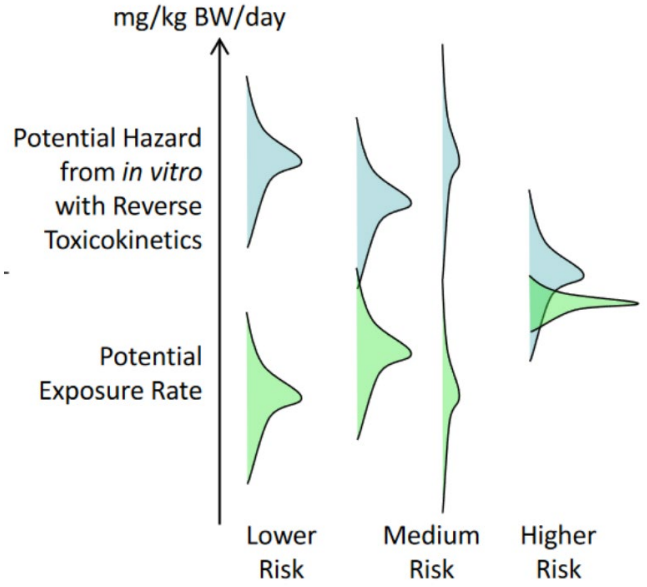
DETERMINING REPRESENTATIVE NATIONAL SURFACE WATER CHEMICAL CONCENTRATION RANGES FOR RISK PRIORITIZATION IN DRINKING WATER

Risa R. Sayre[1,2,3], *Marc Serre*[2], *R. Woodrow Setzer*[1*], *John F. Wambaugh*[1,2]

1: Center for Computational Toxicology and Exposure, U.S. Environmental Protection Agency; 2: Department of Environmental Sciences and Engineering, UNC Chapel Hill; 3: Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee; *Emeritus

WORK IN PROGRESS

With thousands of chemicals in commerce and the environment, efficient tools are needed to support risk prioritization and evaluation.



We present the framework for an open, reproducible workflow to:

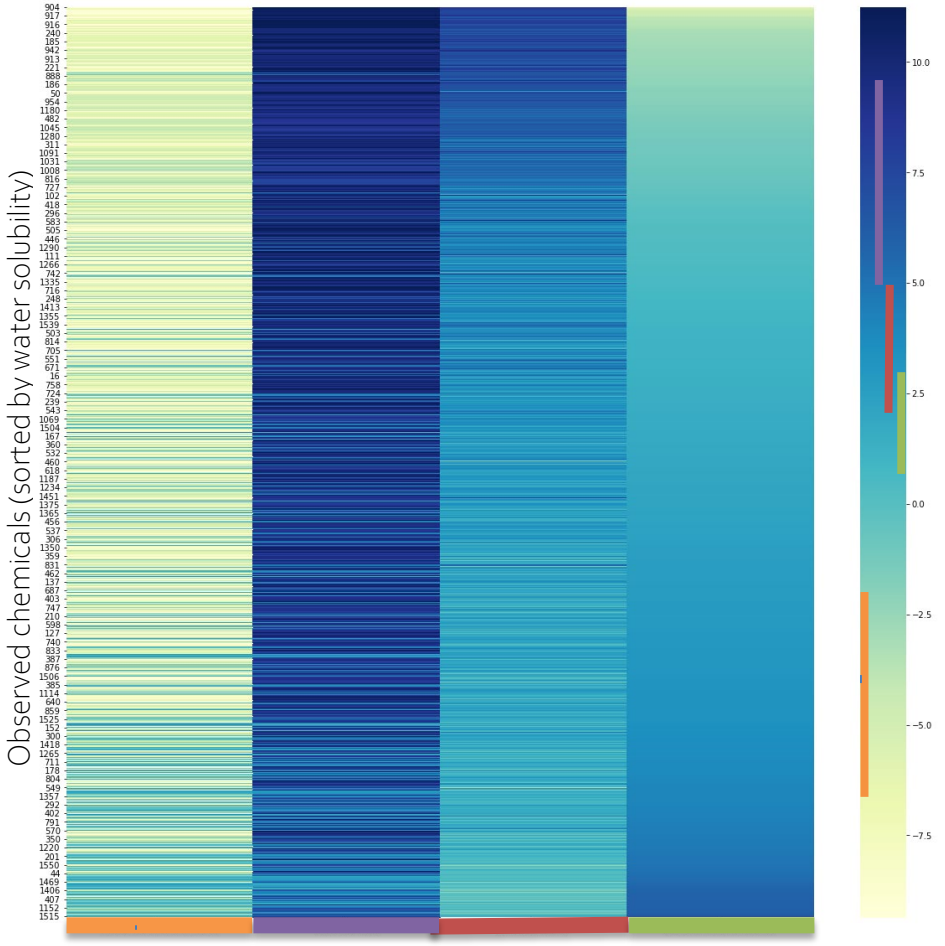
1. Determine representative surface water concentrations for hundreds of organic chemicals in the United States based on monitoring data
2. Calibrate a metamodel to predict representative surface water concentrations for thousands of non-monitored organic chemicals
3. *Estimate removal efficiency of organic chemicals by conventional drinking water treatment processes*
4. *Prioritize organic chemicals based on their estimated risk based on exposure from treated drinking water*

Data availability

Water Quality Portal

https://www.waterqualitydata.us/portal

- Data collected by over 400 state, federal, tribal, and local agencies
- Searched for any organic chemical samples in water for the contiguous U.S. from 2008 to 2018
  - Hundreds of chemical names
  - 2114 of 2270 hydrologic subbasins
  - Millions of rows of data: samples collected for many different reasons under different conditions



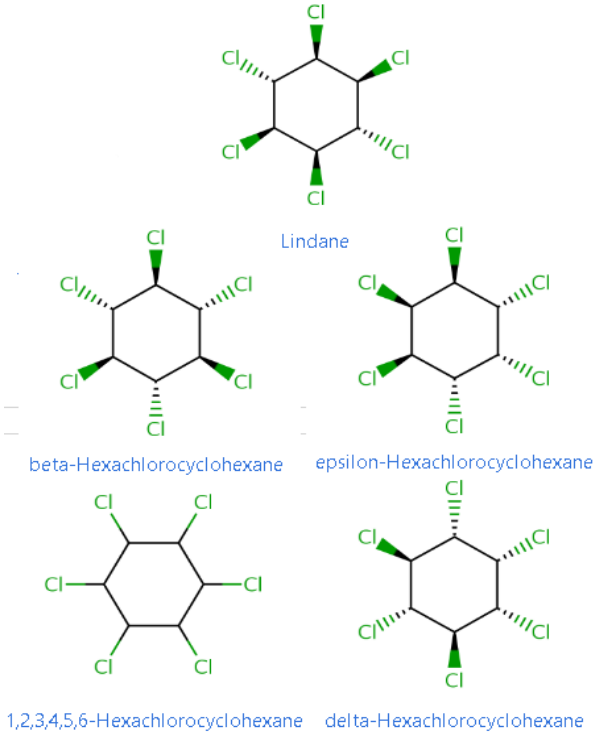
Chemical property space of observation set: vapor pressure (mmHg), octanol:air, octanol:water, water solubility (mg/L), all in log10 space. Lines on the colorbar indicate range of moderate values for the property. All physicochemical property predictions made using OPERA 4.2

Chemical curation

- 1626 names mapped to chemical structures using EPA's Chemicals Dashboard
- 111 names manually curated (usually just a different acceptable spelling)
- 117 not yet mapped
- 311 names that referred to mixtures, ambiguous structures, organometallics manually removed

Harmonized to unique structure → 1404

- about half are pesticides, as identified by EPA Chemicals Dashboard lists



Lindane, a Stockholm persistent organic pollutant, and several of its isomers

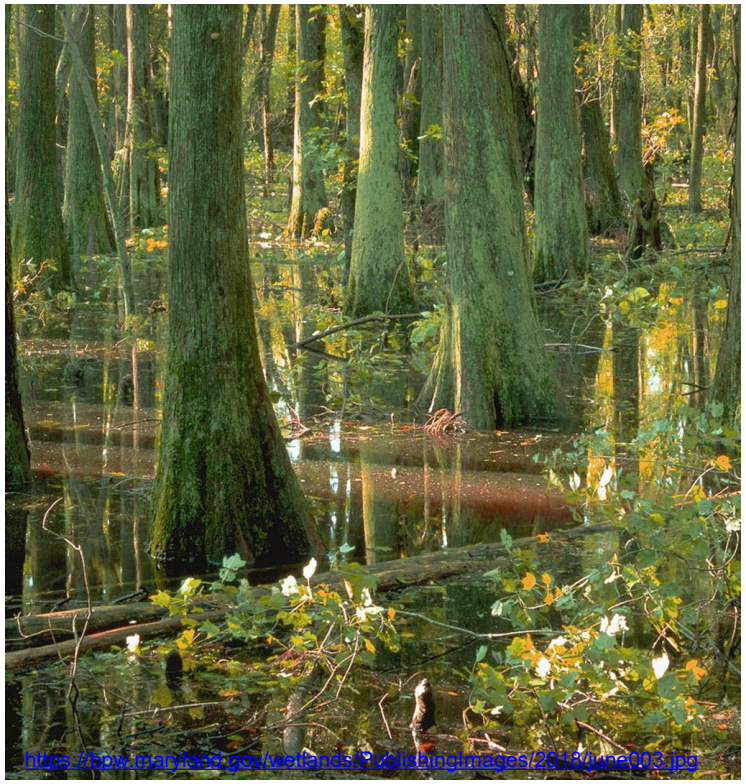
Metadata filters

Excluded sites:

- Not representative of ambient concentrations (Waste-injection well, sewer, finished water)
- Not surface water (Borehole, atmospheric)
- Not fresh water (Ocean, estuary)

Included sites:

Surface (some edge cases like palustrine wetland, hyporheic-zone/Ranney well, stormwater)

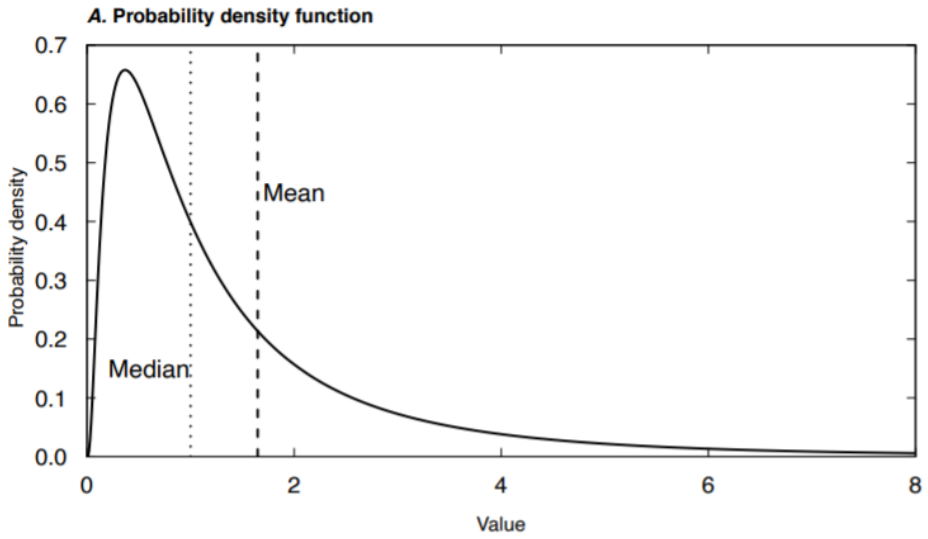


A palustrine wetland

Excluded activities: Not representative of ambient concentrations (blanks, spikes, leachate, initial dilution zone, radiolabeled)

Characteristics of distributions of chemical concentrations in water

- Lower bound of zero
- Presence of outliers
- Positive skewness
- Non-normal
- **Censored data**
- **Seasonal patterns**
- *Autocorrelation*
- *Dependence on environmental variables*

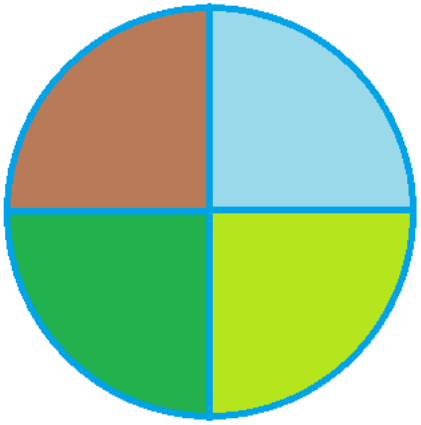


Probability density function (PDF) of a lognormal distribution

From USGS book, Statistical Methods in Water Resources: Section A, Statistical Analysis Book 4, Hydrologic Analysis and Interpretation

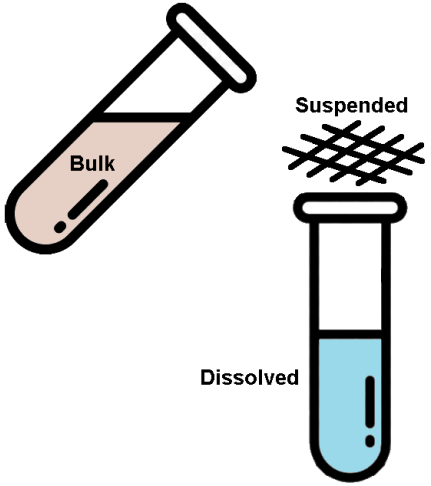
Differences between sample sets

Using two-sample Kolmogorov-Smirnov (KS) test to determine whether concentrations per chemical are "same" or "different", comparing sets by:



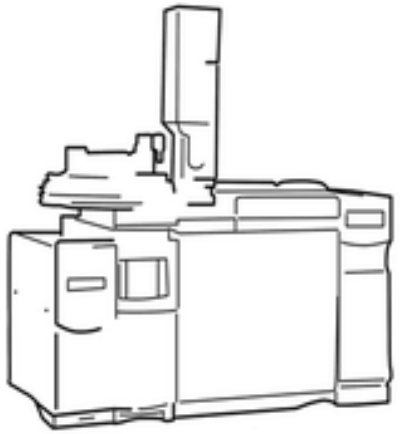
Season

Q1 (1191870) , Q2 (3106021), Q3 (2310940), Q4 (1239780)



Phase

Bulk (2246524), dissolved (4262598), unknown (1337479)

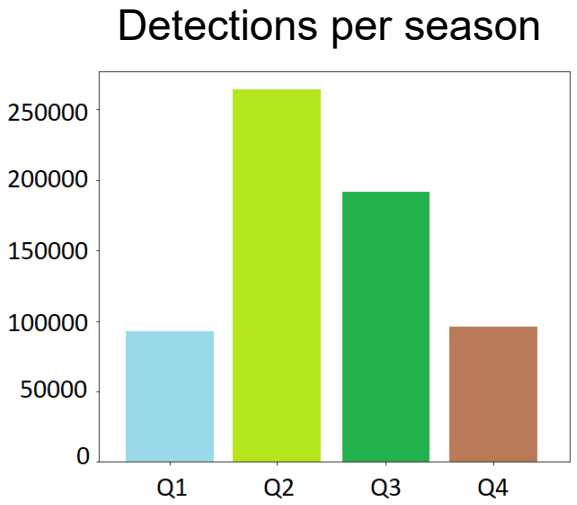


Limit value type

Tech min (3440638), tech quant (940072), reporting min (2262248), unknown (34828)

Seasonal concentration differences

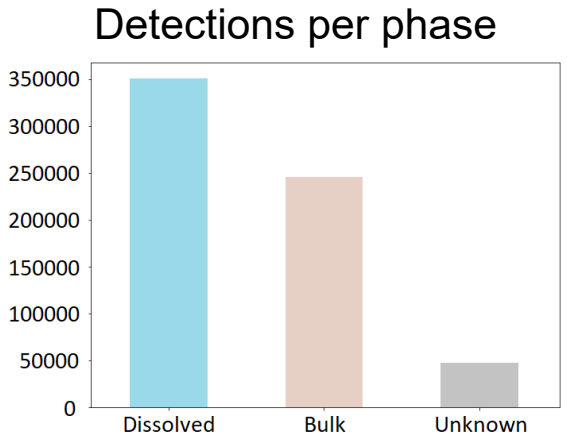
sets	% same	chemicals
Q1, Q2	86.5	1179
Q1, Q3	87.8	1188
Q1, Q4	89.1	1184
Q2, Q3	91.9	1314
Q2, Q4	88.6	1306
Q3, Q4	91.5	1332



Concentrations are **not different** for most chemicals when grouped by season

Phase concentration differences

sets	% same	chemicals
Unknown, dissolved	70.4	273
Unknown, bulk	71.2	245
Dissolved, bulk	67.3	334



Concentrations are different for some chemicals when grouped by phase, so we will build different representative concentrations

Results were independent of Log P

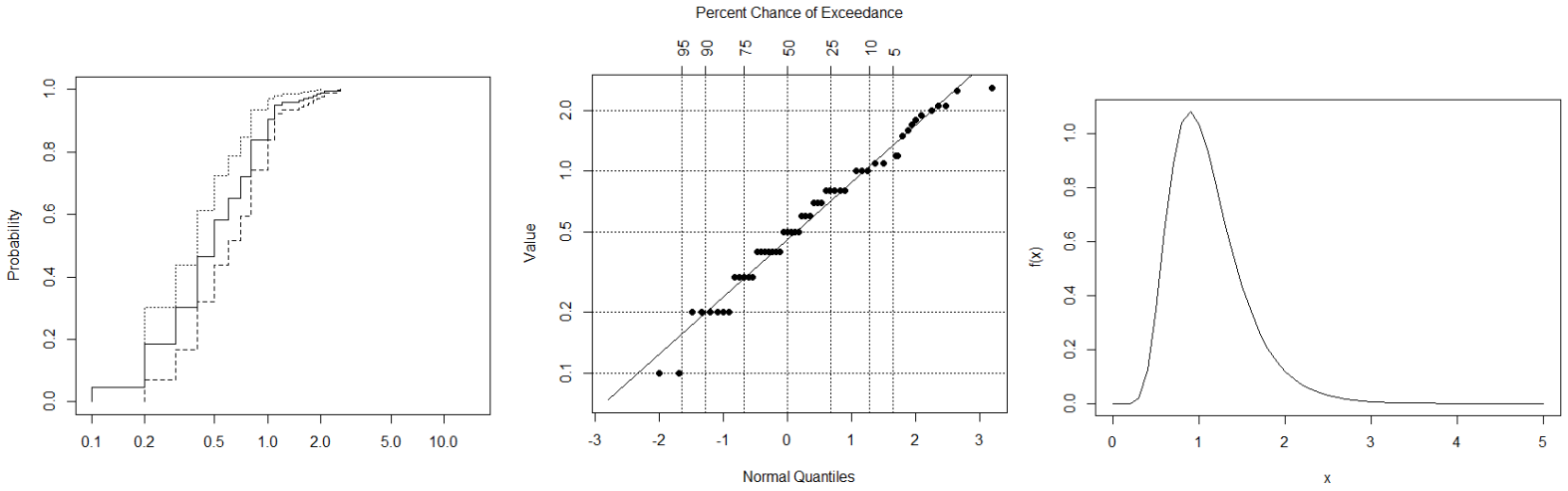
SW concentration distributions

Methods used for calculating summary values on samples with censored data:

Kaplan-Meier (K-M)

Robust regression-on-order statistic (ROS)

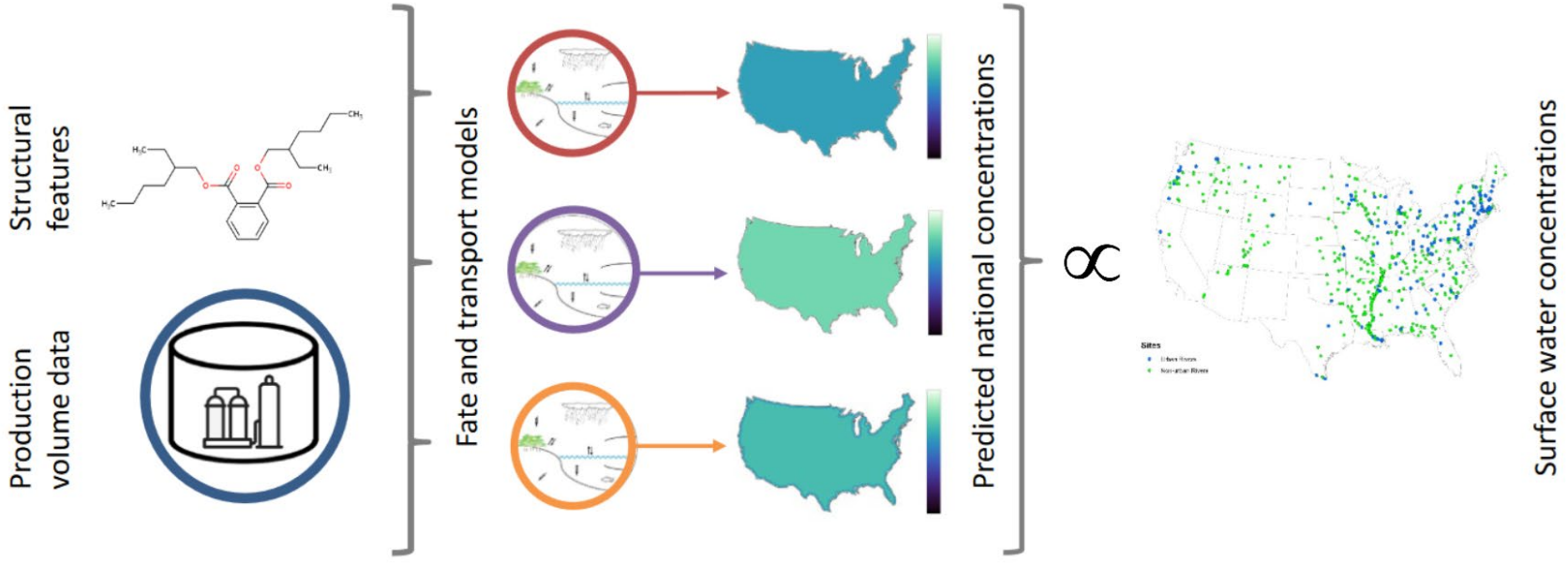
Maximum likelihood estimation (MLE)



Estimated concentration distributions using the three methods above for single chemical dissolved results with multiple censoring limits and 92.6% censored data.

Future work: investigate when each method works best

Metamodel framework



$$\ln y_i = m_0 + \sum_{j=1}^{n_j} \sum_{k=1}^{n_{kj}} m_{jk} \ln\left(l_{ji} p_{ki}\right)$$

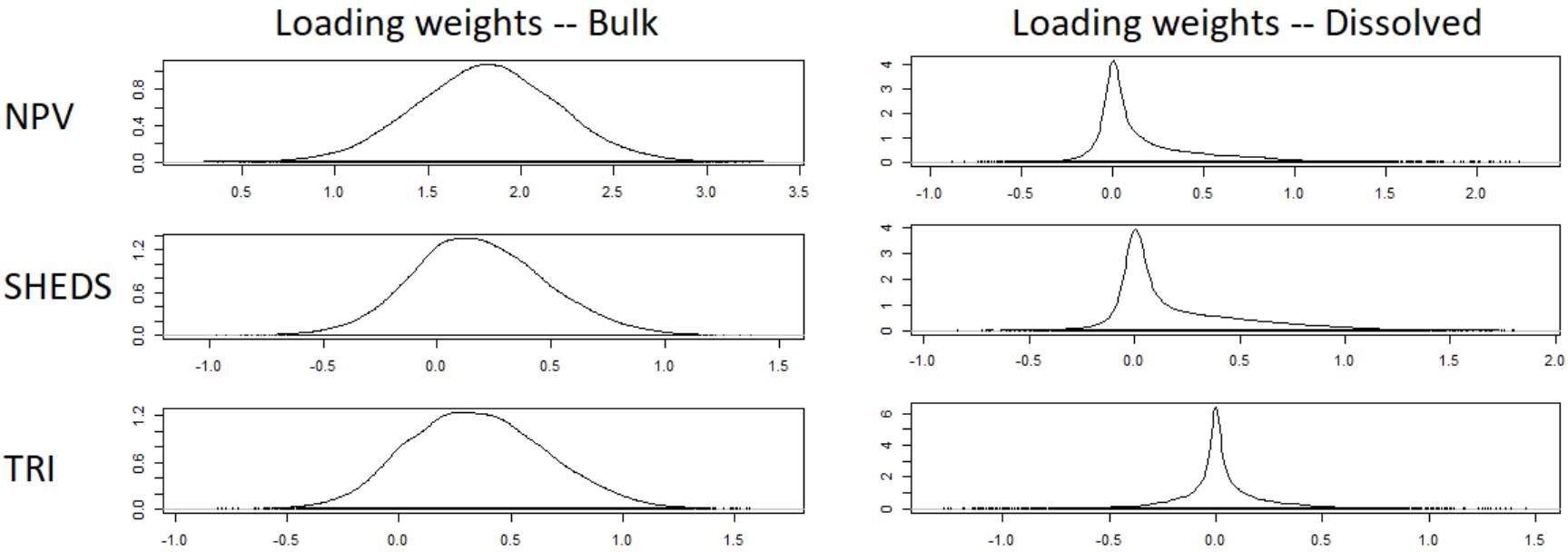
Loading sources (*j*)

- NPV: Chemical Data Reporting under the Toxic Substances Control Act (data reported to the EPA about the mass of chemicals imported into or produced in the U.S. by year) + Pesticides Industry Sales and Usage, 2008 – 2012 Market Estimates (EPA report)
- SHEDS-DTD (Stochastic Human Exposure and Dose Simulation): model simulating the amount of a chemical that goes down the drain based on household usage, reduced by the percentage removed by wastewater treatment
- Toxic Release Inventory (TRI): data reported to the EPA about industrial releases

$$\ln y_i = m_0 + \sum_{j=1}^{n_j} \sum_{k=1}^{n_{kj}} m_{jk} \ln\left(l_{ji}\, p_{ki}\right)$$

Where *i* is a given chemical, *y* is the representative concentration, *j* is a loading source, *l* is the value of that loading given the source and chemical (amount/time) , *k* is an exposure model, *p* is the value of an exposure model (amount/amount/time), and *m* is a model weight

Metamodel loading weights (m_j)



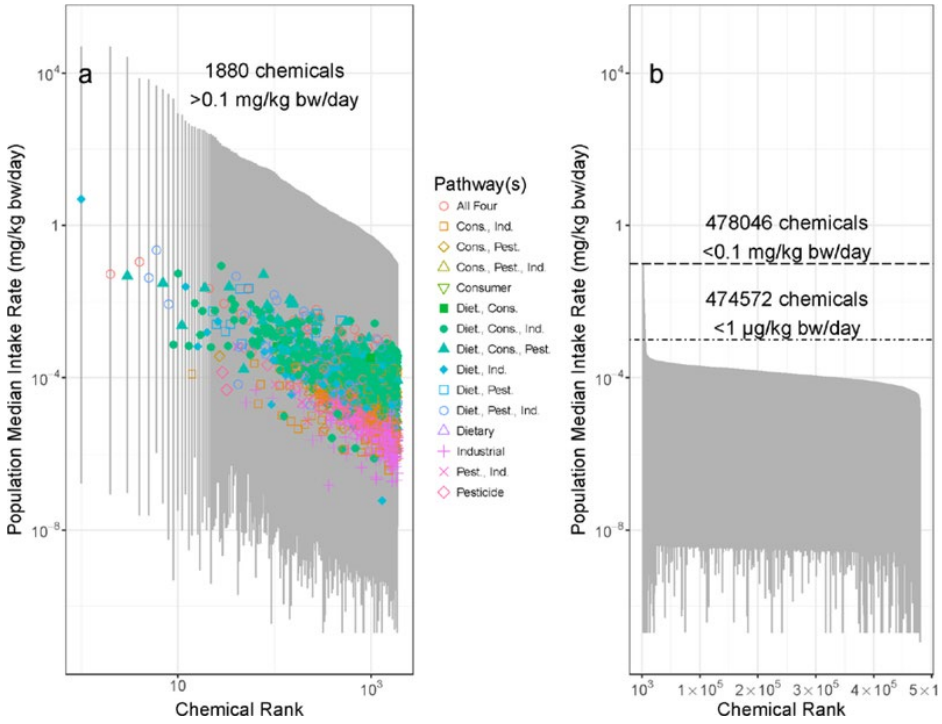
The predictive density for thousands of possible comparisons of the loadings with the MLE representative values for the observed concentrations. The prior probabilities are set to zero. When a model explains the observations for the chemicals, the metamodel is updated to increase (or decrease) the weight of the model.

Fate models (*k*)

MODEL	INPUTS	OUTPUT
E-FAST	Removal from wastewater treatment (%), Estimated flow of the receiving stream	Estimation of Surface Water Exposure Concentrations in Rivers and Streams (µg/L)/(kg/day)
EXAMS metamodel	Estimated flow of the receiving stream, octanol:water, air:water	Annual average dissolved water concentration rate (mg/L)/(kg/hr)
USEtox	Molar mass, octanol:water, air:water, pKa, half-lives in air, water, soil, and sediment, organic carbon:water, vapor pressure, water solubility, bioaccumulation in fish	Bulk or dissolved water (kg/(kg/d))

Metamodel loading*fate weights (m_{jk})

- Once we get *m* from our calibrated metamodel, we can predict values of *y* and their uncertainties
- Predictions possible for as many as 8295 compounds (limited by loading data)
- Predicted concentration ranges can be compared with different risk assessment parameters



Results of previous SEEM model, which was used to prioritize chemicals according to risk from representative median intake and predicted exposure pathway based on near-field contact (Ring, CL, et al. Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways Environ Sci Technol. 2019 Jan 15; 53(2): 719–732. 10.1021/acs.est.8b04056)

Drinking water concentrations

2410	1,1-DICHLOROPROPENE		<	MDL	.005 MG/L		
2420	1,2,3-TRICHLOROBENZENE		<	MDL	.0005 MG/L		
2414	1,2,3-TRICHLOROPROPANE		<	MDL	.0005 MG/L		
2378	1,2,4-TRICHLOROBENZENE		<	MDL	.0005 MG/L		01-01-2007 12-31-2007
2418	1,2,4-TRIMETHYLBENZENE		<	MDL	.0005 MG/L		
2980	1,2-DICHLOROETHANE		<	MDL	.0005 MG/L		01-01-2007 12-31-2007
2983	1,2-DICHLOROPROPANE		<	MDL	.0005 MG/L		01-01-2007 12-31-2007
2424	1,3,5-TRIMETHYLBENZENE		<	MDL	.0005 MG/L		
2412	1,3-DICHLOROPROPANE		<	MDL	.0005 MG/L		
2413	1,3-DICHLOROPROPENE		<	MDL	.0005 MG/L		
2416	2,2-DICHLOROPROPANE		<	MDL	.0005 MG/L		
2990	BENZENE		<	MDL	.0005 MG/L		01-01-2007 12-31-2007
2993	BROMOBENZENE		<	MDL	.0005 MG/L		
2430	BROMOCHLOROMETHANE		<	MDL	.0005 MG/L		
2943	BROMODICHLOROMETHANE					.00167 MG/L	
2942	BROMOFORM		<	MDL	.001 MG/L		
2214	BROMOMETHANE		<	MDL	.0005 MG/L		
2982	CARBON TETRACHLORIDE		<	MDL	.0005 MG/L		01-01-2007 12-31-2007
2989	CHLOROBENZENE		<	MDL	.0005 MG/L		01-01-2007 12-31-2007
2216	CHLOROETHANE		<	MDL	.0005 MG/L		
2941	CHLOROFORM					.0188 MG/L	

- Drinking Water Watch "Other Chemical Results by Analyte" are available for finished water from community drinking water systems across the U.S. that use surface water
- In future work, we'll use these to develop representative drinking water concentrations, and model the percent difference between the two sets of distributions to estimate removal efficiency as a prioritization parameter

Thank you!

For comments, suggestions, or questions please email
sayre.risa@epa.gov