three XGBoost models based on 24 geospatial predictors
Fig. S5. Uncertainties in the three prediction maps of API exceedance
Fig. S6. Uncertainty comparison for the three XGBoost models
Fig. S7. The relative contributions of all geospatial predictors in the three XGBoost models
Fig. S8. SHapley Additive exPlanations (SHAP) plots for the three XGBoost models
Fig. S9. The ROC curves for the three XGBoost models based on the 53 APIs
Fig. S10. The relative contributions of all APIs in the three XGBoost models
Fig. S11. The comparison of occurrence and toxicity between 9 priority and 44 non-priority APIs
Fig. S12. Contributions of API interaction to total mixture toxicity across 1,017 mixtures
Fig. S13. External validation and confusion matrix for the XGBoost model
Fig. S14. The relative importance and coefficient of variation of SHAP values for 24 predictors
Fig. S15. A priori structural equation modeling
Fig. S16. The mass loss rate of 53 APIs under the 72 h exposure conditions
Fig. S17. Comparison between nominal concentration and measured acute concentration

Supplementary Tables

Table S1. The detailed information of the selected 53 APIs
Table S2. Performances of the three XGBoost models based on the 24 geographical parameters
Table S3. Performances of the three XGBoost models based on the 53 APIs
Table S4. Full information of 40 predictors used in the machine learning models
Table S5. The performances of other four models based on the 24 geospatial predictors
Table S6. The optimized hyperparameters in three XGBoost models based on 24 geospatial predictors
Table S7. The performances of other four models based on the 53 APIs
Table S8. The optimized hyperparameters in the three XGBoost models based on the 53 APIs
Table S9. The composition of the BG11 medium
Table S10. The composition of the artificial freshwater for the cultivation for *D. magna*

Supplementary References

Text S1. Organism cultivation

The cultivation protocols for *E. coli*, *R. subcapitata* and *D. magna* were based on the Organization for Economic Cooperation and Development (OECD) guidelines and published literature[1-3]. *E. coli* strain K12 was obtained from the China General Microbiological Culture Collection Center. *R. subcapitata* (FACHB-271) and *Chlorella vulgaris* (FACHB-8) were purchased from the Institute of Hydrobiology, Chinese Academy of Sciences (Wuhan, China). *D. magna* was obtained from our cooperative laboratory in Tsinghua University (Beijing, China).

E. coli was grown on Luria-Bertani (LB) medium (containing 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl with the pH adjusted to 7.0) for 24 h at 37 °C under 180 rpm sharking speed. The culture was harvested in the mid-exponential growth phase and diluted to approximately 5×10^6 colony-forming units/mL (approximately 0.06 of optical absorbance at 600 nm, OD_{600}) for the exposure experiments. *R. subcapitata* was grown in BG11 media (Table S9), and harvested during the exponential growth phase and diluted to initial concentrations of 5×10^4 cells/mL for the exposure experiments. *D. magna* was cultured in artificial freshwater (Table S10), and fed with *Chlorella vulgaris* (FACHB-8). 20 *D. magna* neonates (aged less than 24 h) were selected for the exposure experiments. *R. subcapitata* and *D. magna* were cultured in a growth chamber at 22 °C with constant light (4000 lx) provided under a 16/8 h light/dark cycle.

Text S2. Quality assurance and quality control

(1) Setup of exposure experiment

To minimize the mass loss of APIs due to the adsorption, the exposure experiments were performed in 96-well glass-coated polypropylene microplates[4] or glass culture dishes.

During the exposure periods, all glass containers were randomly placed in the incubator and chamber, and frequently rearranged during the exposure periods to ensure homogeneous exposure conditions and to reduce uncertainty. Based on our previous experience, we chose dimethyl sulfoxide (DMSO) as the solvent, as 0.1% DMSO has been shown to be an effective co-solvent and has no significant ($p > 0.05$) toxicity to the test species[4,5].

(2) Evaluation of chemical stability

The potential for chemical mass loss during exposure experiments is a recognized phenomenon, arising from pathways such as hydrolysis, oxidation, photo-degradation, volatilization, and adsorption onto container surfaces[5,6]. Consequently, we conducted an experiment to evaluate the chemical stability of 53 APIs under experimental conditions, following our recent study[5]. Specifically, 0.1 mg/L solutions of each API were prepared in glass vessels and subsequently incubated in a growth chamber under conditions identical to those employed in the exposure experiments. API stability was then determined by quantifying API concentrations at 0 and 72 hours, followed by calculating the resultant mass loss. API concentrations were quantified using an online solid phase extraction (SPE)-liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Agilent 6470, CA, USA); the detailed description of the analytical method is available in our previous study[7]. For the quantification of 53 APIs, calibration curves were established using five standard concentrations per analyte, with coefficients of determination (R^2) exceeding 0.99 for all curves. The limit of detection and limit of quantification for the 53 APIs were 0.01-0.1 and 2.4-10.0 ng/L, respectively.

In accordance with the OECD Guideline[6], an API is classified as stable if its mass loss remains below 20%. Of the 53 APIs tested, 49 exhibited stabilities, with mass loss remaining

below this threshold (Fig. S16). Conversely, tetracycline, clarithromycin, erythromycin, and tylosin exhibited mass losses of 23.7%, 31.7%, 32.3%, and 33.7%, respectively. The result was generally consistent with a previous study that reported a 33.0% mass loss for erythromycin in a chemical stability experiment[6]. Notably, the nine priory APIs identified in the main text (Section of Prioritization of pharmaceuticals), which cumulatively accounted for 54%–90% of the mixture toxicity across the three species, did not include these four unstable APIs. This suggests that the observed mass loss has a relatively limited impact on the overall risk assessment of API mixtures. While acknowledging the inherent challenges associated with mass loss during exposure periods, future research could explore the utilization of specialized equipment designed to sustain consistent API concentrations in solutions.

(3) Verification of chemical concentration

In this study, we prepared 1,017 experimental solutions based on the actual concentrations of 53 APIs detected at the 1,017 riverine sites. To minimize potential preparation errors, we used certified single‑analyte liquid standards for each API (10, 100, or 1000 mg/L; vendor certified by Alta Scientific Co. Ltd. (Tianjing, China) or J&K Chemical Ltd. (Beijing, China)) instead of solid standards. Mixtures were subsequently prepared via serial volumetric dilution employing calibrated Eppendorf pipettes. Each API was first dissolved in DMSO (final ≤0.1% v/v as co-solvent) and then diluted into ultrapure water. All mixtures were freshly prepared immediately prior to testing.

Given the significant analytical burden in simultaneous quantitative analysis of all APIs in 1,017 solutions, we opted to perform a targeted validation of nominal–measured agreement. Specifically, we first combined the single-analyte liquid standards to prepare a

53-API stock mixture, then diluted this stock with ultrapure water to yield six working mixtures with per-API nominal concentrations of 17, 24, 41, 63, 79 and 137 ng/L, covering the central range of target concentrations represented in the 1,017 mixtures. API concentrations were immediately quantified using an online SPE-LC-MS/MS as described above Section. All analytes were above their LOQs, and measured concentrations closely matched nominal values: deviations were within ±6% for all analyte–mixture pairs, with 83% within ±3% (Fig. S17). These results indicate good concordance between nominal design and actual exposure concentrations, supporting the validity of the toxicity assessments based on the prepared mixtures.

Text S3. The selection and processing of predictive variables

Initially, we compiled a set of 40 geospatial predictors for the development of machine learning models to predict the API exceedance in global rivers. Selection of these predictors was based on two key criteria: (1) evidence from prior studies indicating a plausible direct or indirect influence on API accumulation; and (2) availability of dataset from peer-reviewed publications or open-access databases. These predictors encompass a broad spectrum of categories, including socioeconomic, climatic, API source proximity, soil, hydrochemical, and topographical parameters. Comprehensive details on variable descriptions, data sources, and spatial resolutions are presented in Table S4. The subsequent subsections outline the selection criteria and data processing workflow:

(1) Socioeconomic variables

We selected six socioeconomic indicators, including population density, human development index, urban extent, human footprint, gross domestic product (GDP), and GDP per capita, to characterize the drivers of API discharge strength. Population growth, rapid

urbanization, and economic expansion are typically associated with elevated API production, consumption, and emissions[8].

(2) Climate variables

We used nine common climate variables for climatic conditions, including temperature, precipitation, solar radiation, wind speed, water vapor pressure, potential evapotranspiration (PET), actual evapotranspiration (AET), priestley-Taylor alpha coefficient (AET/PET), and aridity index. Climate conditions can influence runoff and river discharge, thereby affecting the environmental fate, transport, and dilution of chemicals in rivers[9].

(3) Topography variables

We selected topographic wetness index, terrain ruggedness index, compound topographic index, stream power index, elevation, and slope as proxies for topographic conditions. Topographic features can modulate river flow volume and direction, thereby affecting the chemical transport and accumulation[10].

(4) Soil variables

We selected soil organic carbon, pH, cation exchange capacity, and water content as proxies for soil properties. These soil properties may affect the dissolved concentrations and bioavailability of chemicals[11].

(5) Hydrochemistry variables

We selected river flow, river discharge, salinity, water temperature, total dissolved solid, adaptive capacity, and ecological vulnerability as proxies for hydrochemical conditions. These conditions can regulate the dilution, residence time, transport, adsorption, and degradation of chemicals in rivers[12]. These variables also serve as proxies for the matrix effects of natural waters, as factors such as salinity, temperature, total dissolved solid, and

discharge modulate the bioavailability of APIs through adsorption, complexation, and dilution dynamics. While globally gridded data for certain matrix parameters (e.g., dissolved organic matter and pH) remain unavailable, these selected variables capture the most readily measurable hydrochemical drivers of bioavailability at the global scale.

(6) API source proximity

These distance-to-source covariates are not intended to simulate specific transport pathways; rather, they serve as integrated spatial proximity proxies based on the principle that riverine grid cells closer to a given source have a higher probability of receiving API inputs through any combination of hydrological, subsurface, or atmospheric routes. Following previous studies[10,13-15], we calculated the Euclidean distance from each riverine grid cell centroid to the nearest point source for eight API source categories: wastewater treatment plants (WWTPs), landfills, aquaculture ponds, cropland, pasture, and livestock facilities (pig, duck, and chicken farms). WWTP effluents represent an important pathway for APIs to enter rivers due to incomplete removal during treatment[10,16]; consequently, proximity to WWTPs is positively correlated with the likelihood of API-contaminated discharge. Landfills may release APIs via leachate, with the probability of a given riverine grid cell being affected decreasing with distance. Aquaculture ponds are associated with direct application of veterinary APIs, and adjacent water bodies are more likely to receive residues through overflow or seepage[17]. Cropland[11] and pasture[18] receive reclaimed water, manure, and biosolids containing APIs; thus, the probability of runoff-mediated API transport into nearby streams is highest in grid cells immediately adjacent to these agricultural sources. Livestock farms extensively use antibiotics, and APIs excreted by animals can reach nearby riverine systems through manure storage, effluents, and land application[19,20].

To ensure consistency across all variables, we resampled all 40 variable datasets to match a uniform spatial resolution of 2 km. Gridded datasets with lower pixel resolutions were resampled using the Nearest method, while finer-resolution datasets were resampled by the mean aggregation method[11]. A 2 km gridded map encompassing 223,152 cells was then generated across the global river network using ArcGIS software, with Natural Earth database (https://www.naturalearthdata.com/) as the base map. Afterwards, the values of all 40 predictors corresponding to each grid cell were extracted based on their geographic coordinates and utilized as prediction dataset. For the grid cells with available toxicity data, we calculated the geometric mean of API mixture toxicity, then binarized it as API exceedance (≥5% inhibition/mortality) or non-exceedance (<5%). This binary variable served as the response variable, and the extracted 40 parameters were used as predictive variables for the machine learning classification model. Following the identification and removal of variables exhibiting high multicollinearity (variance inflation factor > 5) and substantial null values, recursive feature elimination[21] was further applied to select an optimal subset of 24 predictors for subsequent machine learning model development.

Together, these 24 predictors—including the distance-to-source covariates—constitute the feature space from which the machine learning model learns the empirical, nonlinear relationships with observed API mixture toxicity in these riverine grid cells. Crucially, the machine learning model does not simulate any physical transport or attenuation process; it captures the statistical associations between the environmental/source-proximity characteristics of a grid cell and its API toxicity. In this context, the distance-to-source covariates encode a simple but powerful principle: the closer a riverine grid cell is to a source, the higher the probability that it receives API inputs, whether through direct hydrological

connectivity, overland runoff, subsurface flow, or short-range atmospheric deposition. The machine learning model treats this as one dimension of a high-dimensional feature space, enabling it to learn how distance interacts with other variables (e.g., precipitation, slope, and discharge) to modulate API toxicity.

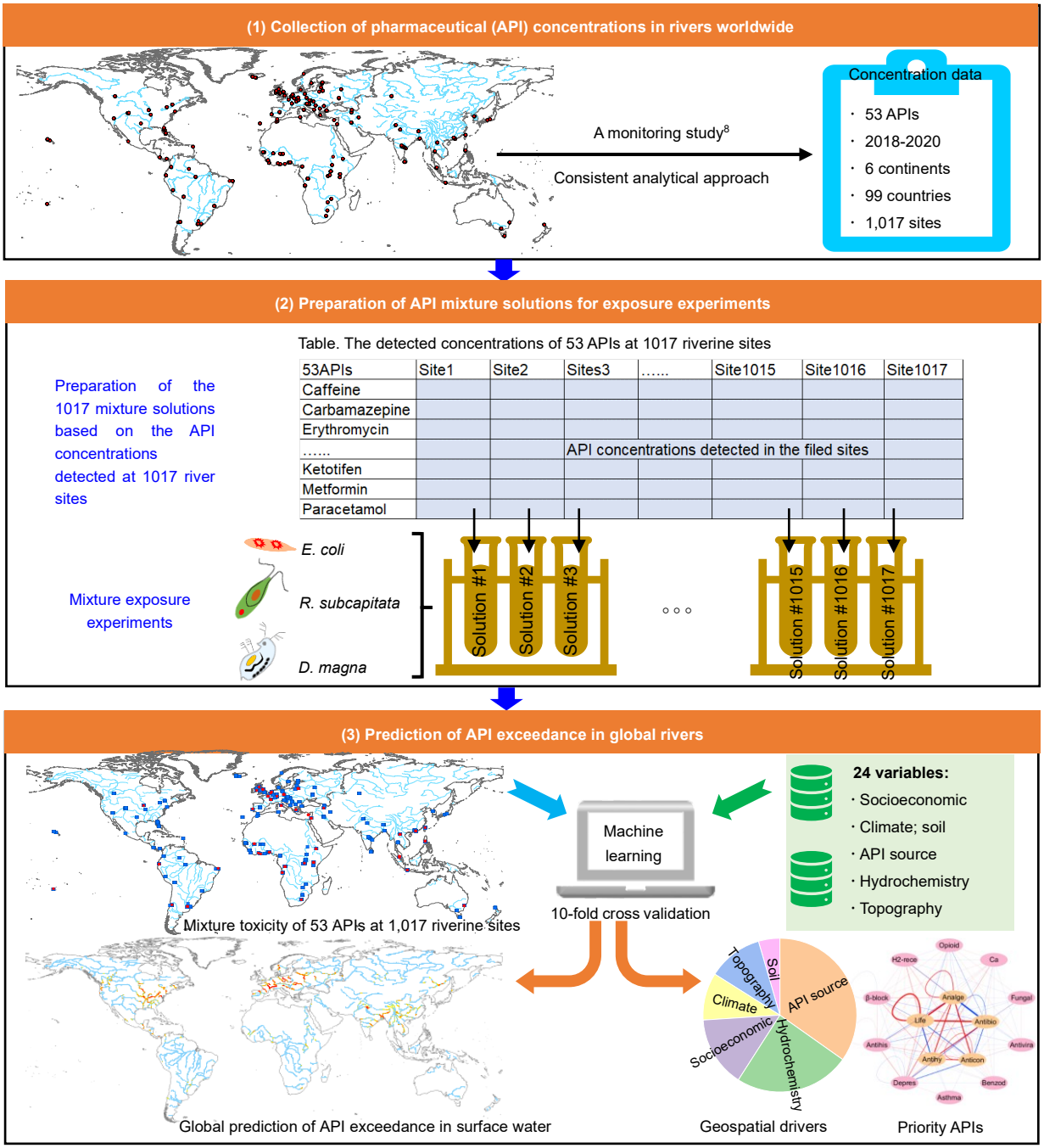


Fig. S1. Flowchart for addressing the toxicity assessment of pharmaceutical (API) mixture in global rivers. (1) Collection of measured concentrations of 53 APIs at 1,017 riverine sites in 99 countries on six continents from a monitoring study[8]. (2) Preparation of 1,017 mixture solutions containing combinations of the 53 APIs based on their actual detected concentrations at the 1,017 riverine sites for the exposure experiments to *E. coli*, *R. subcapitata*, and *D. magna*. (3) Development of machine learning models based on the 1,017 measurements and 24 predictors to generate the global maps of API exceedance at 2 km resolution, and to identify the geospatial drivers and priority APIs.

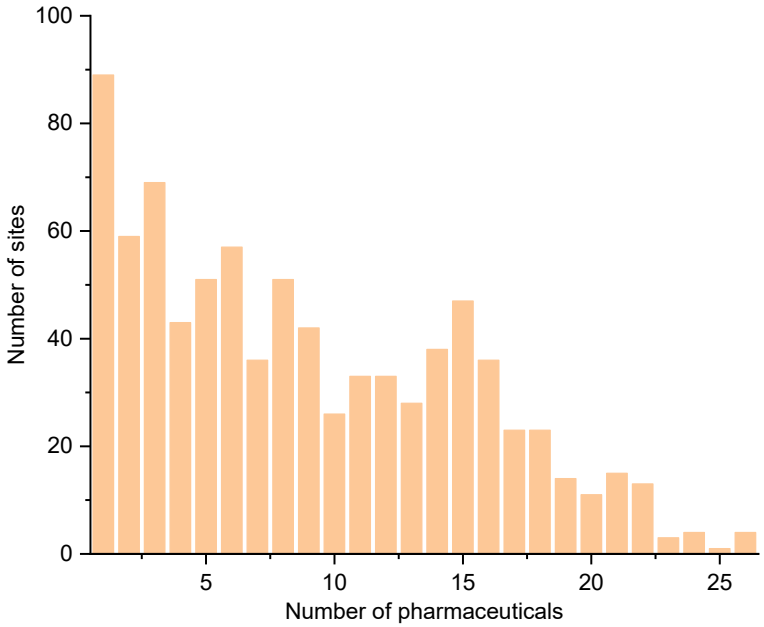


Fig. S2. The number of APIs with different modes of action detected at the 1,017 riverine sites. The modes of action for the 53 APIs were collected from the DrugBank Database (https://go.drugbank.com/).

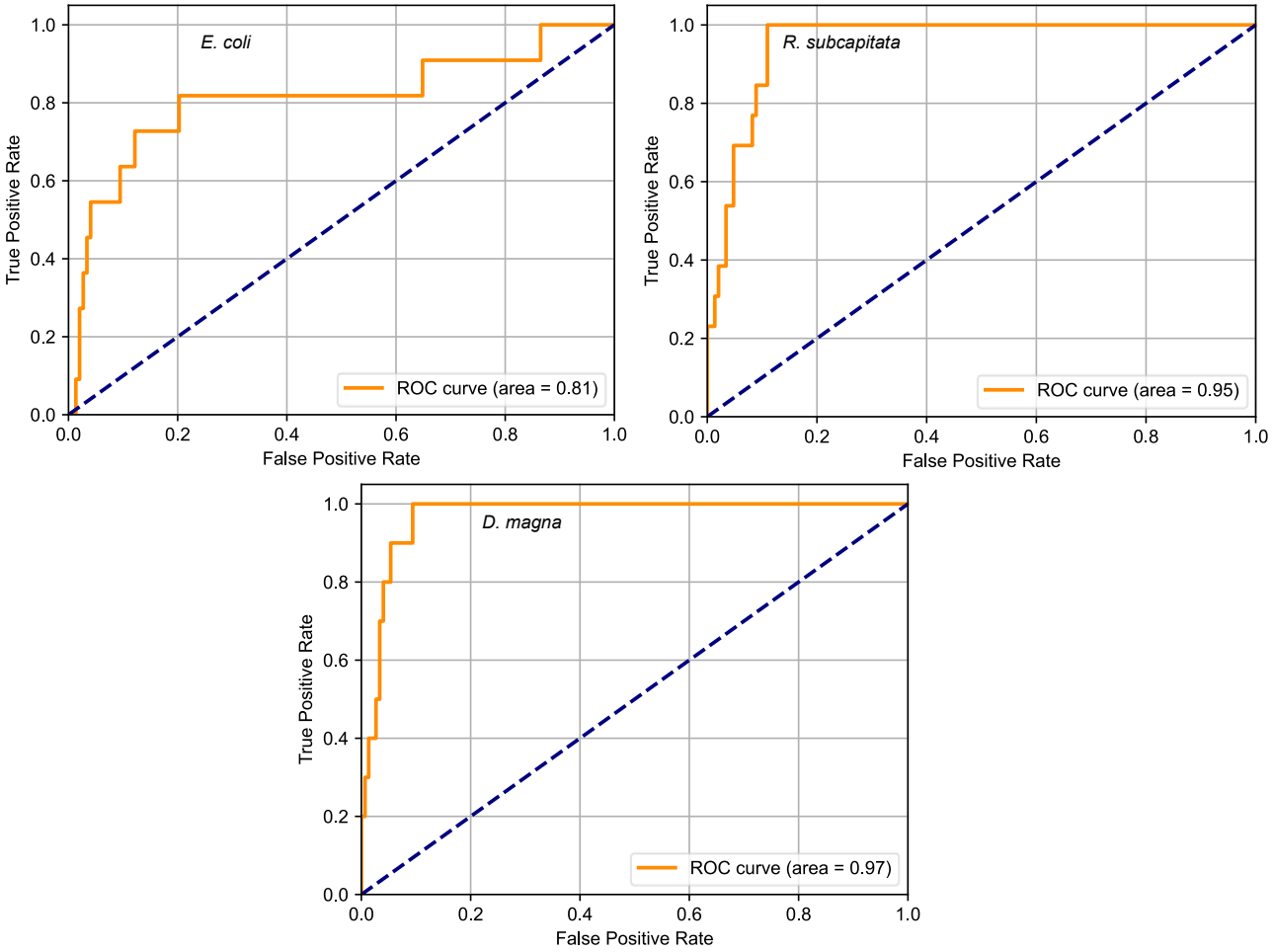


Fig. S3. The ROC curves for the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna*, based on 24 geospatial predictors.

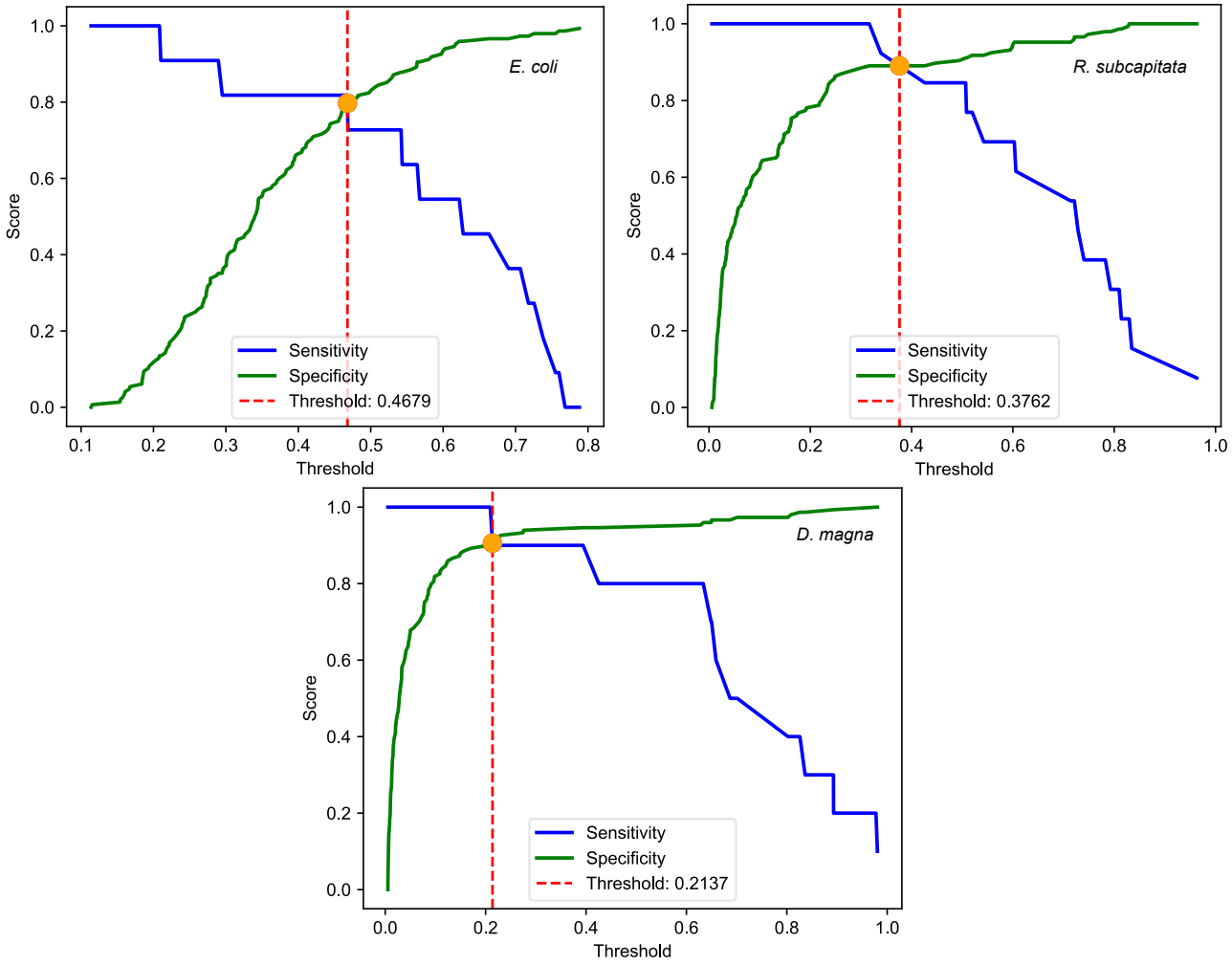


Fig. S4. The probability cutoffs in the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna*. A probability cutoff is the threshold where sensitivity and specificity are equal, indicating the same ability of the XGBoost model in classifying the toxicity of API mixtures greater or less than 5% inhibition/lethality.

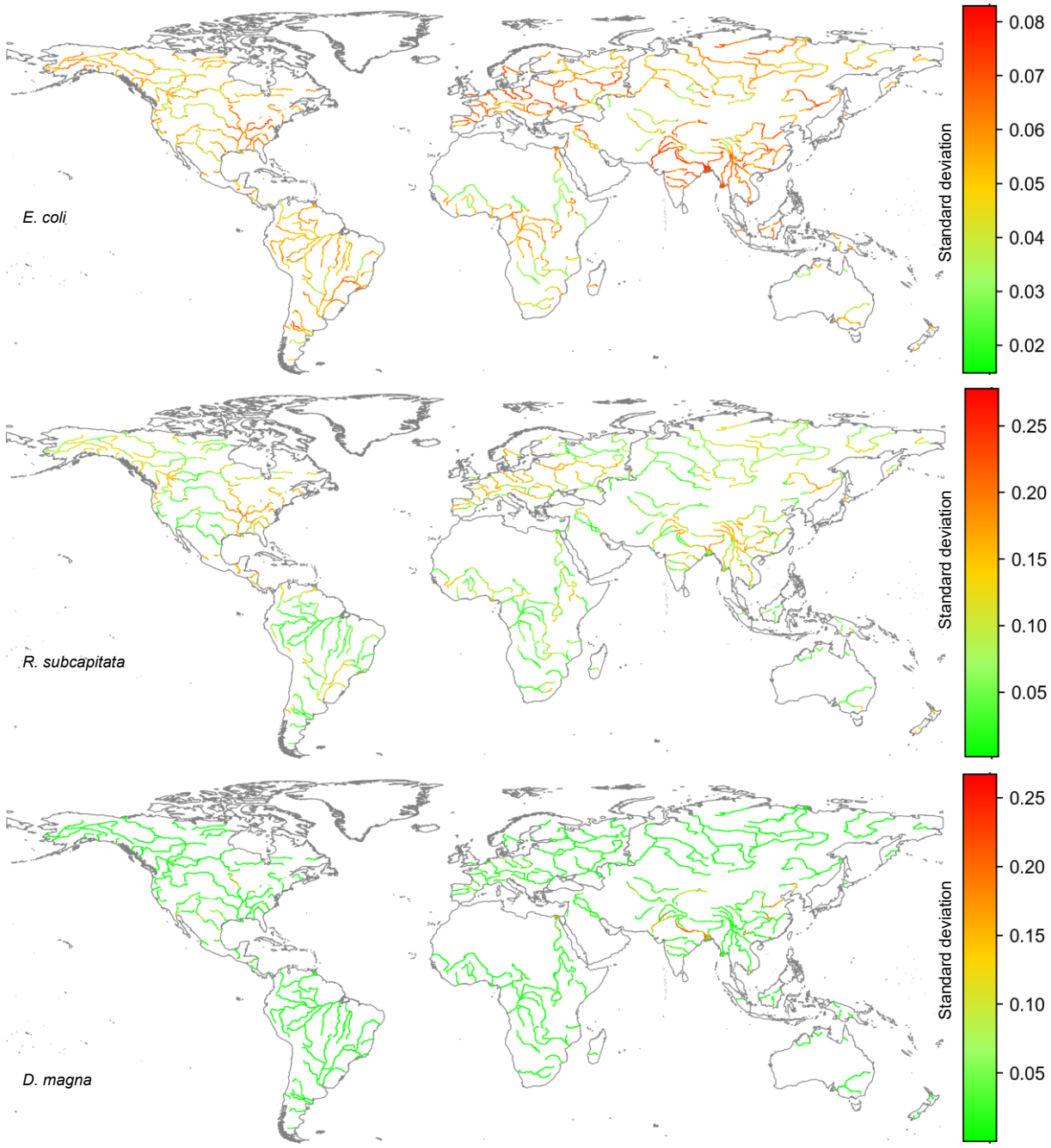


Fig. S5. Uncertainties in the three prediction maps of API exceedance for *E. coli*, *R. subcapitata*, and *D. magna*. Uncertainty was evaluated by calculating the standard deviation. A lower value of the standard deviation indicates lower uncertainty.

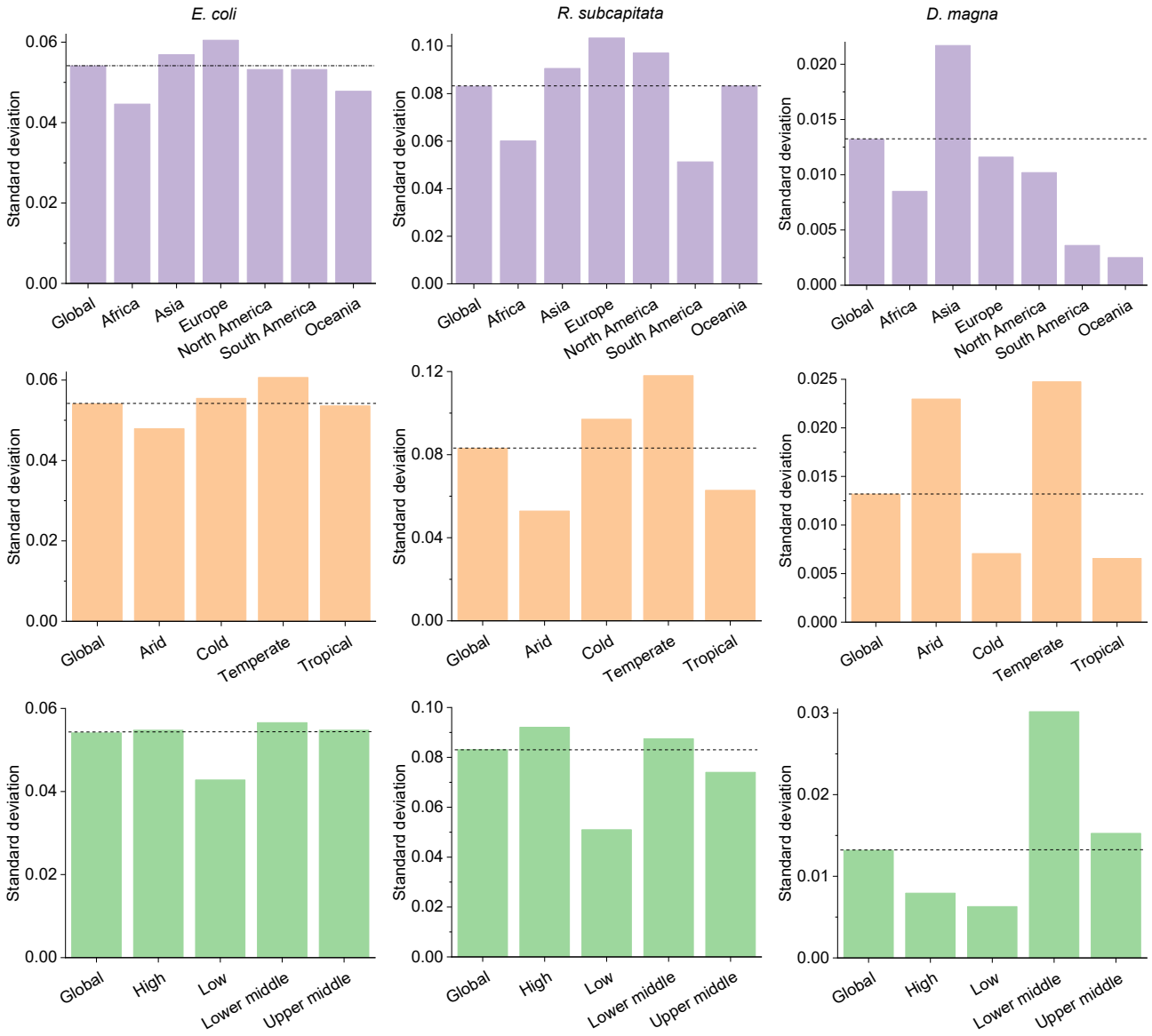


Fig. S6. Uncertainty comparison for the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna* across different continents (top), climate zones (middle) and income groups (bottom). Model uncertainty was evaluated by calculating the average standard deviation for each continent, climate zone and income group. A lower standard deviation value indicates lower uncertainty.

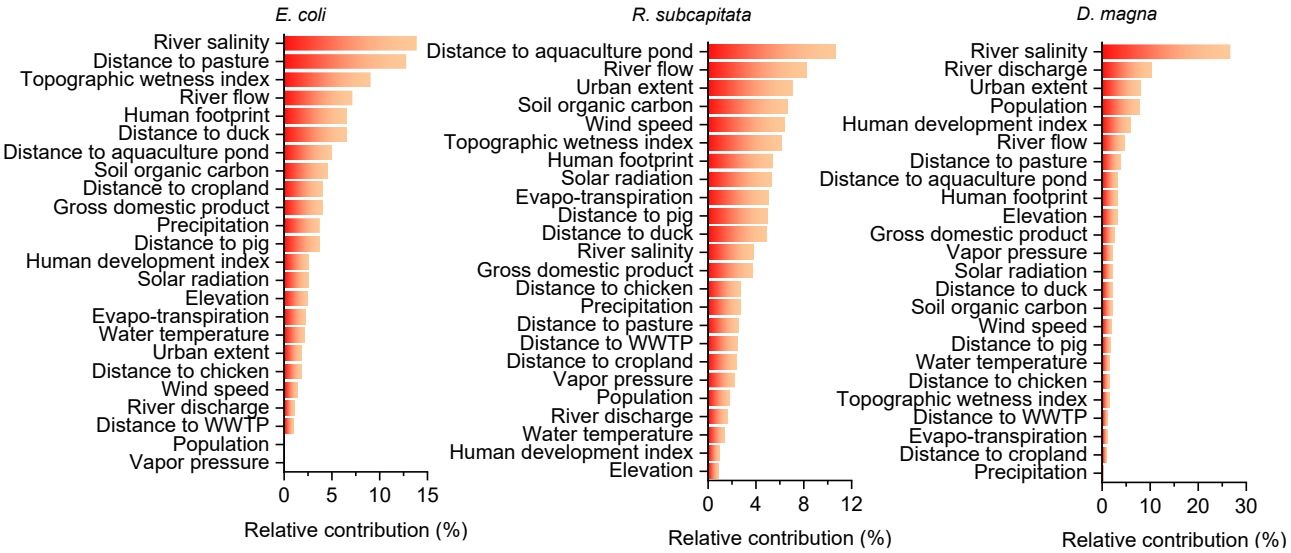


Fig. S7. The relative contributions of all geospatial predictors in the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna*.

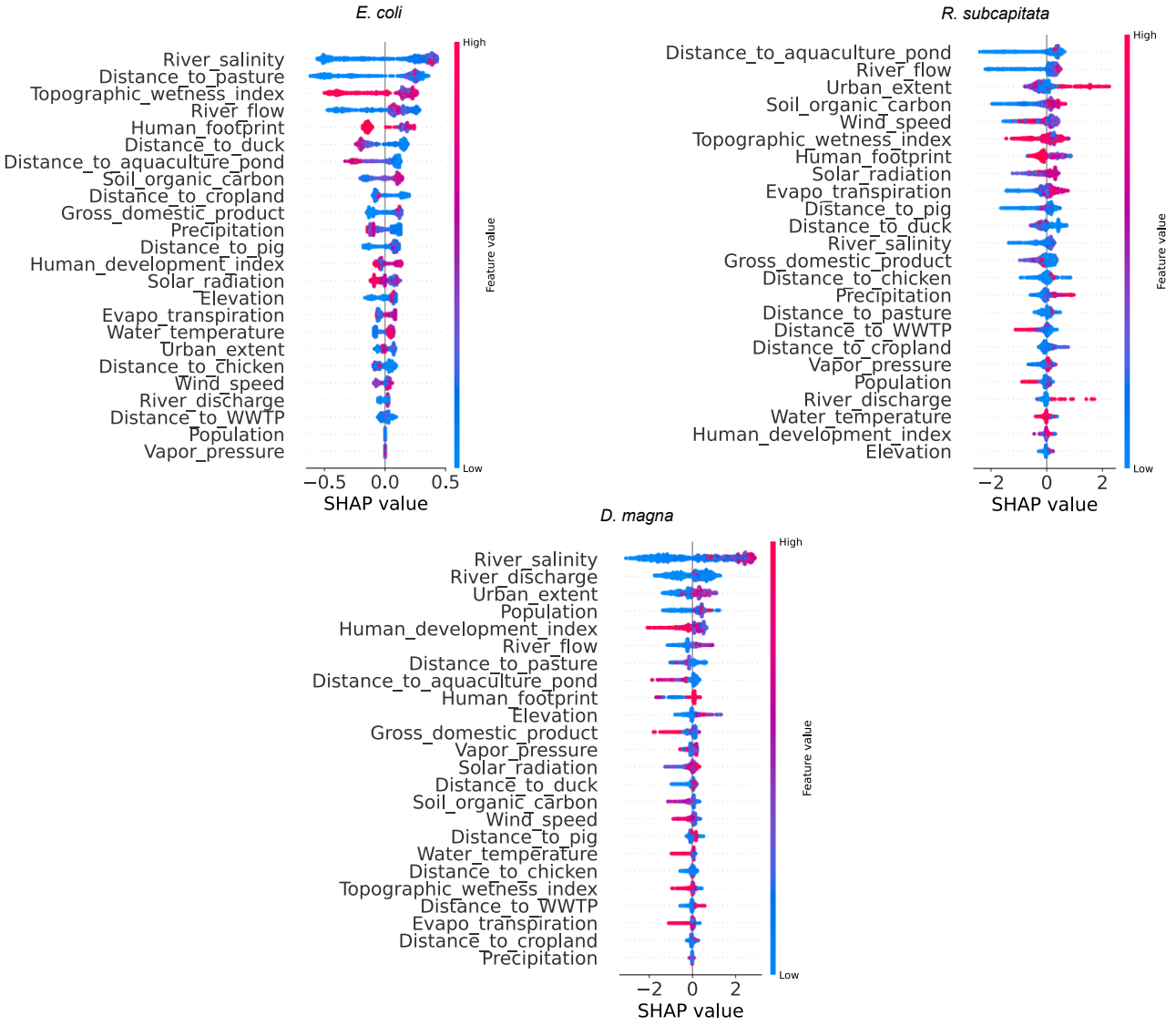


Fig. S8. SHapley Additive exPlanations (SHAP) plots for the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna*. SHAP values greater than 0 indicate a positive effect, and vice versa.

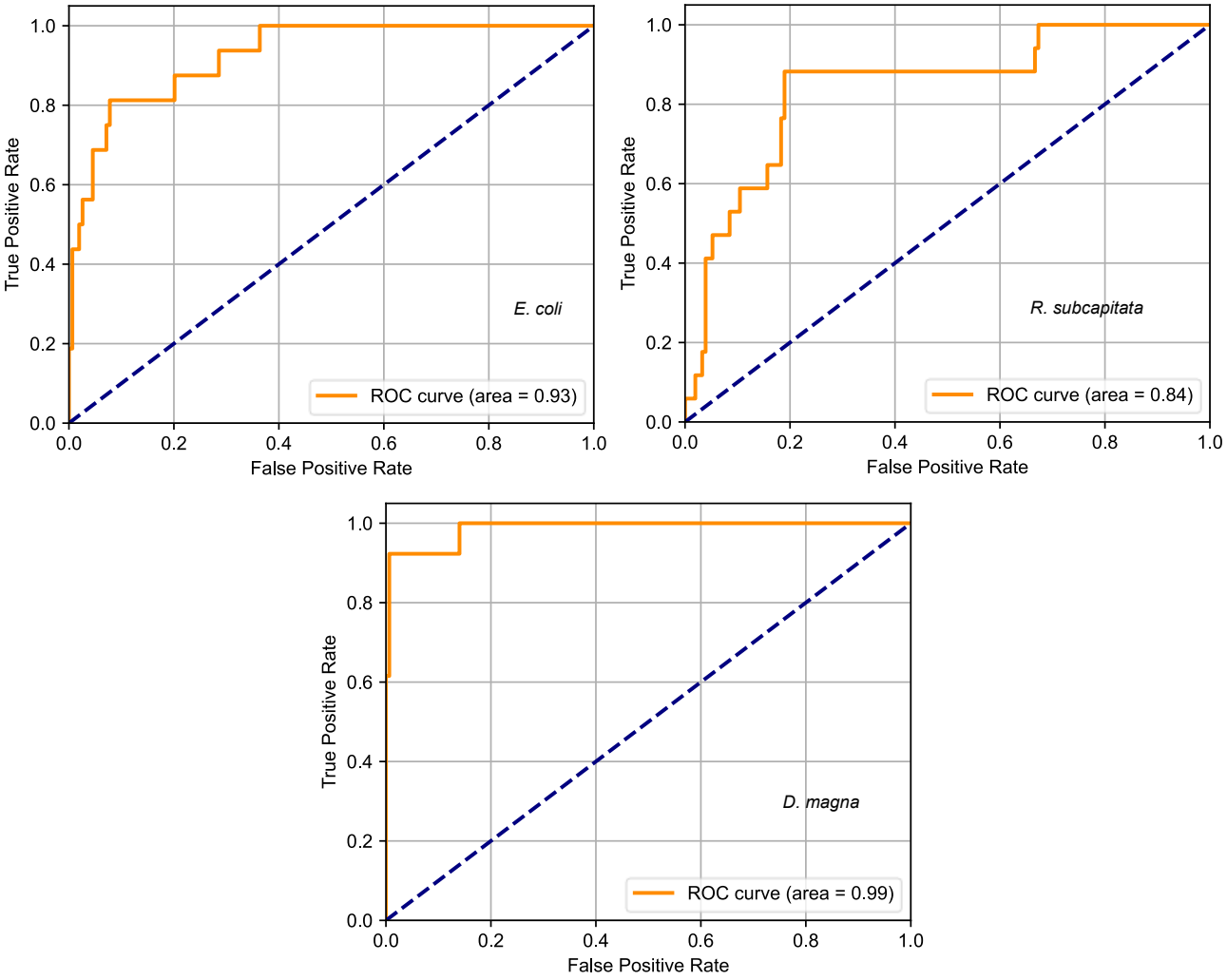


Fig. S9. The ROC curves for the three XGBoost models for predicting the API mixture toxicity to *E. coli*, *R. subcapitata* and *D. magna*, based on the 53 APIs.

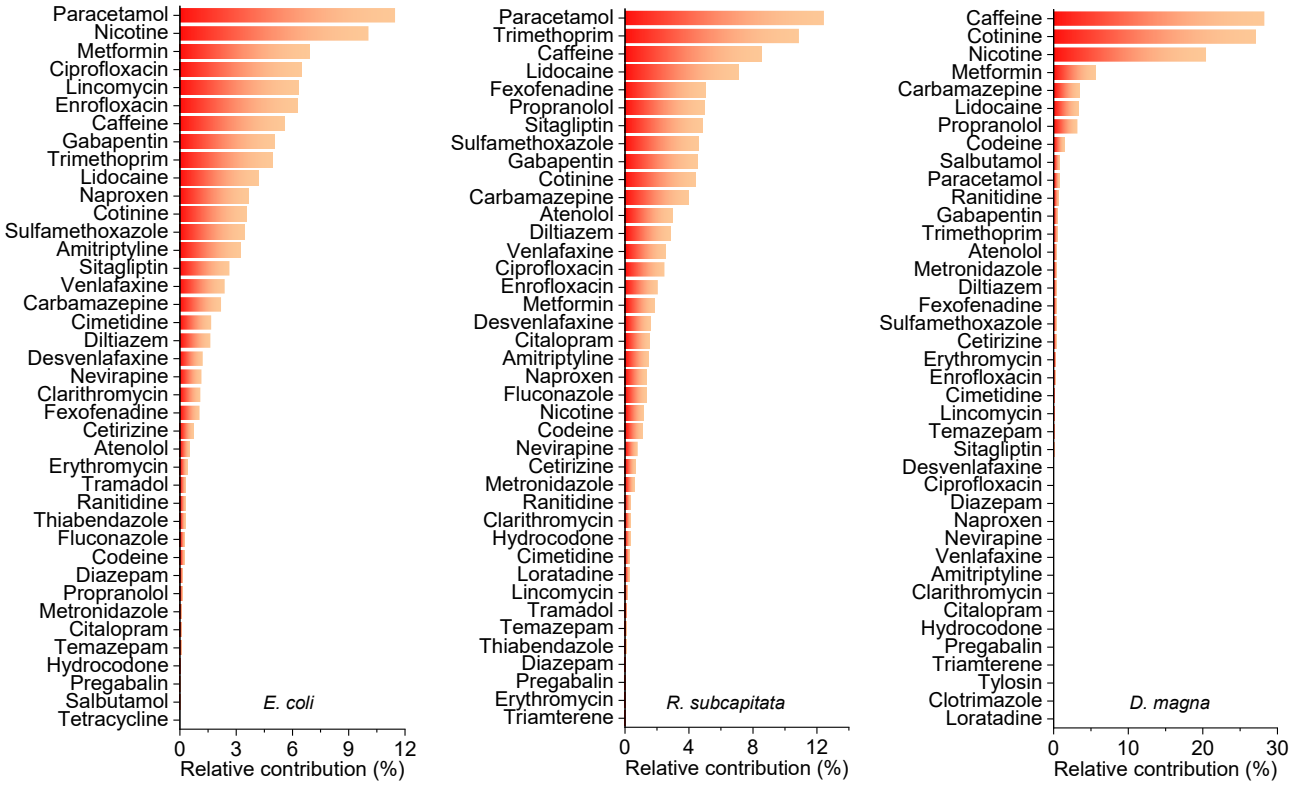


Fig. S10. The relative contributions of all APIs in the three XGBoost models for predicting the API mixture toxicity to *E. coli*, *R. subcapitata* and *D. magna*.

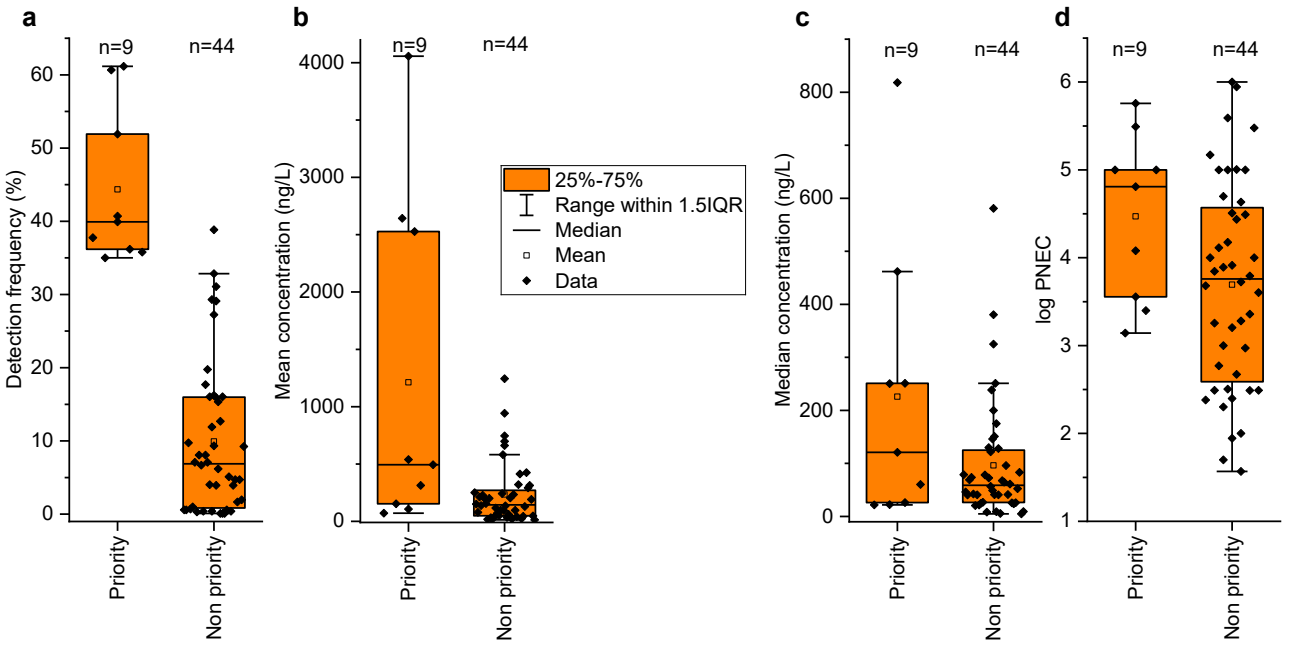


Fig. S11. The comparison of occurrence and toxicity between 9 priority and 44 non-priority APIs. The detection frequency (a), mean (b) and median (c) concentrations at 1,017 river sites. (d) The predicted no effect concentrations (PNEC), which is originated from a previous study[22]. Lower PNEC values indicate higher toxicity to aquatic organisms.

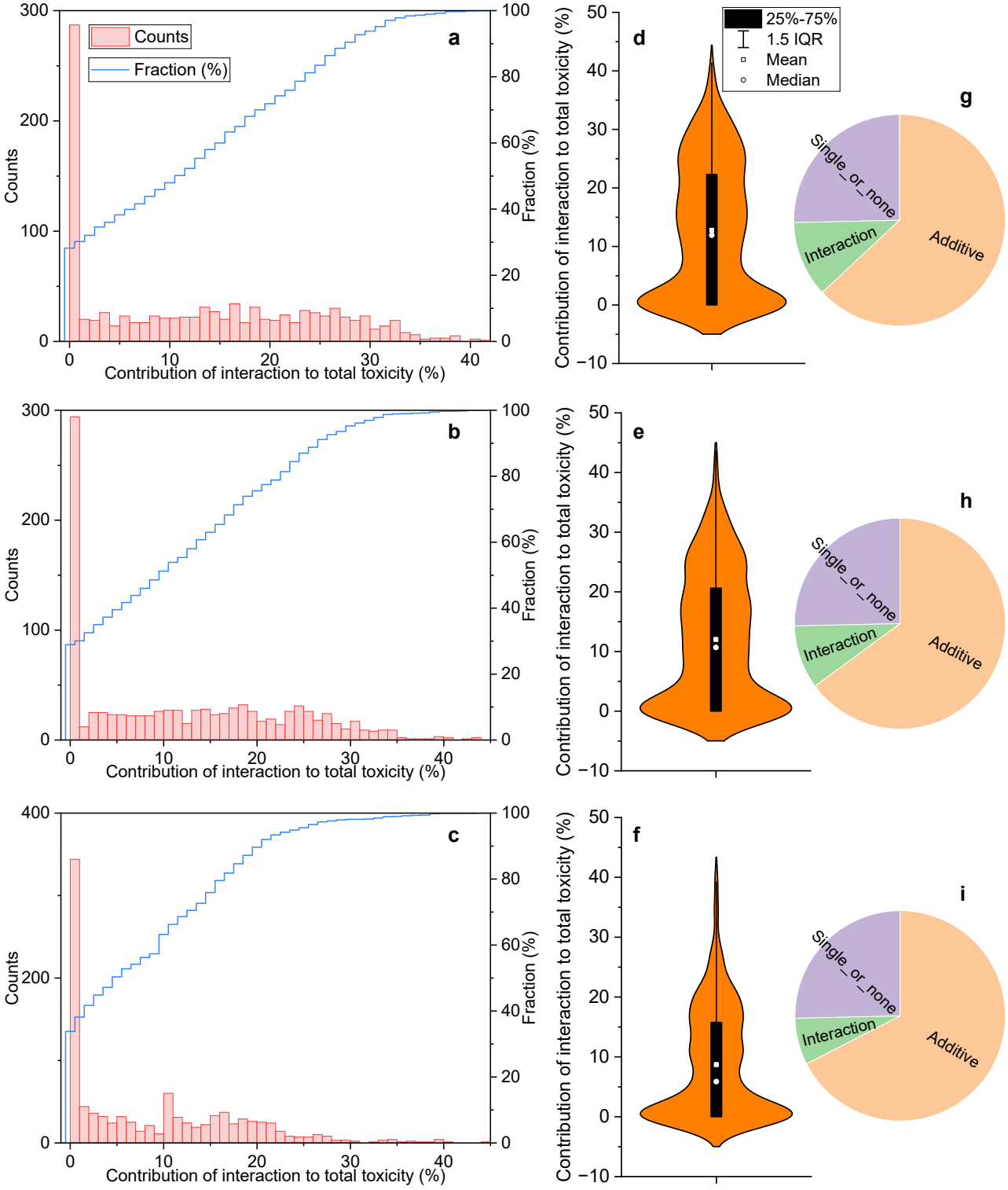


Fig. S12. Contributions of API interaction to total mixture toxicity across 1,017 mixtures for *E. coli* (top), *R. subcapitata* (middle), and *D. magna* (bottom), shown as probability distributions (a-c), box plots (d-f) and pie charts (g-i). Single_or_none: none or only one API in this mixture. Additive: no interaction.

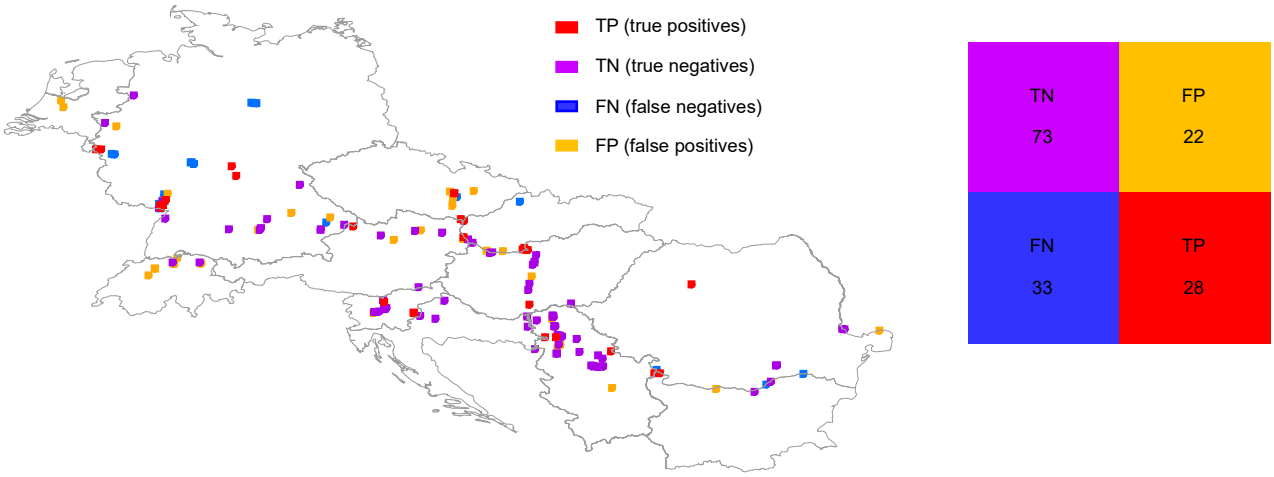


Fig. S13. External validation (left) and confusion matrix (right) for the XGBoost model with an independent dataset comprising 156 riverine grid cells (2 km) across 12 European countries. TP (true positives) are samples correctly identified as exceedance; FN (false negatives) are samples incorrectly identified as non-exceedance; TN (true negatives) are samples correctly identified as non-exceedance; FP (false positives) are samples incorrectly identified as exceedance. For this external validation, we first collected the detected concentrations of 138 APIs from 187 river sites in the Danube and Rhine River Basins across 12 European countries, as reported in a recent monitoring study[23]. Following the method described in the main text, we then prepared the 187 mixture solutions containing combinations of the 138 APIs based on their actual detected concentrations at the 187 riverine sites. Utilizing these 187 mixture solutions, we conducted acute mixture toxicity tests on *R. subcapitata*. The geometric mean values of API mixture toxicity were calculated for each 2 × 2 km riverine grid cells (n = 156), forming an independent dataset alongside 24 geospatial predictors. Our XGBoost model, trained on 53 APIs, achieved an accuracy of 0.65 on this independent dataset containing 138 APIs. It is important to note that our XGBoost models were trained to learn the statistical relationship between 24 predictors and the mixture toxicity of 53 APIs (i.e., response variable "*y*"). Because this 138-API dataset is not a comparable validation set but rather a different response variable "*y*" altogether, the resulting accuracy of 0.65 should not be interpreted as a failure of our model robustness. This external dataset of 138 APIs is used only to extend the limitations discussion, highlighting that considering only 53 APIs may underestimate the overall API risk.

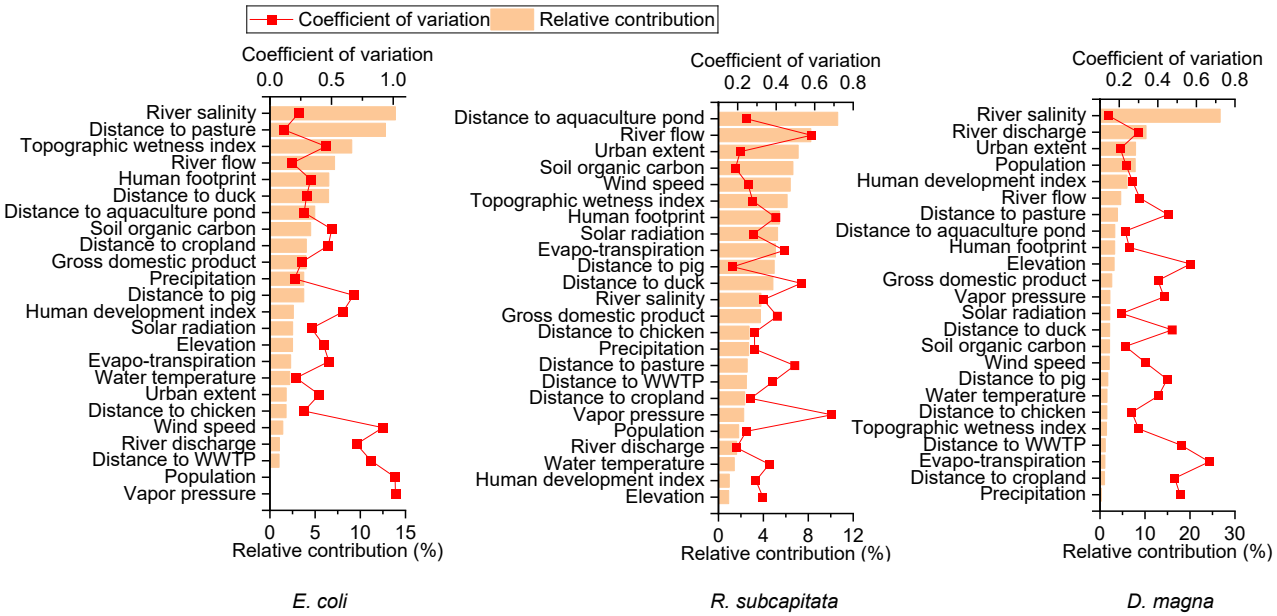


Fig. S14. The relative importance and coefficient of variation of SHAP values for 24 predictors in the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna*.

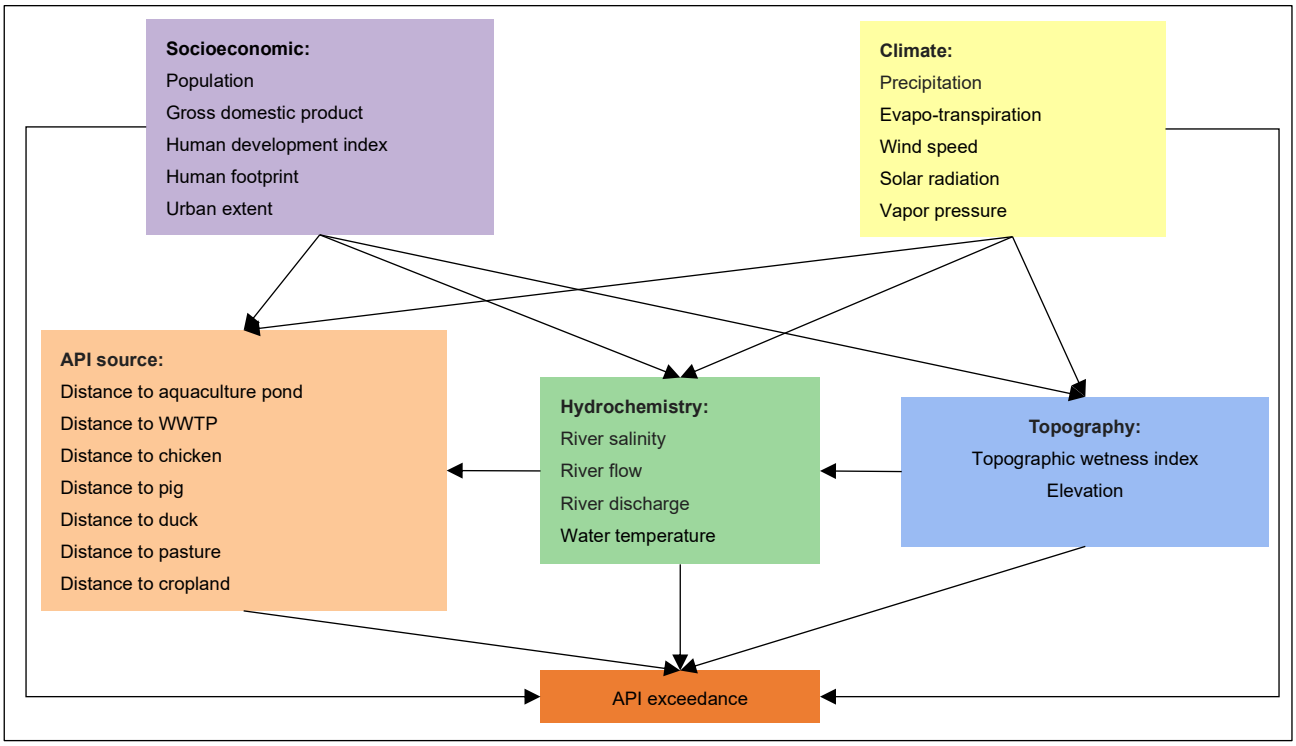


Fig. S15. A priori structural equation modeling (SEM) of associations among socioeconomic (population, gross domestic product, human development index, human footprint, and urban extent), climate (precipitation, Evapo-transpiration, wind speed, solar radiation, and vapor pressure), hydrochemistry (river salinity, river flow, river discharge, and water temperature), topography (topographic wetness index and elevation), and API source proximity (inverse distance to the nearest source; distance to aquaculture pond, distance to WWTP, distance to chicken, distance to pig, distance to duck, distance to pasture, and distance to cropland) parameters and the corresponding relationships with the distribution and exceedance of APIs in rivers.

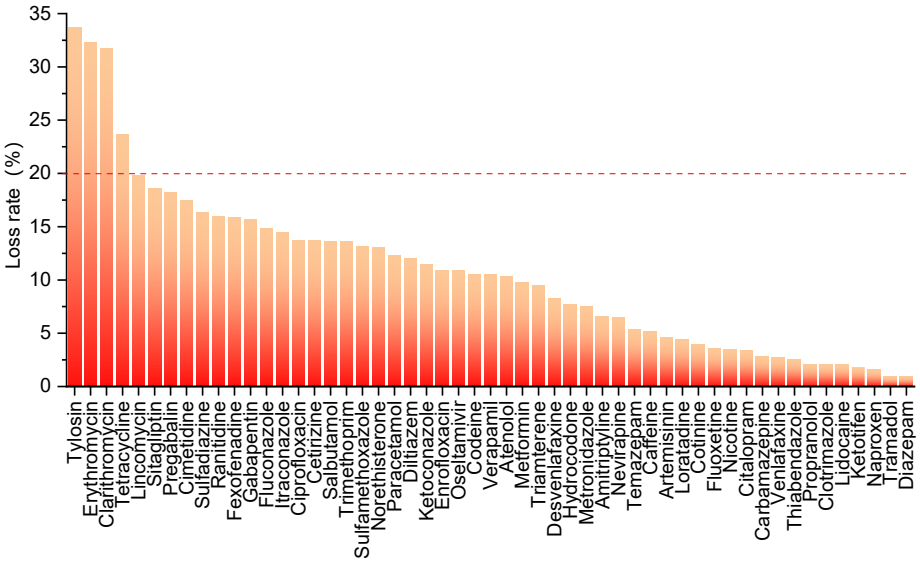


Fig. S16. The mass loss rate of 53 APIs under the 72 h exposure conditions.

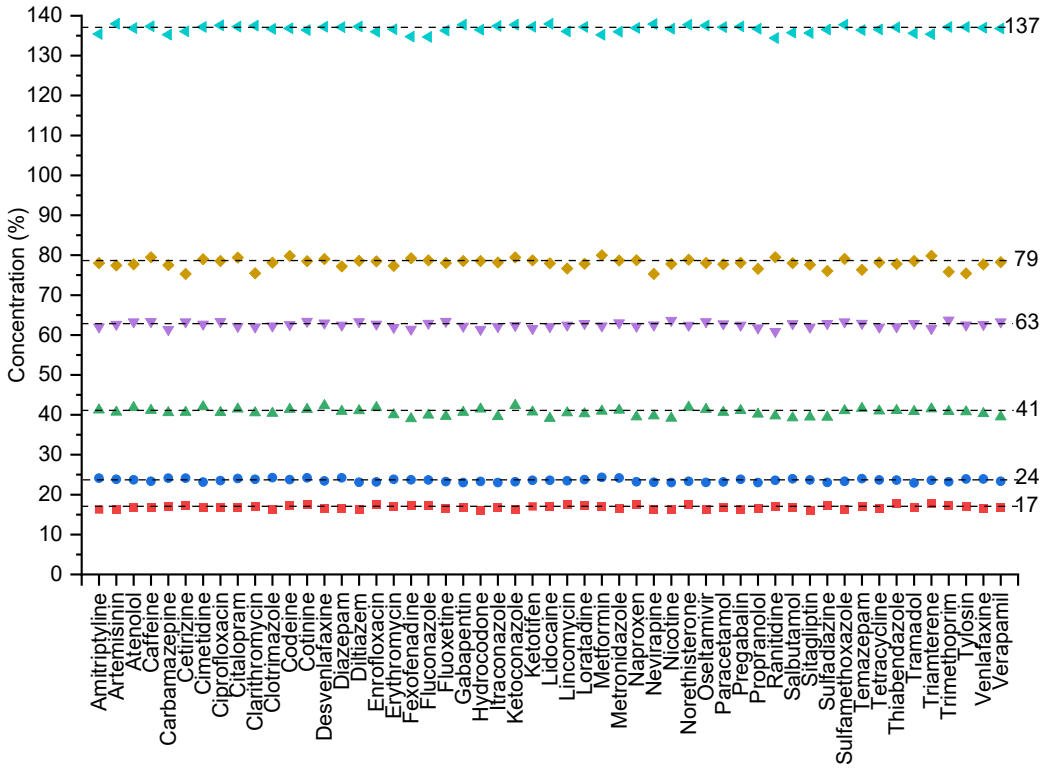


Fig. S17. Comparison between nominal concentration (17, 24, 41, 63, 79 and 137 ng/L) and measured acute concentration of 53 APIs in a verification experiment.

Table S1. The detailed information of the selected 53 APIs.

ID	API	Therapeutic class	Class abbr
1	Amitriptyline	Antidepressant	Depres
2	Artemisinin	Antimalarial	Antimal
3	Atenolol	Beta-blocker	β-block
4	Caffeine	Non-pharmaceutical stimulants/lifestyle	Life
5	Carbamazepine	Anticonvulsants	Anticon
6	Cetirizine	Antihistamine	Antihis
7	Cimetidine	Histamine H2 receptor antagonist	H2-rece
8	Ciprofloxacin	Antibiotics	Antibio
9	Citalopram	Antidepressant	Depres
10	Clarithromycin	Antibiotics	Antibio
11	Clotrimazole	Antifungal	Fungal
12	Codeine	Opioid pain medication	Opioid
13	Cotinine	Non-pharmaceutical stimulants/lifestyle	Life
14	Desvenlafaxine	Antidepressant	Depres
15	Diazepam	Benzodiazepine	Benzod
16	Diltiazem	Calcium channel blocker	Ca
17	Enrofloxacin	Antibiotics	Antibio
18	Erythromycin	Antibiotics	Antibio
19	Fexofenadine	Antihistamine	Antihis
20	Fluconazole	Antifungal	Fungal

21	Fluoxetine	Antidepressant	Depres
22	Gabapentin	Anticonvulsants	Anticon
23	Hydrocodone	Opioid pain medication	Opioid
24	Itraconazole	Antifungal	Fungal
25	Ketoconazole	Antifungal	Fungal
26	Ketotifen	Antihistamine	Antihis
27	Lidocaine	Analgesic	Analge
28	Lincomycin	Antibiotics	Antibio
29	Loratadine	Antihistamine	Antihis
30	Metformin	Antihyperglycemic	Antihy
31	Metronidazole	Antibiotics	Antibio
32	Naproxen	Analgesic	Analge
33	Nevirapine	Antiviral	Antiv
34	Nicotine	Non-pharmaceutical stimulants/lifestyle	Life
35	Norethisterone	Oral contraceptive	Oral-con
36	Oseltamivir	Antiviral	Antiv
37	Paracetamol	Analgesic	Analge
38	Pregabalin	Anticonvulsants	Anticon
39	Propranolol	Beta-blocker	β-block
40	Ranitidine	Histamine H2 receptor antagonist	H2-rece
41	Salbutamol	β2 adrenergic receptor agonist (anti-asthma)	Anti-asthma
42	Sitagliptin	Antihyperglycemic	Antihy

43	Sulfadiazine	Antibiotics	Antibio
44	Sulfamethoxazole	Antibiotics	Antibio
45	Temazepam	Benzodiazepine	Benzod
46	Tetracycline	Antibiotics	Antibio
47	Thiabendazole	Antifungal	Fungal
48	Tramadol	Opioid pain medication	Opioid
49	Triamterene	Diuretic	Diure
50	Trimethoprim	Antibiotics	Antibio
51	Tylosin	Antibiotics	Antibio
52	Venlafaxine	Antidepressant	Depres
53	Verapamil	Calcium channel blocker	Ca

Table S2. Performances of the three XGBoost models for predicting the API exceedance to *E. coli*, *R. subcapitata* and *D. magna*, based on the 24 geographical parameters.

	E. coli	*R. subcapitata*	*D. magna*
Accuracy	0.8239	0.9057	0.9434
AUC	0.8102	0.9547	0.9698
Sensitivity	0.7273	0.8462	0.8000
Specificity	0.8311	0.9110	0.9530

Table S3. Performances of the three XGBoost models for predicting the API mixture toxicity to *E. coli*, *R. subcapitata* and *D. magna*, based on the 53 APIs.

	E. coli	*R. subcapitata*	*D. magna*
Accuracy	0.9118	0.8118	0.9706
AUC	0.9274	0.8416	0.9873
Sensitivity	0.8125	0.8824	0.9231
Specificity	0.9221	0.8039	0.9745

Table S4. Full information of 40 predictors used in the machine learning models.

Group	Variables (abbreviation)	Resolution	Source
Climate	Air temperature, precipitation, solar radiation, wind speed, water vapor pressure	1 km	24
Climate	Aridity index, potential evapotranspiration (PET), actual evapotranspiration (AET), priestley-Taylor alpha coefficient (AET/PET)	1 km	25
Socioeconomic	Human footprint	1 km	26
Socioeconomic	Gross domestic product (GDP)	1 km	27
Socioeconomic	Urban extent	1 km	28
Socioeconomic	GDP per capita (GDP_per), Human development index (HDI)	1 km	29
Socioeconomic	Population density	1 km	30
Hydrochemistry	River flow, discharge, water temperature, salinity, total dissolved solid	10 km	31
Hydrochemistry	Adaptive capacity, ecological vulnerability	1 km	26
Topography	Elevation, slope	1 km	32
Topography	Topographic wetness index	1 km	33
Topography	Terrain ruggedness index, compound topographic index, stream power index	250 m	34
Soil	Soil water content	1 km	35
Soil	Soil pH, cation exchange capacity	1 km	36
Soil	Soil organic carbon	1 km	37
API source	Distance to the nearest site with 8 potential sources of API, including WWTP, landfill, aquaculture pond, cropland, pasture, pig, duck, and chicken farms	Distance (m)	38-43

Table S5. The performances of other four models, including random forest (RF), support vector machine (SVM), K-nearest neighbor (KNN), and neural network (NN), for predicting the API exceedance to *D. magna*, *R. subcapitata* and *E. coli*, based on the 24 geospatial predictors.

		RF	SVM	KNN	NN
D. magna	Accuracy	0.9434	0.8491	0.8931	0.9057
	AUC	0.9591	0.9268	0.9711	0.9228
	Sensitivity	0.6000	0.9000	1.0000	0.6000
	Specificity	0.9664	0.8456	0.8859	0.9262
R. subcapitata	Accuracy	0.9119	0.8239	0.8428	0.8994
	AUC	0.9305	0.8720	0.9370	0.8688
	Sensitivity	0.5385	0.6923	1.0000	0.6923
	Specificity	0.9452	0.8356	0.8288	0.9178
E. coli	Accuracy	0.9119	0.7610	0.7736	0.8679
	AUC	0.8004	0.6904	0.7110	0.8170
	Sensitivity	0.4545	0.5455	0.5455	0.5455
	Specificity	0.9459	0.7770	0.7905	0.8919

Table S6. The optimized hyperparameters in the three XGBoost models for predicting the API exceedance to *D. magna*, *R. subcapitata* and *E. coli*, based on the 24 geospatial predictors.

Hyperparameter	*D. magna*	*R. subcapitata*	*E. coli*
n_estimators	100	120	75
max_depth	6	6	6
learning_rate	0.05	0.045	0.042
min_child_weight	1	1	6
gamma	0	0	0
subsample	1	1	0.1
colsample_bytree	1	1	0.9

Table S7. The performances of other four models, including random forest (RF), support vector machine (SVM), K-nearest neighbor (KNN), and neural network (NN), for predicting API mixture toxicity to *D. magna*, *R. subcapitata* and *E. coli*, based on the 53 APIs.

		RF	SVM	KNN	NN
D. magna	Accuracy	0.9706	0.9471	0.9765	0.9765
	AUC	0.9824	0.9735	0.9515	0.9187
	Sensitivity	0.8462	0.5385	0.9231	0.8462
	Specificity	0.9809	0.9809	0.9809	0.9873
R. subcapitata	Accuracy	0.8765	0.4235	0.7471	0.7176
	AUC	0.8483	0.6874	0.7468	0.7785
	Sensitivity	0.5294	0.8235	0.7059	0.8824
	Specificity	0.9150	0.3791	0.7516	0.6993
E. coli	Accuracy	0.8647	0.8882	0.8294	0.7941
	AUC	0.8955	0.7163	0.8472	0.6213
	Sensitivity	0.5625	0.2500	0.8750	0.5000
	Specificity	0.8961	0.9545	0.8247	0.8247

Table S8. The optimized hyperparameters in the three XGBoost models for predicting the API exceedance to *D. magna*, *R. subcapitata* and *E. coli*, based on the 53 APIs.

Hyperparameter	*D. magna*	*R. subcapitata*	*E. coli*
n_estimators	300	110	160
max_depth	6	6	6
learning_rate	0.01	0.022	0.07
min_child_weight	1	1	1
gamma	0	0	0
subsample	1	1	1
colsample_bytree	1	1	1

Table S9. The composition of the BG11 medium[1].

	Substance	Concentration (g/L)
	$NaNO_3$	1.500
	$MgSO_4 \cdot 7H_2O$	0.075
	K_2HPO_4	0.040
	$CaCl_2 \cdot 2H_2O$	0.036
	Na_2CO_3	0.020
	Citric acid	0.006
	Ferric ammonium citrate	0.006
	$EDTANa_2$	0.001
A5	H_3BO_3	2.860
	$MnCl_2 \cdot 4H_2O$	1.860
	$Na_2MoO_4 \cdot 2H_2O$	0.390
	$ZnSO_4 \cdot 7H_2O$	0.220
	$CuSO_4 \cdot 5H_2O$	0.080
	$Co(NO_3)_2 \cdot 6H_2O$	0.050

Table S10. The composition of the artificial freshwater for the cultivation for *D. magna*[2].

Stock solutions (Single substance)		Volume of stock solution
Substance	Initial concentration	added to 1 L water
Calcium chloride dihydrate	11.76 g/L	25 mL
Magnesium sulfate heptahydrate	4.93 g/L	25 mL
Sodium bicarbonate	2.59 g/L	25 mL
Potassium chloride	0.23 g/L	25 mL

References

1 OECD. *Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test*. (2011).

2 OECD. *Test No. 202: Daphnia sp. Acute Immobilisation Test*. (2004).

3 Li, X., Shi, J., Sun, H. & Lin, Z. Hormetic dose-dependent response about typical antibiotics and their mixtures on plasmid conjugative transfer of Escherichia coli and its relationship with toxic effects on growth. *Ecotoxicol. Environ. Saf.* **205**, 111300 (2020). https://doi.org/https://doi.org/10.1016/j.ecoenv.2020.111300

4 Chen, J. *et al.* A machine-learning approach clarifies interactions between contaminants of emerging concern. *One Earth* **5**, 1239-1249 (2022). https://doi.org/https://doi.org/10.1016/j.oneear.2022.10.006

5 Chen, J. *et al.* Mitigating pesticide mixture hazard in global surface waters through agricultural management. *One Earth* **8**, 101163 (2025). https://doi.org/https://doi.org/10.1016/j.oneear.2024.11.017

6 Rodea-Palomares, I. *et al.* Hidden drivers of low-dose pharmaceutical pollutant mixtures revealed by the novel GSA-QHTS screening method. *Sci. Adv.* **2**, e1601272 (2016). https://doi.org/10.1126/sciadv.1601272

7 Zhong, M. *et al.* Automated online solid-phase extraction liquid chromatography tandem mass spectrometry investigation for simultaneous quantification of per- and polyfluoroalkyl substances, pharmaceuticals and personal care products, and organophosphorus flame retardants in environmental waters. *J. Chromatogr. A* **1602**, 350-358 (2019). https://doi.org/https://doi.org/10.1016/j.chroma.2019.06.012

8 Wilkinson, J. L. *et al.* Pharmaceutical pollution of the world's rivers. *Proc. Natl. Acad.*

Sci. U.S.A. **119**, e2113947119 (2022). https://doi.org/doi:10.1073/pnas.2113947119

9 Li, S. *et al.* Antibiotics in global rivers. *Nati. Sci. Open* **1**, 20220029 (2022). https://doi.org/https://doi.org/10.1360/nso/20220029

10 Zhao, L. *et al.* Unveiling PFAS hazard in European surface waters using an interpretable machine-learning model. *Environ. Int.*, **199**, 109504 (2025). https://doi.org/https://doi.org/10.1016/j.envint.2025.109504

11 Zhao, F. *et al.* Reducing risks of antibiotics to crop production requires land system intensification within thresholds. *Nat. Commun.* **14**, 6094 (2023). https://doi.org/10.1038/s41467-023-41258-x

12 Sheikholeslami, R. & Hall, J. W. Global patterns and key drivers of stream nitrogen concentration: A machine learning approach. *Sci. Total Environ.* **868**, 161623 (2023). https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.161623

13 Tokranov, A. K. *et al.* Predictions of groundwater PFAS occurrence at drinking water supply depths in the United States. *Science* **386**, 748-755 (2024). https://doi.org/doi:10.1126/science.ado6638

14 Wang, J., Shao, S., Gao, Q., Wang, B. & Zhang, Y. Mapping PFAS exceedance risk in China's surface water: A machine learning approach informed by source distribution. *Environ. Sci. Technol.* **60**, 13462-13472 (2026). https://doi.org/10.1021/acs.est.6c00574

15 Chen, J. *et al.* Global mapping of pharmaceutical ecological risk in rivers using machine learning: drivers, hotspots, and compounded water stress. *Water Res.* **291**, 125184 (2026). https://doi.org/https://doi.org/10.1016/j.watres.2025.125184

16 Hughes, S. R., Kay, P. & Brown, L. E. Global synthesis and critical evaluation of

pharmaceutical data sets collected from river systems. *Environ. Sci. Technol.* **47**, 661-677 (2013). https://doi.org/10.1021/es3030148

17 Ehalt Macedo, H., Lehner, B., Nicell, J. A., Khan, U. & Klein, E. Y. Antibiotics in the global river system arising from human consumption. *PNAS Nexus* **4** (2025). https://doi.org/10.1093/pnasnexus/pgaf096

18 Du, Z. *et al.* Annual global grided livestock mapping from 1961 to 2021. *Earth Syst. Sci. Data* **17**, 5543-5556 (2025). https://doi.org/10.5194/essd-17-5543-2025

19 Liu, Y. *et al.* Profiles, drivers, and prioritization of antibiotics in China's major rivers. *J. Hazard. Mater.* **477**, 135399 (2024). https://doi.org/https://doi.org/10.1016/j.jhazmat.2024.135399

20 Hu, J. *et al.* Animal production predominantly contributes to antibiotic profiles in the Yangtze River. *Water Res.* **242**, 120214 (2023). https://doi.org/https://doi.org/10.1016/j.watres.2023.120214

21 Lombard, M. A. *et al.* Estimating lithium concentrations in groundwater used as drinking water for the conterminous united states. *Environ. Sci. Technol.* **58**, 1255-1264 (2024). https://doi.org/10.1021/acs.est.3c03315

22 Bouzas-Monroy, A., Wilkinson, J. L., Melling, M. & Boxall, A. B. A. Assessment of the potential ecotoxicological effects of pharmaceuticals in the world's rivers. *Environ. Toxicol. Chem.* **41**, 2008-2020 (2022). https://doi.org/https://doi.org/10.1002/etc.5355

23 Finckh, S. *et al.* Mapping chemical footprints of organic micropollutants in European streams. *Environ. Int.* **183**, 108371 (2024). https://doi.org/https://doi.org/10.1016/j.envint.2023.108371

24 Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces

for global land areas. *Int. J. Climatol.* **37**, 4302-4315 (2017). https://doi.org/https://doi.org/10.1002/joc.5086

25 Trabucco, A. Z., Robert. *Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v3. figshare. Dataset. https://doi.org/10.6084/m9.figshare.7504448.v4*, 2019).

26 Varis, O., Taka, M. & Kummu, M. The planet's stressed river basins: Too much pressure or too little adaptive capacity? *Earth's Future* **7**, 1118-1135 (2019). https://doi.org/https://doi.org/10.1029/2019EF001239

27 Chen, J. *et al.* Global 1 km × 1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data. *Sci. Data* **9**, 202 (2022). https://doi.org/10.1038/s41597-022-01322-5

28 Corbane, C. P., et al., GHSL Data Package 2019, EUR 29788 EN, Publications Office of the European Union, Luxembourg, (2019), ISBN 978-92-76-13186-1, doi:10.2760/290498, JRC 117104.

29 Kummu, M., Taka, M. & Guillaume, J. H. A. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Sci. Data* **5**, 180004 (2018). https://doi.org/10.1038/sdata.2018.4

30 Network, C. f. I. E. S. I. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 to 2020 Revision of UN WPP Country Totals, Revision 11.: Columbia University (2018).

31 Jones, E. R. *et al.* DynQual v1.0: a high-resolution global surface water quality model. *Geosci. Model Dev.* **16**, 4481-4500 (2023). https://doi.org/10.5194/gmd-16-4481-2023

32 Amatulli, G. *et al.* A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci. Data* **5**, 180040 (2018). https://doi.org/10.1038/sdata.2018.40

33 Hengl, T. Global DEM derivatives at 250 m, 1 km and 2 km based on the MERIT DEM (1.0) (2018). https://doi.org/10.5281/zenodo.1447210.

34 Amatulli, G., McInerney, D., Sethi, T., Strobl, P. & Domisch, S. Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. *Sci. Data* **7**, 162 (2020). https://doi.org/10.1038/s41597-020-0479-6

35 Trabucco, A.Z., Robert J. Global high-resolution soil-water balance (2019). https://doi.org/10.6084/m9.figshare.7707605.v3.

36 Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *Plos One* **12**, e0169748 (2017). https://doi.org/10.1371/journal.pone.0169748

37 FAO. Global Soil Organic Carbon Map v1.5 (GSOC). https://data.apps.fao.org/catalog/dataset/7730e747-eb73-49c9-bfe6-84ebae718743. (2019).

38 Ehalt Macedo, H. *et al.* Distribution and characteristics of wastewater treatment plants within the global river network. *Earth Syst. Sci. Data* **14**, 559-577 (2022). https://doi.org/10.5194/essd-14-559-2022

39 Xu, Y. *et al.* Global mapping of human-transformed dike-pond systems. *Remote Sens. Environ.* **313**, 114354 (2024). https://doi.org/https://doi.org/10.1016/j.rse.2024.114354

40 Potapov, P. *et al.* Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat. Food* **3**, 19-28 (2022). https://doi.org/10.1038/s43016-021-00429-z

41 Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**, 1-22 (2004). https://doi.org/https://doi.org/10.1016/j.jhydrol.2004.03.028

42 Gilbert, M. *et al.* Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **5**, 180227 (2018). https://doi.org/10.1038/sdata.2018.227

43 *World Landfills Database*. *https://www.geoslices.com/landfills*.