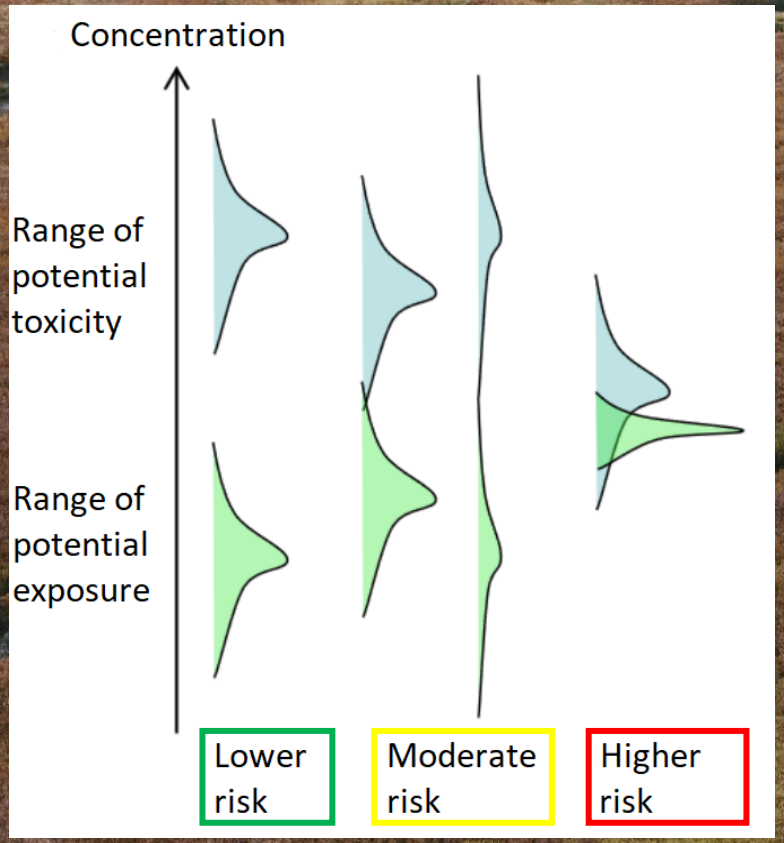
**Risa Sayre**[1,2,3], *R. Woodrow Setzer*[1*], *Marc Serre*[2], *John F. Wambaugh*[1,2]

# Determining Representative National Surface Water Chemical Concentration Ranges for Risk Prioritization

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

This presentation does not necessarily reflect U.S. EPA policy.



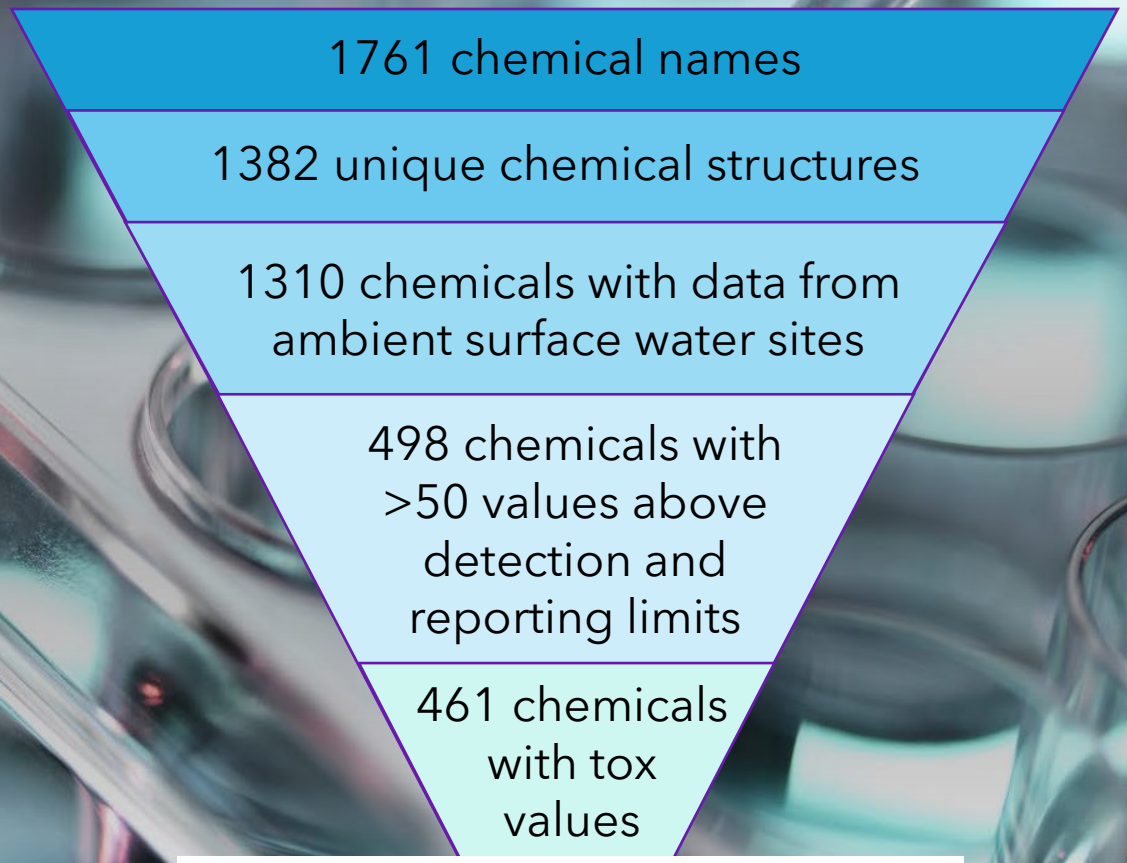

Even a highly uncertain estimate can provide enough information to rank chemicals for risk prioritization.

# Nation-scale values for risk prioritization

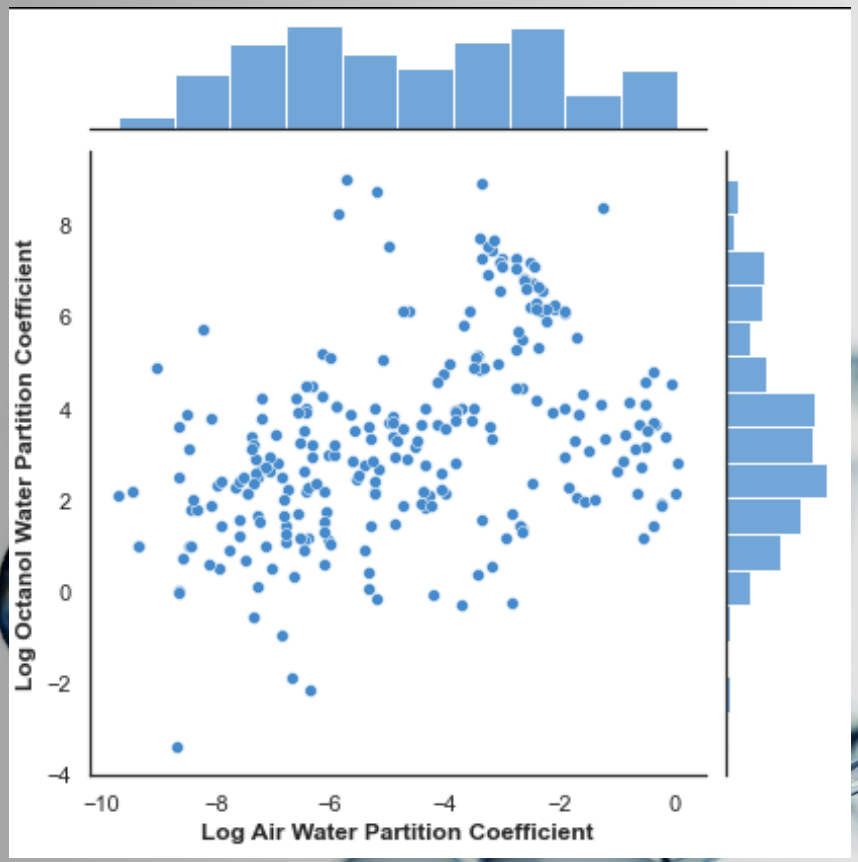
National average surface water concentrations of anthropogenic chemicals can be used to estimate one type of potential **exposure**: a component needed to identify the most important chemicals for a thorough risk evaluation.

# Data availability

I designed a Python workflow to pull, filter, and standardize National Water Quality Portal records from 2008 to 2018 for all organic chemicals from sites or activities that would yield ambient concentrations from across the U.S. (2114 of 2270 hydrologic subbasins)




The number of chemicals for which estimates could be made decreased with each data cleaning step.

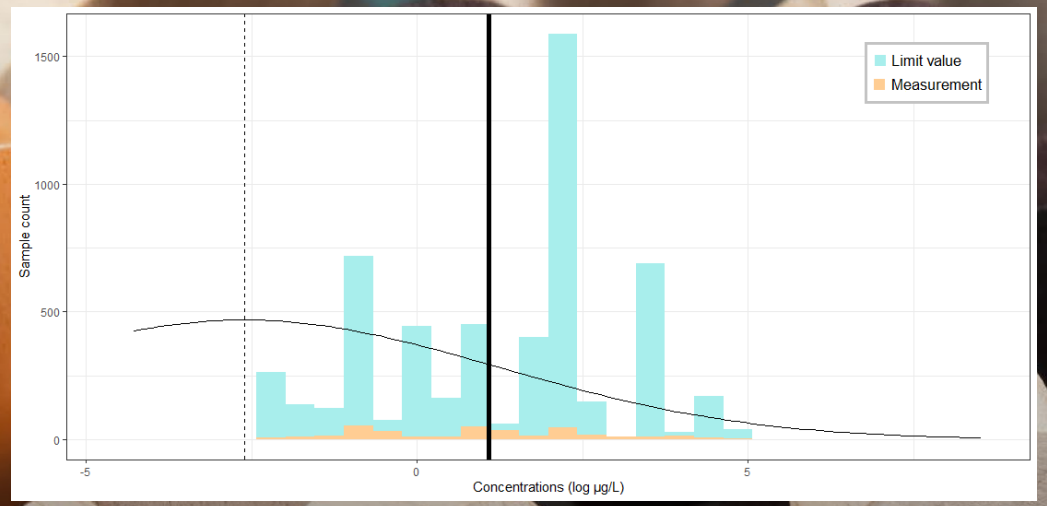


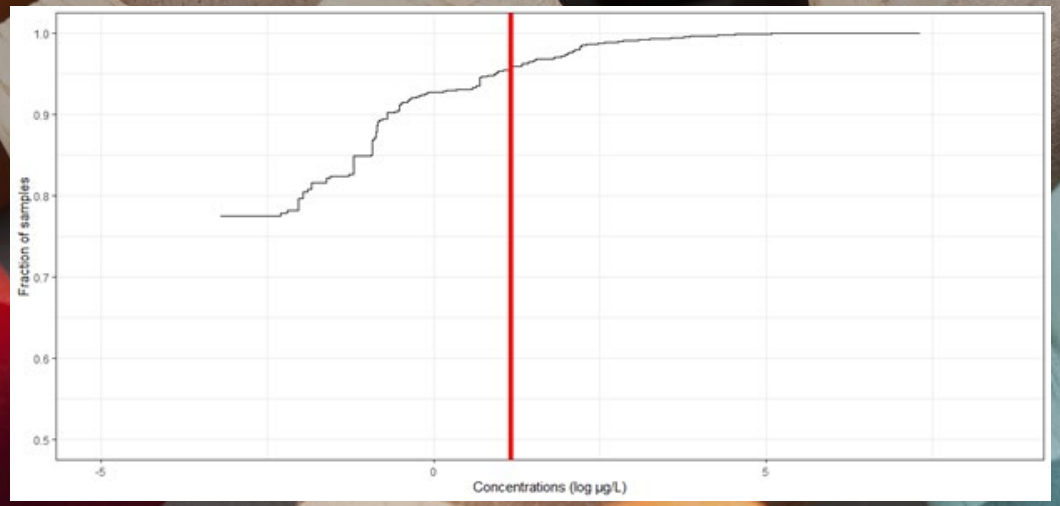

There is higher representation of high log P (defined as >5) chemicals in this set than for chemicals in general.

# Chemical space

The range of air:water and octanol:water (log P) partition coefficients in this set is similar to the range in the EPA's DSSTox, a database of almost 800,000 chemicals.

50% are pesticides, 12% are PCBs, 14% are pharmaceuticals, and the remainder are chemicals form industrial or biological wastewater




Di (2-ethylhexyl) phthalate has high estimation uncertainty due to a large proportion of nondetects, high limit values and a few extreme observations

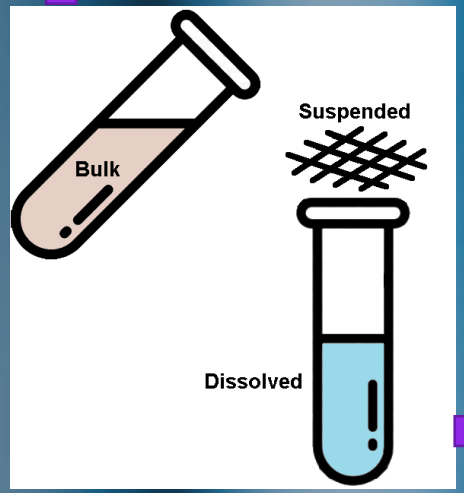
# Representative values

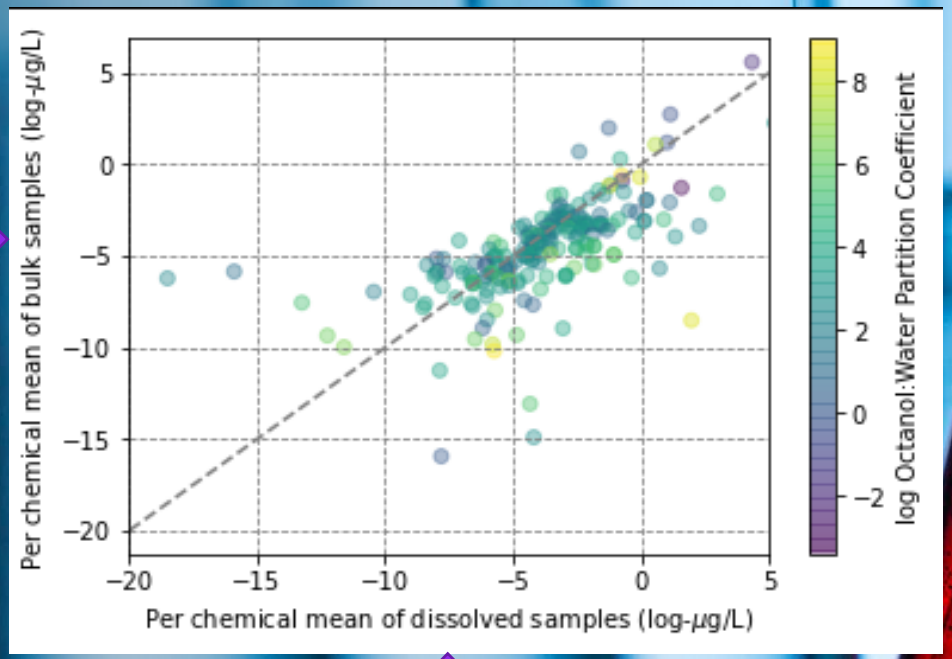
Records for most chemicals were over 80% censored (below a detection or reporting limit).

We needed a method that could incorporate knowledge from both discrete values and left-censoring limits.

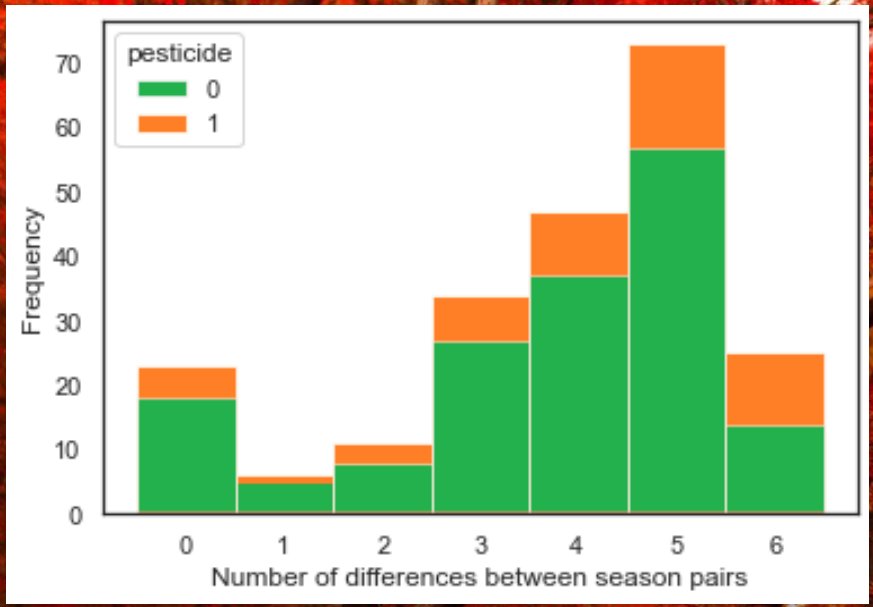
We compare the *mean* of the *lognormal* maximum likelihood estimate (MLE; black, upper plot) and the 95$^{th}$ percentile of the empirical cumulative distribution function (KM; red, lower plot)

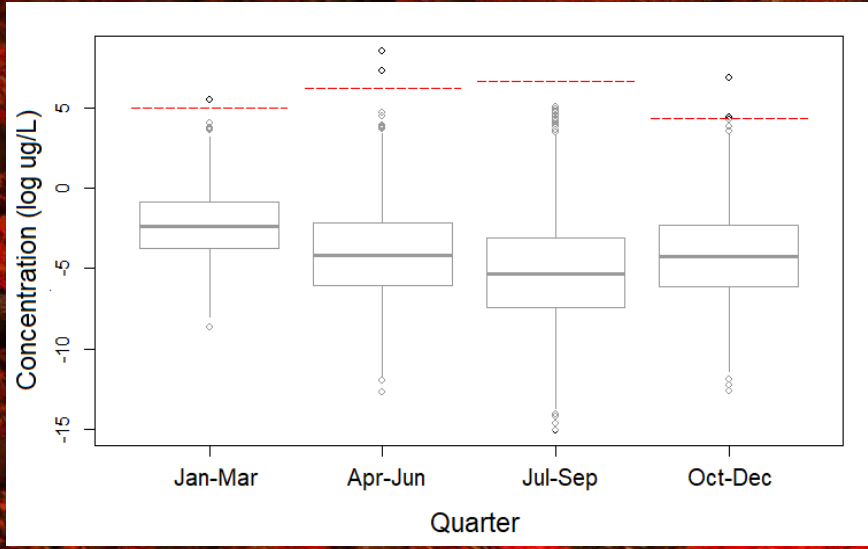
# Different sample types had similar means but different distributions




For most chemicals, the MLE mean bulk concentration is similar to the MLE mean dissolved concentration, regardless of the chemical's affinity for organic matter (log P).



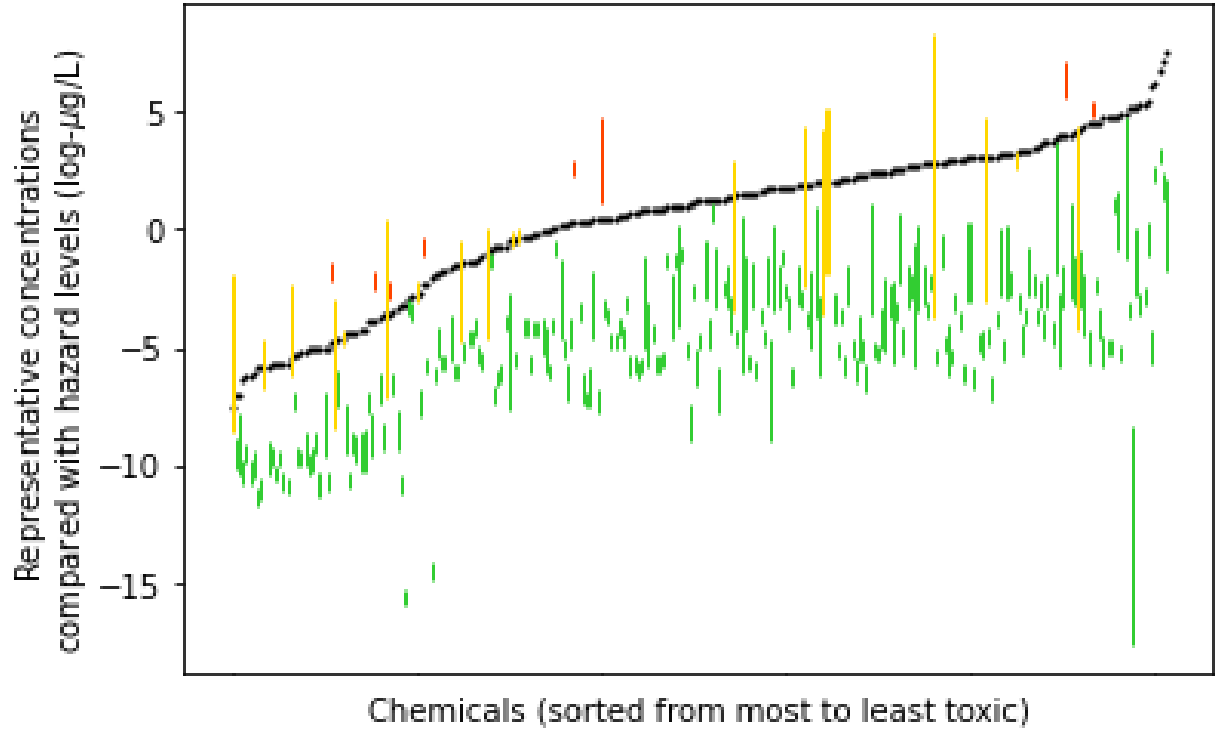

# Seasonal concentrations

For chemicals with samples taken in all quarters, a log-rank test on all season pairs showed differences between empirical distributions for some chemicals, but no season-pair trend across chemicals, even for pesticides.

Example (at left): Di (2-ethylhexyl) phthalate has statistically similar concentrations in Apr-Jun and Oct-Dec, but not in the five other season-pairs. Dashed red lines indicate the highest limit value.

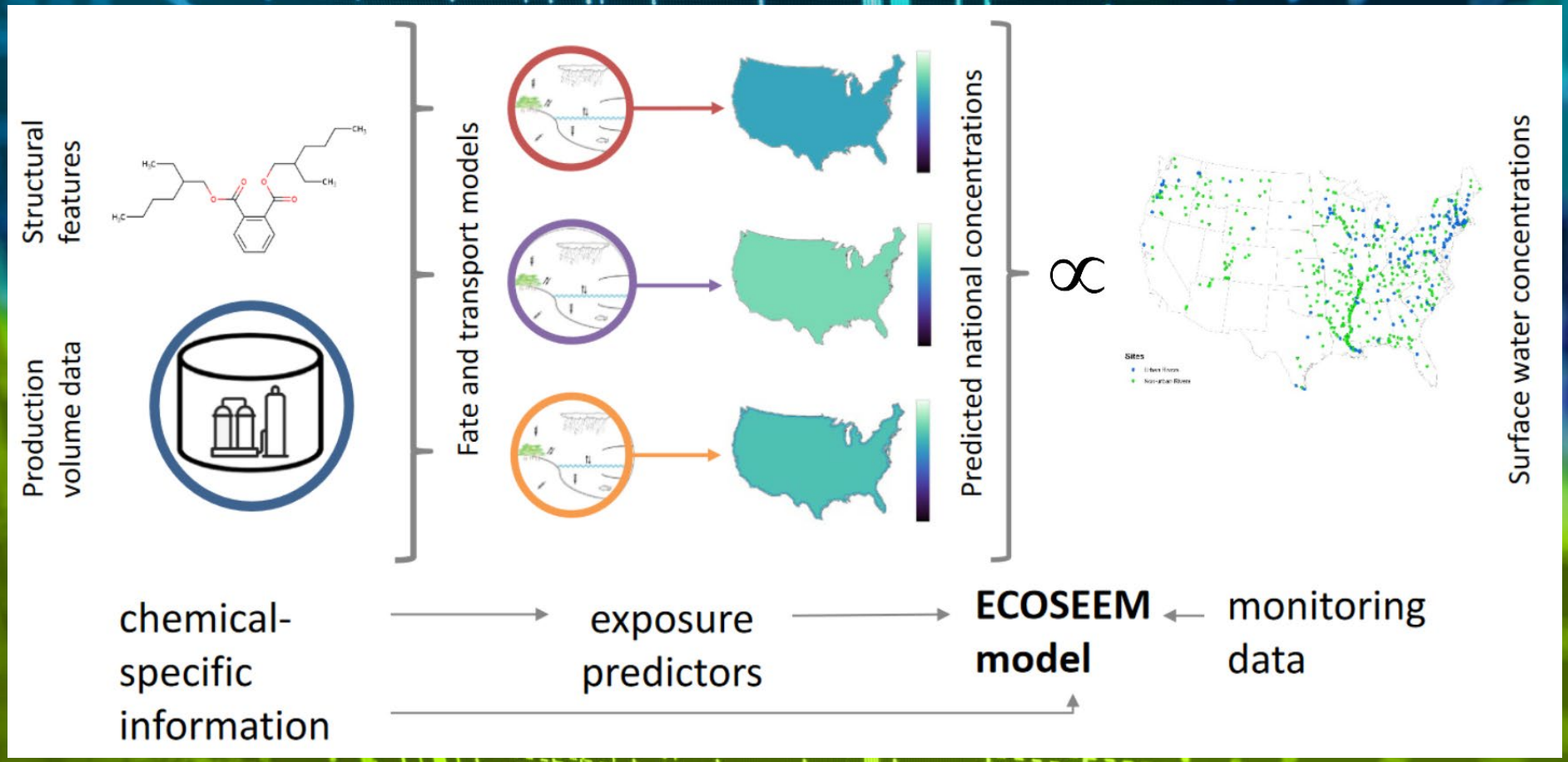
# Representative ranges can be compared to hazard concentrations in *Daphnia magna*

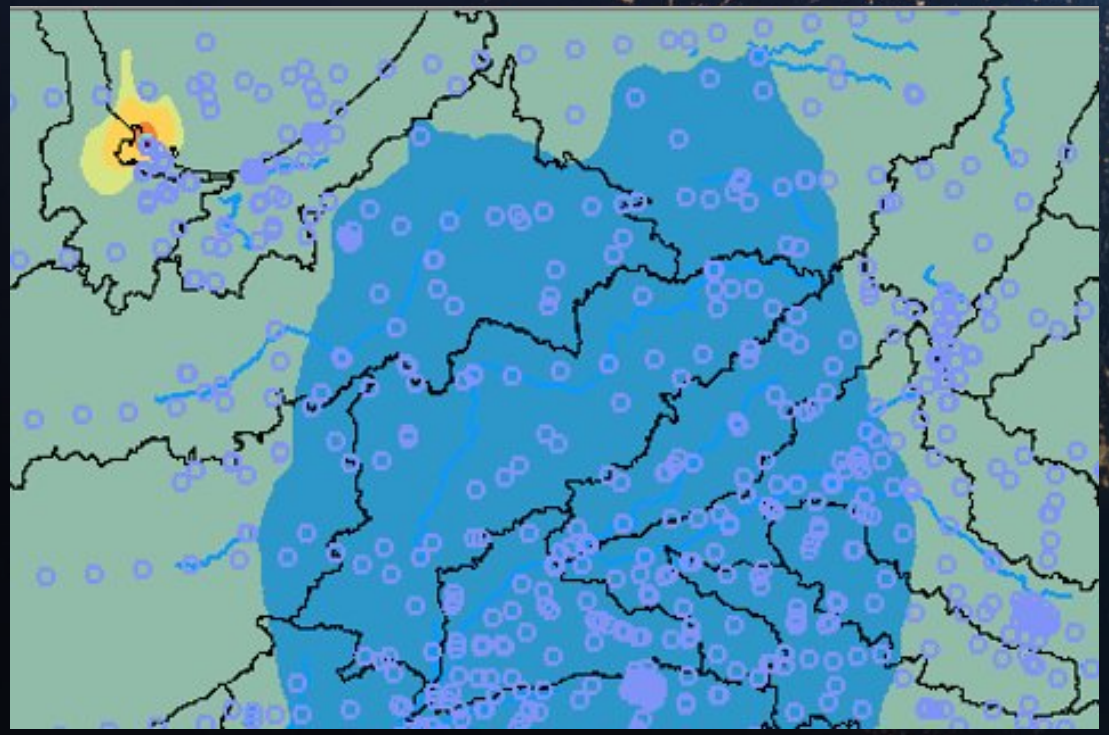
The uncertainty in the MLE mean + standard deviation is still less than the range of toxicities and can help prioritize risk to *Daphnia*, a sensitive organism.




Chemicals where the estimated average exceeds the toxic level are the highest priority (red). The most toxic chemicals are not always the highest priority, due to low exposures.

# Representative ranges calibrate Bayesian metamodel





Estimate of Di (2-ethylhexyl) phthalate concentrations in northern Indiana (red: high to blue: low). Black lines are watershed boundaries, and circle colors indicate estimation uncertainty (all high at this stage, unfortunately).

# Future work: Improved estimates for high-priority chemicals

Evaluate spatiotemporal autocorrelation using Bayesian Maximum Entropy to develop a time-varying surface of concentration estimates including other variables like land use and hydrology

# Thank you!

sayre.risa@epa.gov