Bayesian metamodel to estimate risk for thousands of chemicals in surface water

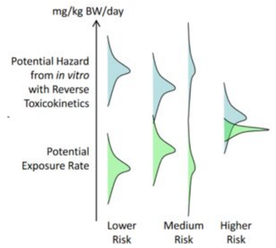
Risa R. Sayre[1,2,3], Jon Arnot[4,5], Kristin Isaacs[1], Peter Fantke[6], Marc Serre[2], John Wambaugh[1]

1: U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure; 2: Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill; 3:Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee; 4: ARC Arnot Research and Consulting; 5: Department of Physical and Environmental Sciences, University of Toronto Scarborough; 6: Department of Management Engineering, Technical University of Denmark

Risa R. Sayre | sayre.risa@epa.gov | ORCiD 0000-0002-6173-8020

1. Introduction

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



Knowledge gap: Inconsistent data availability for concentrations in surface water to develop exposure estimates.

Proposed solution: Development of an open, reproducible workflow to:

1. Determine representative surface water concentrations for hundreds of organic chemicals in the United States based on already available monitoring data
2. Calibrate a metamodel to predict representative surface water concentrations for thousands of non-monitored organic chemicals
3. Prioritize organic chemicals based on the relationship between concentration ranges and predicted no-effect concentrations for freshwater standard test species

2. Method overview

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

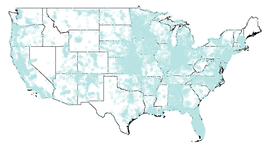
Where *i* is a given chemical, *y* is the representative concentration, *j* is a source of environmental loading data, *l* is the value of that loading given the source and chemical (amount/time), *k* is an exposure model, *p* is the value of a fate prediction model (amount/amount/time), and *m* is a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

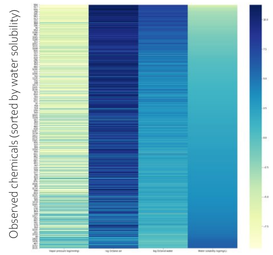
Each coefficient m_{jk} represents a range of weights describing how well model *k* explains observed concentration *y* given the information from loading data source *j* in 20,000 different attempted metamodels.

PAPER ID: 3432976 ACS ENVR

3. Surface water concentrations *y*

The Water Quality Portal provides concentrations of organic chemicals in surface water sampled from 2008 to 2018 covering broad spatial and physicochemical property ranges.





Because over 80% of samples were below varying quantification limits, representative ranges of dissolved (196 chemicals) and bulk (252 chemicals) concentrations were developed using Maximum Likelihood Estimation (MLE).

Upper right: Sampling sites of observation set represent 2114 of 2270 hydrologic subbasins. Lower left: Chemical property space (log10, calculated using OPERA 2.4) of observation set: vapor pressure (mmHg), octanol:air, octanol:water, water solubility (mg/L).

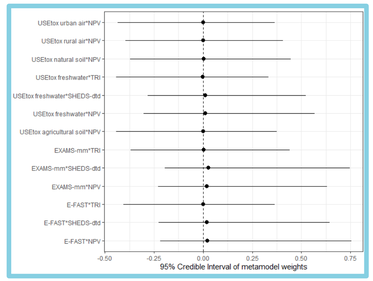
4. Environmental loading data sources *j*

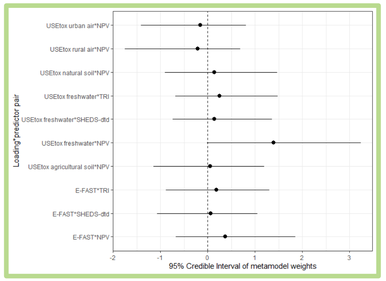
- NPV: Chemical Data Reporting under the Toxic Substances Control Act (reported to the EPA about mass imported into or produced in the U.S. by year) + Pesticides Industry Sales and Usage, 2008 – 2012 Market Estimates (2017 U.S. EPA OPP report)
- SHEDS (Stochastic Human Exposure and Dose Simulation – Down the Drain): model simulating the amount of a chemical that goes down the drain based on household usage, reduced by the percentage removed by wastewater treatment
- TRI (Toxic Release Inventory): data reported to EPA about industrial releases

5. Fate prediction 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, fish bioaccumulation	kg bulk or dissolved chemical in freshwater at steady-state per daily kg emitted (kg/(kg/d))

6. Metamodel loading*fate coefficients (m_{jk})

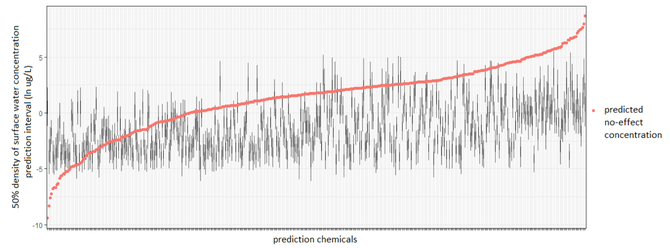




95% Credible intervals of metamodel coefficients [left: dissolved, right: bulk]

No tested loading*fate pair was predictive of observed dissolved concentrations. The most informative pair for bulk concentrations was USEtox freshwater model using loadings from NPV.

7. Prioritization based on ecotoxicity estimates



8. References

Arnot, Jon A., *et al.* Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. Environmental health perspectives 120.11 (2012): 1565-1570. || Barber, M. Craig, *et al.* Developing and applying metamodels of high resolution process-based simulations for high throughput exposure assessment of organic chemicals in riverine ecosystems. Sci Total Environ. 2017 Dec 15;605-606:471-481. || Henderson, A.D., *et al.* 2011. USEtox fate and ecotoxicity factors for comparative assessment of toxic emissions in life cycle analysis: sensitivity to key chemical properties. The International Journal of Life Cycle Assessment 16, 701-709. || Isaacs, Kristin K., *et al.* "SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources." Environmental science & technology 48.21 (2014): 12750-12759. || Kavlock, Robert J., *et al.* "Accelerating the pace of chemical risk assessment." Chemical research in toxicology 31.5 (2018): 287-290. || Mansouri, K., *et al.* Open-source QSAR models for pKa prediction using multiple machine learning approaches. *J Cheminform* 11, 60 (2019). || Martin, T.M., P. Harten, R. Venkatapathy, S. Das and D.M. Young. (2008). "A Hierarchical Clustering Methodology for the Estimation of Toxicity." *Toxicology Mechanisms and Methods*, 18, 2: 251–266. || Read, E. K., *et al.* (2017). Water quality data for national-scale aquatic research: The Water Quality Portal. *Water Resources Research*, *53*(2), 1735–1745. || U.S. EPA. "Access CDR Data". Chemical Data Reporting Under the Toxic Substances Control Act. https://www.epa.gov/chemical-data-reporting/access-cdr-data || U.S. EPA. E-FAST - Exposure and Fate Assessment Screening Tool Version 2014. https://www.epa.gov/tsca-screening-tools/e-fast-exposure-and-fate-assessment-screening-tool-version-2014 || Wambaugh, John F., et al. "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals." Environmental science & technology (2014).