Case Study: Using spatial data to predict PFAS in fish tissue for sampling prioritization in the Columbia River Basin

Nicole “Nikki” M. DeLuca

Postdoctoral Physical Scientist
U.S. EPA, Office of Research and Development
Center for Public Health and Environmental Assessment

Disclaimer

The views expressed in this presentation are those of the author and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Core Partner Group

- Columbia River Inter-Tribal Fish Commission
- Confederated Salish and Kootenai Tribes
- Confederated Tribes of the Grand Ronde
- Confederated Tribes of the Umatilla Indian Reservation
- Yakama Nation
- Confederated Tribes of Warm Springs
- Kalispel Tribe of Indians
- Nez Perce Tribe
- Columbia River Basin Restoration Working Group
- Oregon Health Authority
- Oregon Department of Environmental Quality
- Washington Department of Ecology
- Washington Department of Health
- Montana Department of Environmental Quality
- Missoula Valley Water Quality District
- Idaho Department of Environmental Quality
- US Geological Survey
- USEPA Regions 8, 9, 10
- USEPA Office of Research and Developement
- USEPA Office of Enforcement and Compliance Assurance

EPA Office of Research & Development

- Elaine Cohen Hubal
- Ashley Mullikin
- Ana Rappold

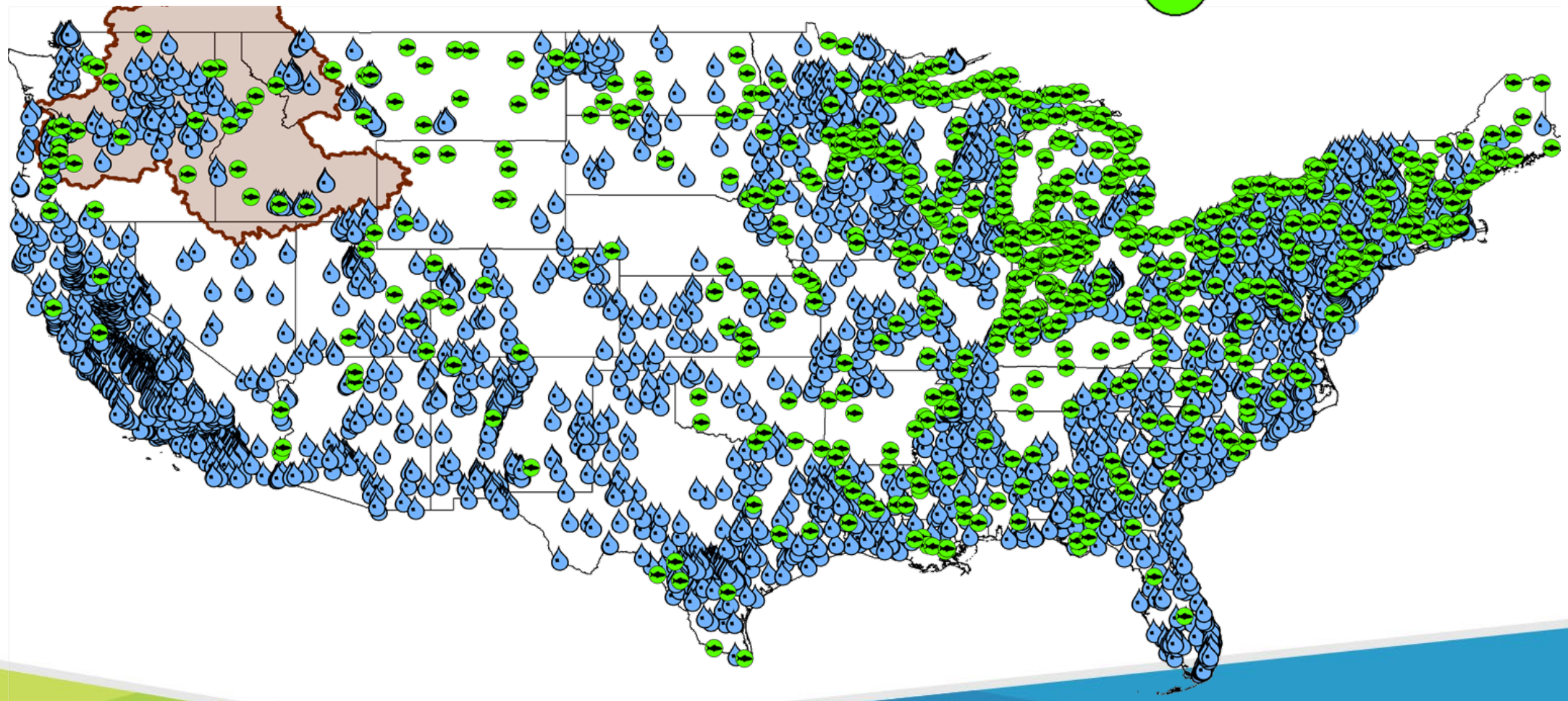
EPA Region 8

- Peter Brumm

EPA Region 10

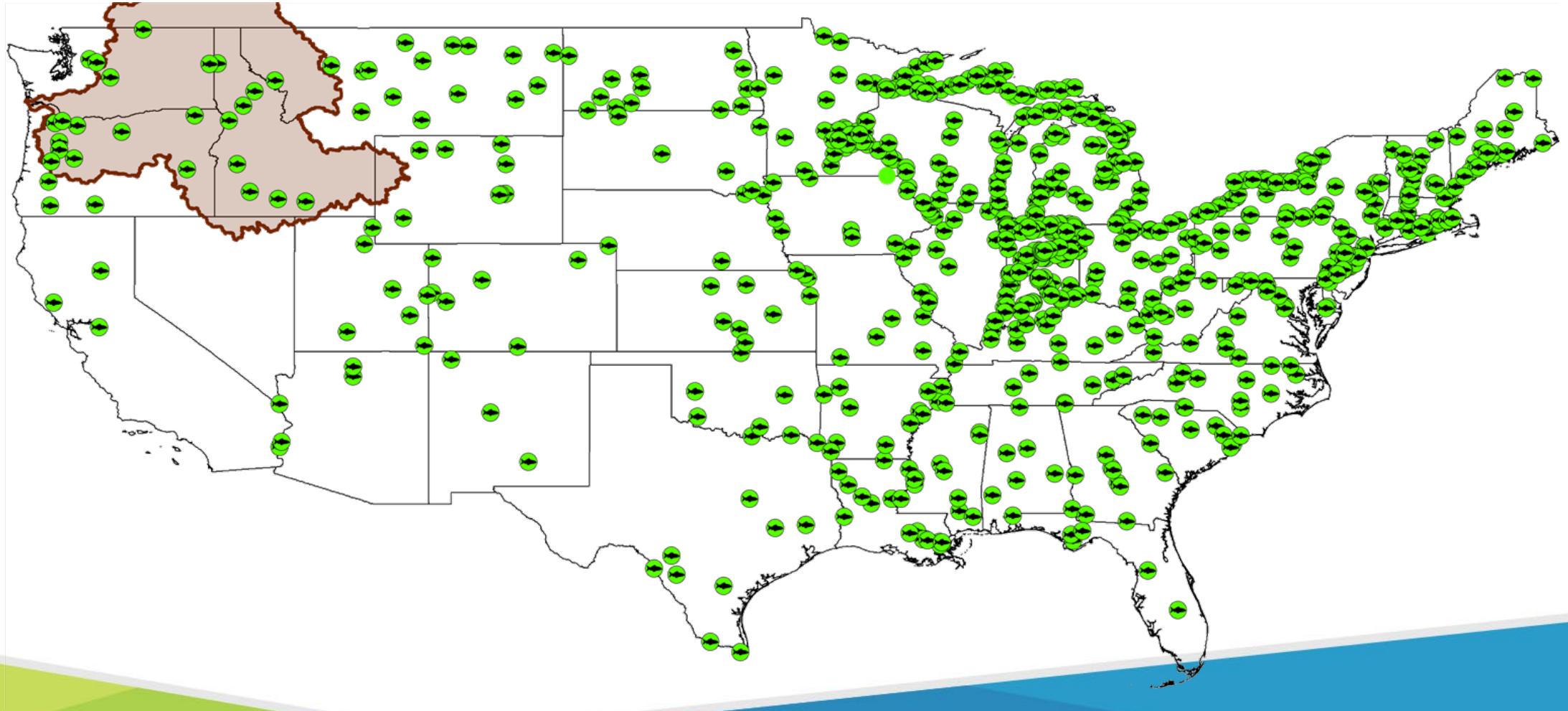
- Ashley Zanolli

Existing PFAS Occurrence Data from PFAS Analytic Tools (as of April 2023)



Existing PFAS Occurrence Data from PFAS Analytic Tools (as of April 2023)

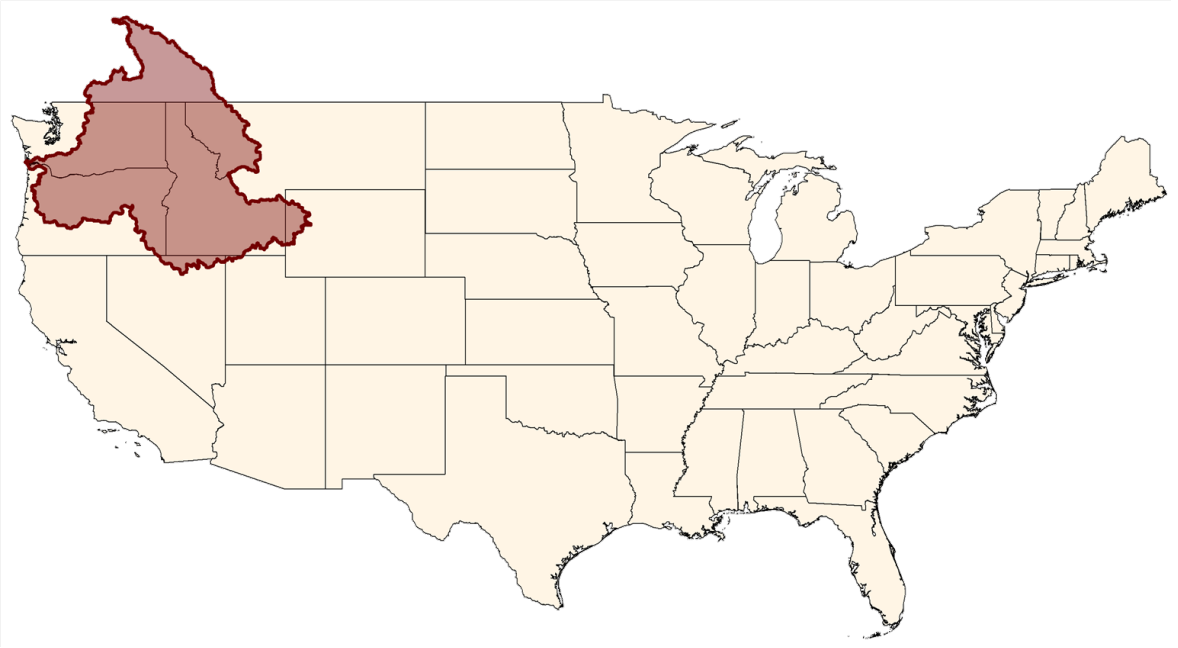
Fish Tissue



Columbia River Basin

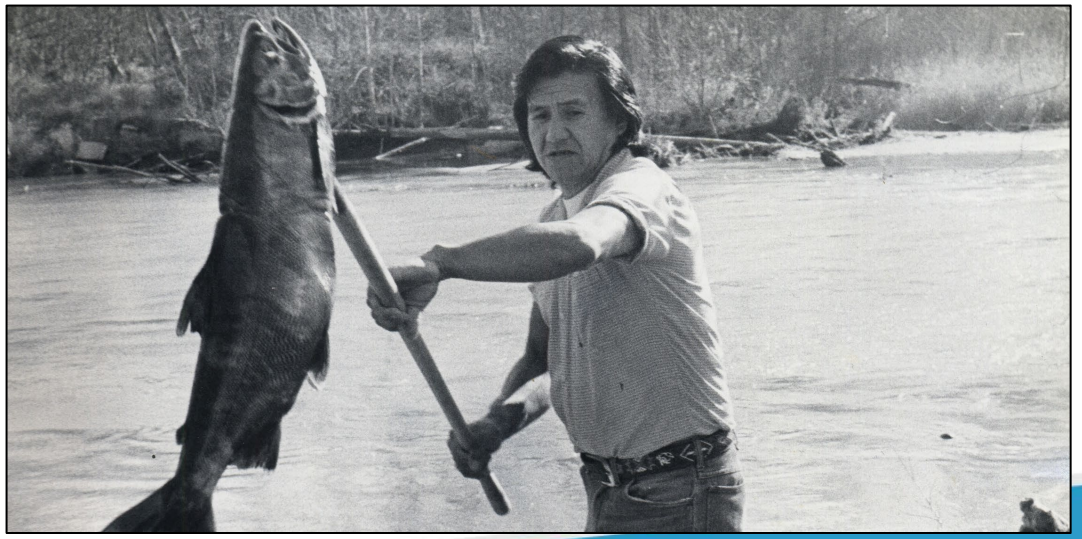
One of the largest watersheds in the U.S. (~666,700 sq. km)

- 16 federally recognized tribal reservations
- Washington, Oregon, Idaho, Montana, British Columbia
  - Wyoming, Nevada, and Utah



Environmental Justice

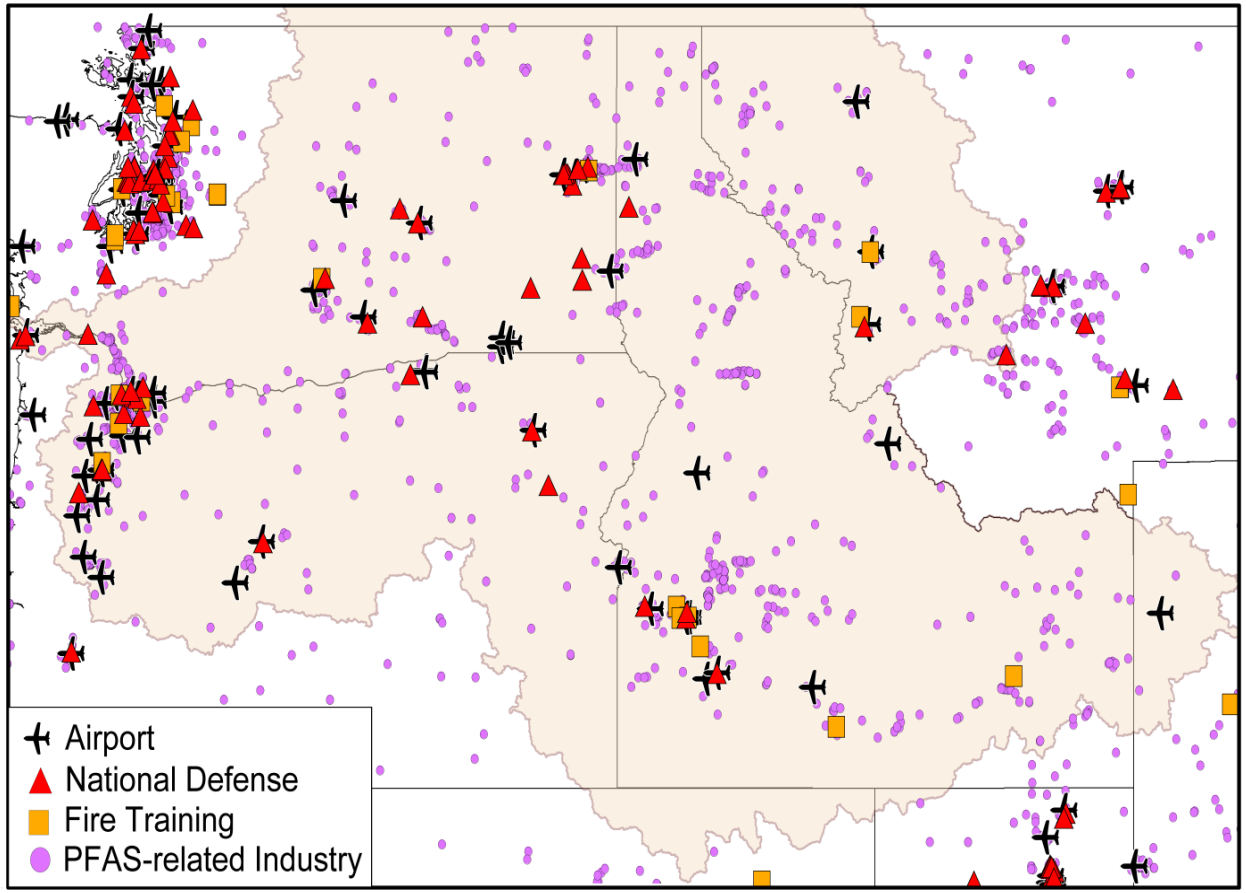
- Contaminated fish pose a disproportionately high exposure risk for high fish-consuming populations
- Tribal people consume **9 to 12 times more fish** than U.S. general population (Columbia River Inter-Tribal Fish Commission and USEPA, 1994)



Northwest Indian Fisheries Commission

Motivation

- Jurisdictions around the world are working to efficiently identify and characterize the extent of PFAS contamination and human exposure.
- Tribes and States in EPA Regions 8 and 10 sought an efficient and cost-efficient way to prioritize site investigation and screen for PFAS contamination.
- Challenges include:
  - Many potential sources
  - Unknown facilities' PFAS use
  - Uncertain fate and transport
  - Limited PFAS measurements



DeLuca et al., 2023

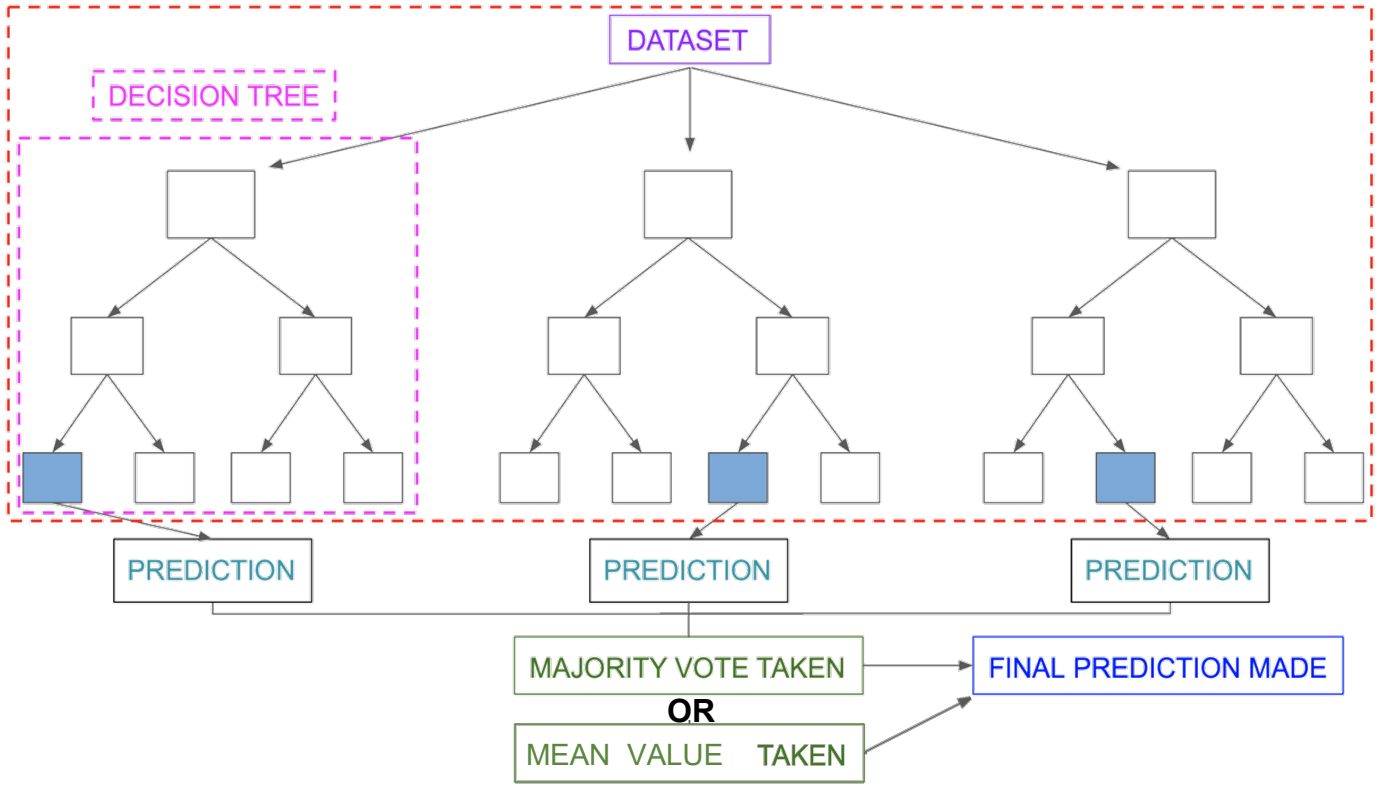
Methodology

- Use widely available geospatial data and limited existing PFAS measurements in fish tissue to develop a model that can predict PFAS concentrations and identify potential hotspots in Washington and Oregon states

- Random forest modeling
  - Regression – predict concentration values
  - Classification – predict above/below a threshold concentration

Random Forest

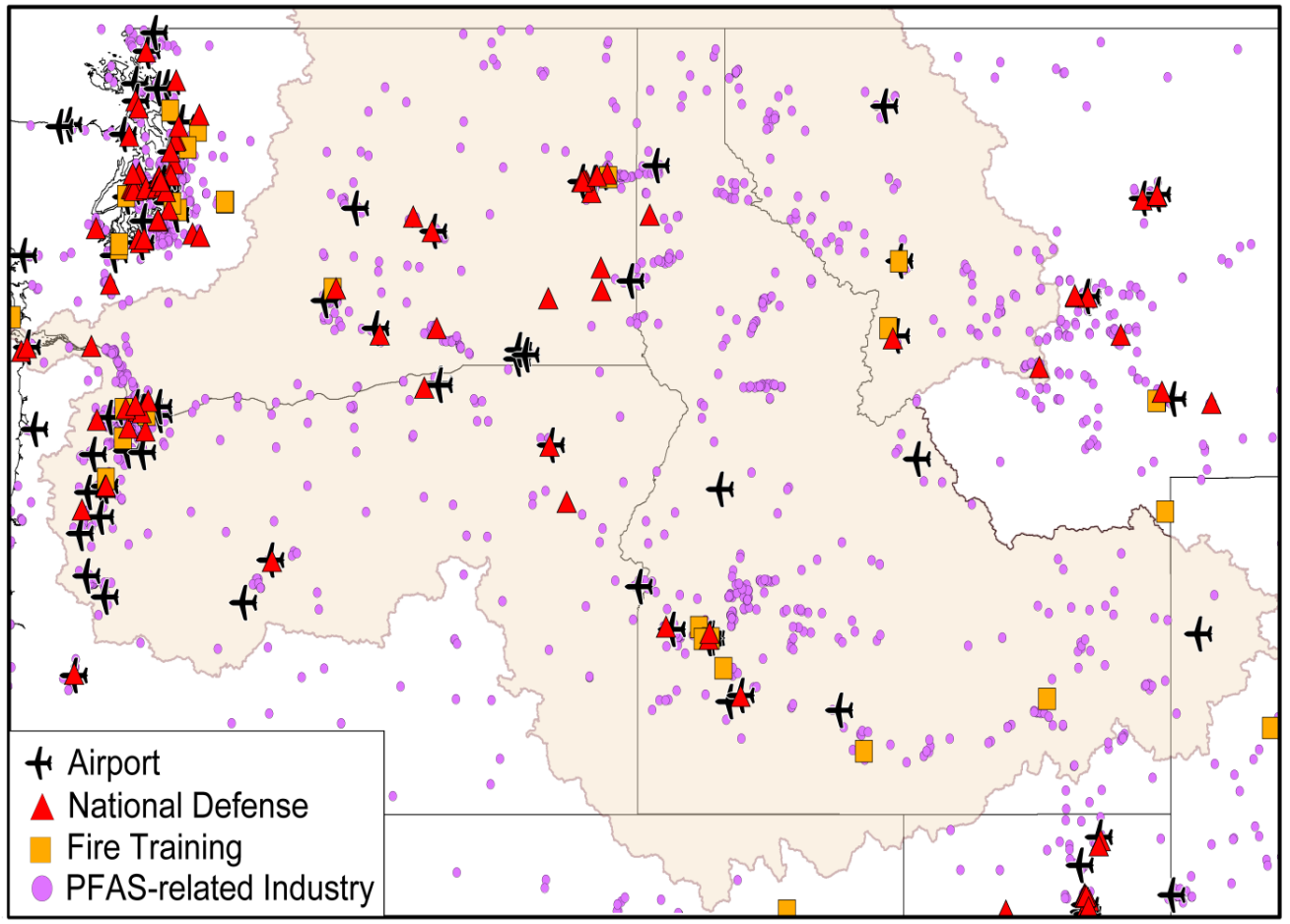
Source: section.io/engineering-education/introduction-to-random-forest-in-machine-learning/



Potential Sources from PFAS Analytic Tools (as of April 2023)

PFAS-related Industries (EPA ECHO)

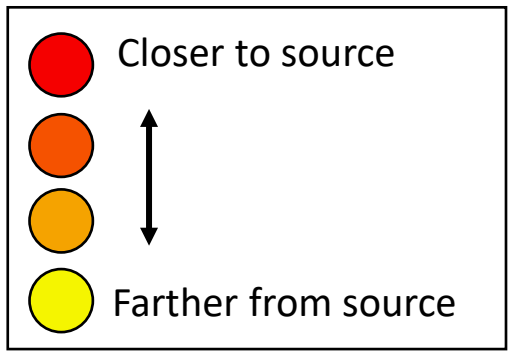
- Fire Training
- Airports
- National Defense
- Mining and Refining
- Landfills
- Metal Coating
- Metal Machinery Manufacturing
- Industrial Gas
- Glass Products
- Furniture and Carpeting
- Electronics
- Consumer Products
- Cleaning Product Manufacturing
- Chemical Manufacturing
- Cement Manufacturing
- Petroleum
- Industrial Gas
- Paints and Coatings
- Oil and Gas
- Plastics and Resins
- Printing
- Paper Mills
- Textiles

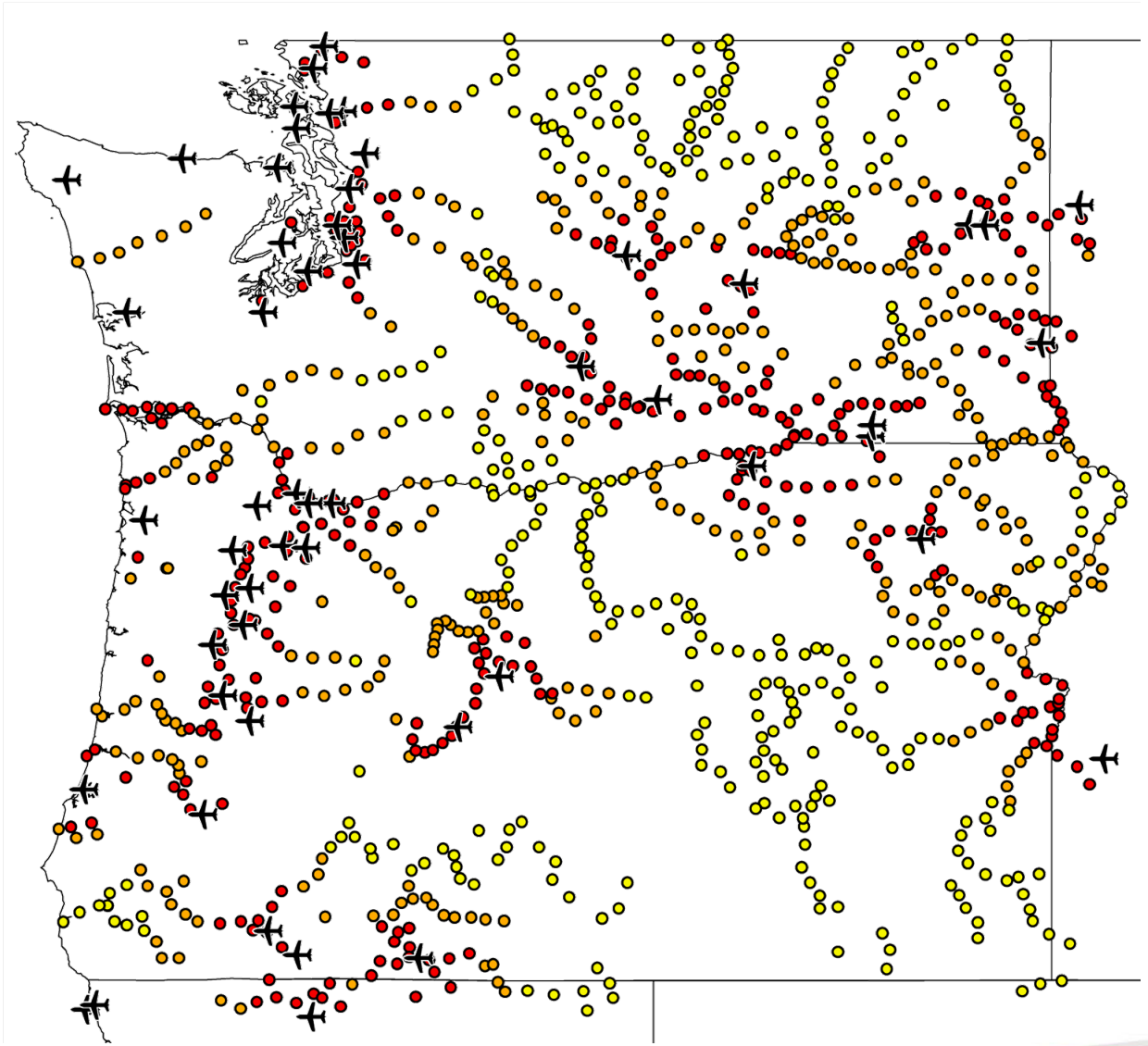


DeLuca et al., 2023

Quantifying Spatial Data

- For each industry, calculated distances from points along waterbodies to nearest potential source
- Example – Distance to nearest airport

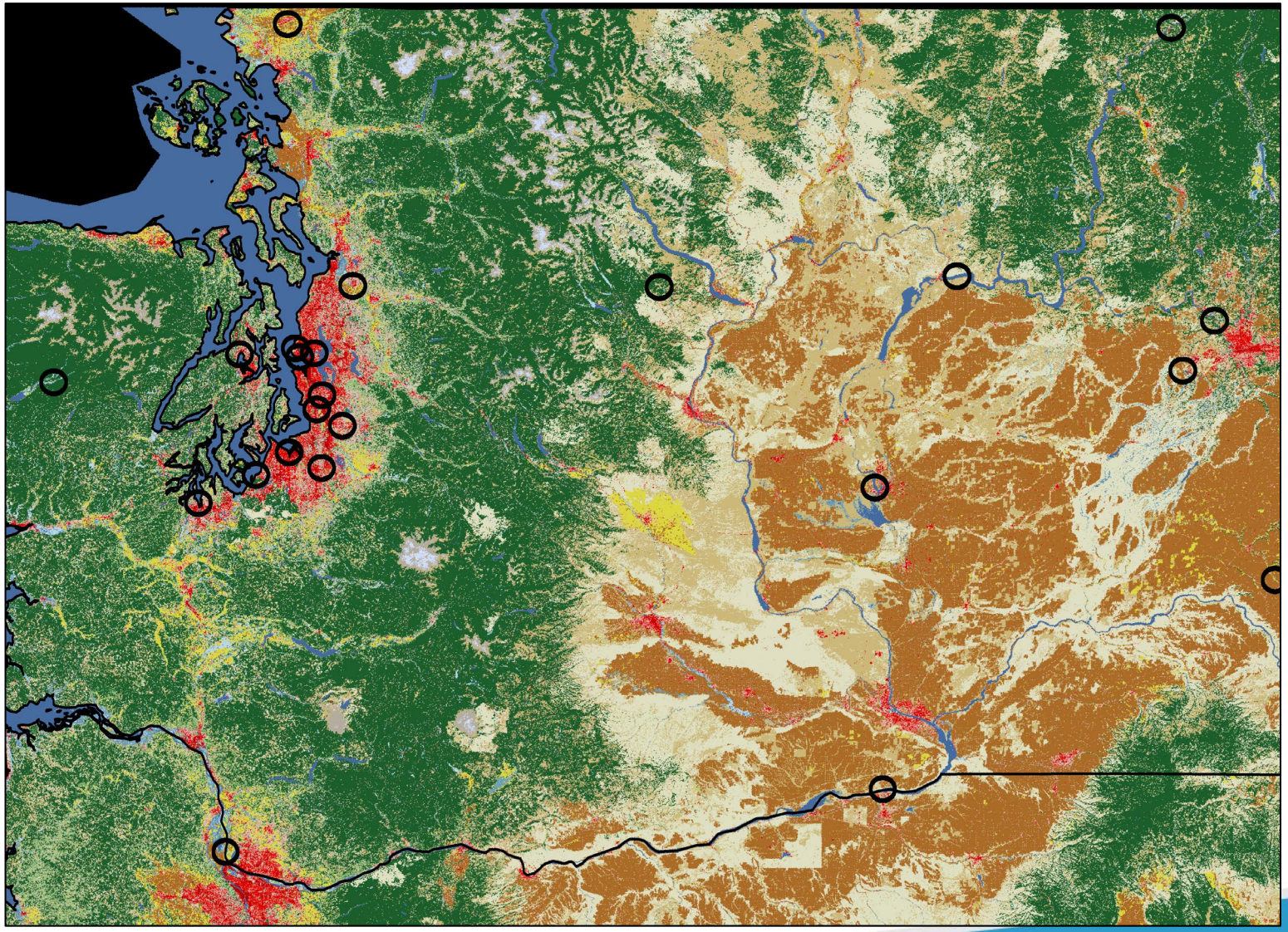




DeLuca et al., 2023

Land Cover

- USGS National Land Cover Database (NLCD)
- Calculate within 5 km buffer
  - % Developed Land
  - % Agricultural Land
  - % Natural Land



5 km buffer around waterbody points

DeLuca et al., 2023

Existing Fish Tissue PFAS Measurements

Data Sources

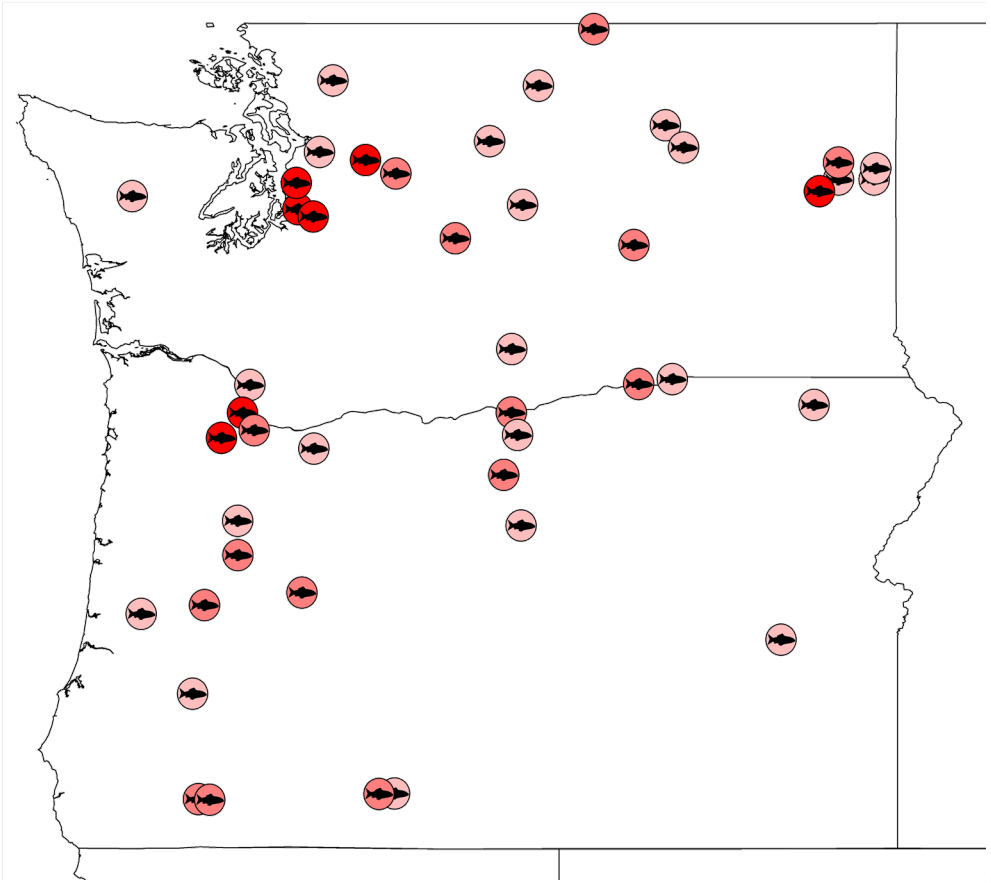
- EPA's PFAS Analytic Tools
  - Water Quality Portal
  - EPA National Rivers and Streams Assessment
- Washington Department of Ecology

Data Summary

- Fillet samples (n=45)
- 2008 - 2019
- Fish Species:
  - Brook trout, sea trout
  - Brown bullhead
  - Channel catfish
  - Common carp
  - Cuthroat trout
  - Largemouth bass
  - Largescale sucker
  - Mountain whitefish
  - Northern pikeminnow
  - Peamouth
  - Pumpkinseed
  - Rainbow trout, redband trout, steelhead
  - Smallmouth bass
  - Tench
  - Tyee sucker
  - Walleye
  - Yellow perch

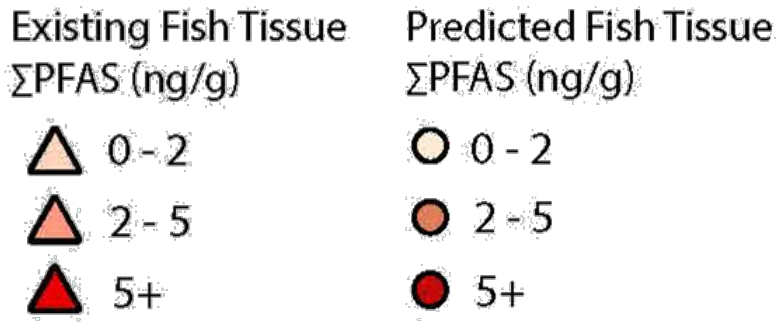
∑PFAS Chemicals

- PFOS
- PFOA
- PFNA
- PFDA
- PFDoA
- PFHxS
- PFHxA
- PFHpA
- PFUnA
- PFBS



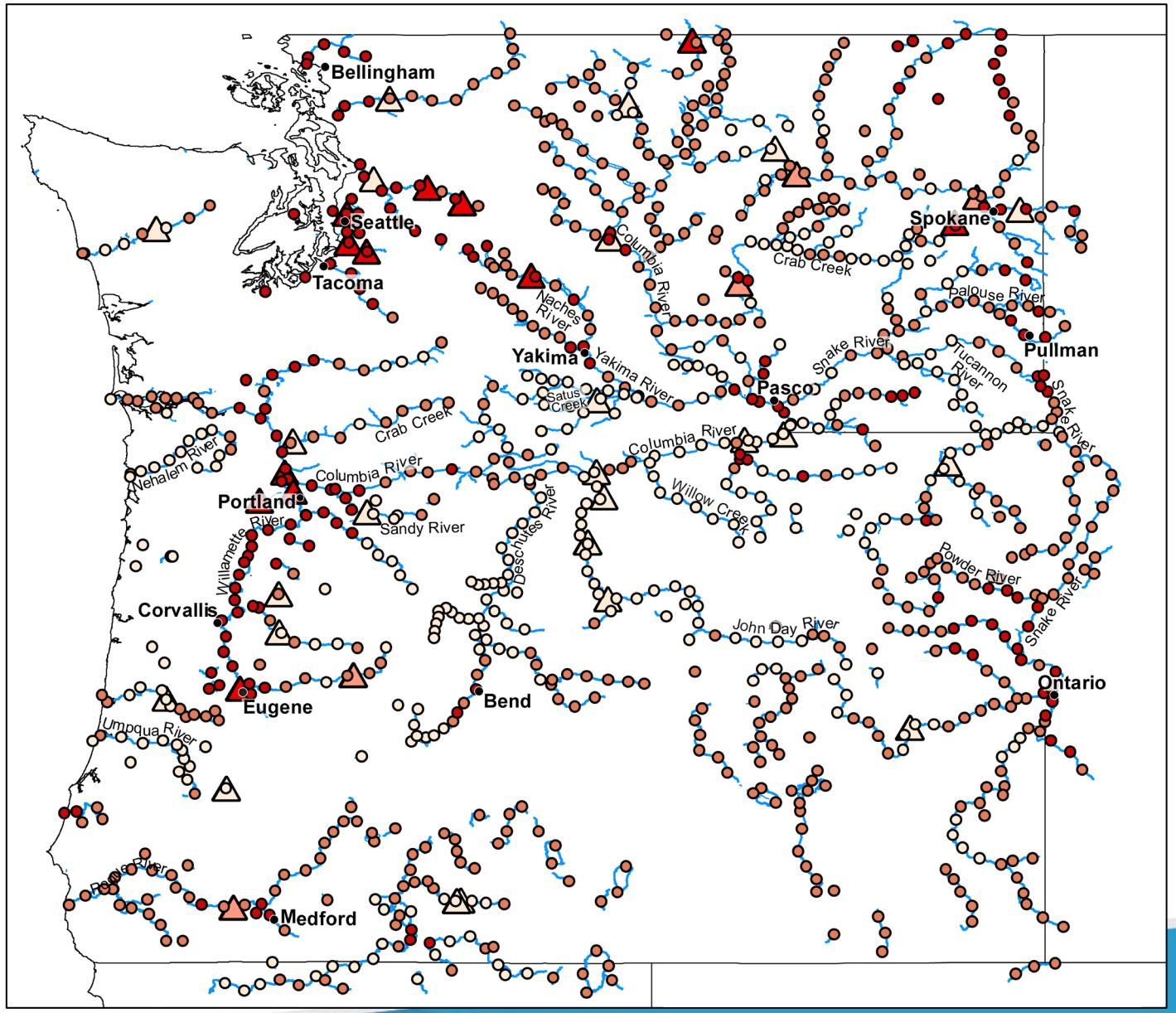
DeLuca et al., 2023

Results - Regression



- Monte Carlo model evaluation
  - 100 random splits of dataset into training (80%) and test (20%) data

    Mean MAE: 7.26 ng/g
    Mean RMSE: 164.65 ng/g
    Mean Bias: 1.14 ng/g



Results - Classification

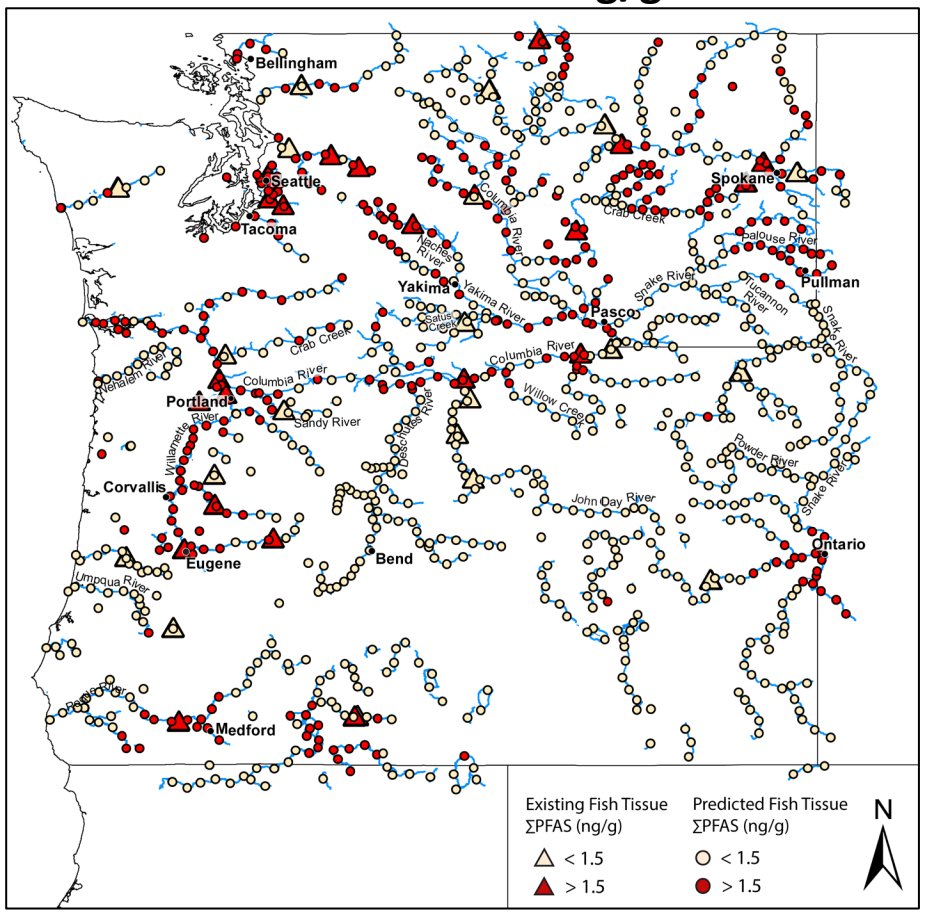
- Monte Carlo model evaluation
  - 3 ng/g threshold

    Mean AUC: 0.72
    Mean Accuracy: 80.2%
    Mean Sensitivity: 74.6%
    Mean Specificity: 84.4%

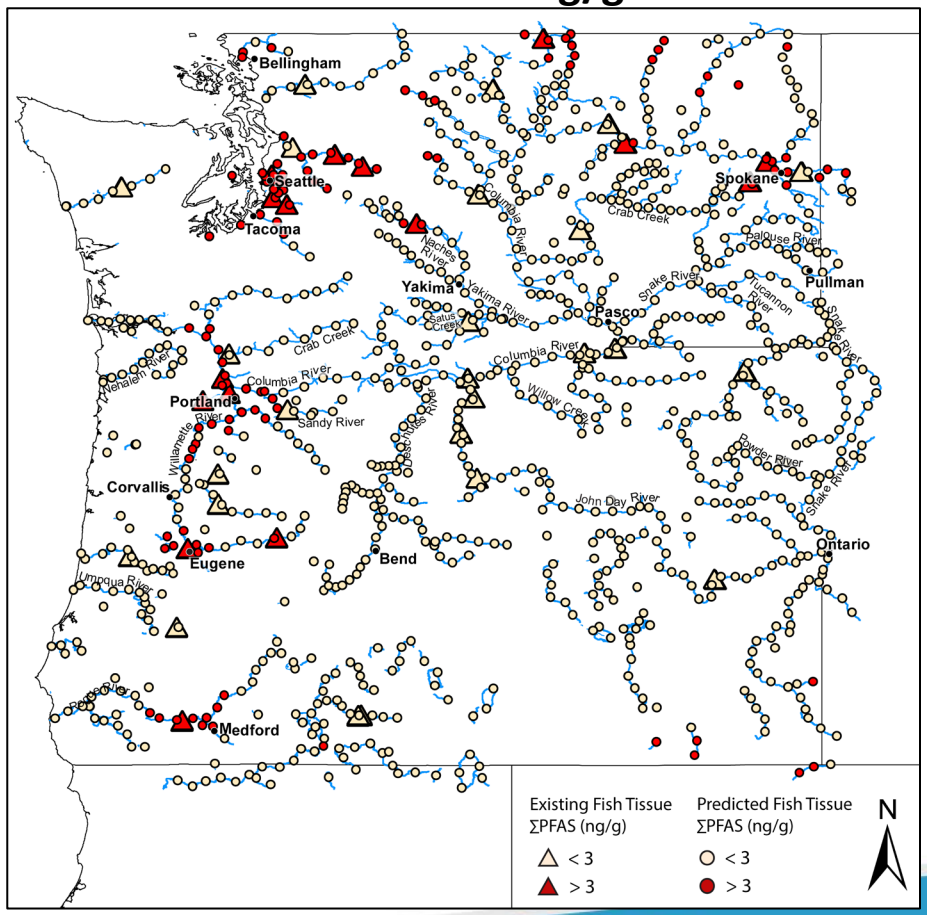
  - 1.5 ng/g threshold

    Mean AUC: 0.63
    Mean Accuracy: 71.0%
    Mean Sensitivity: 65.1%
    Mean Specificity: 79.4%

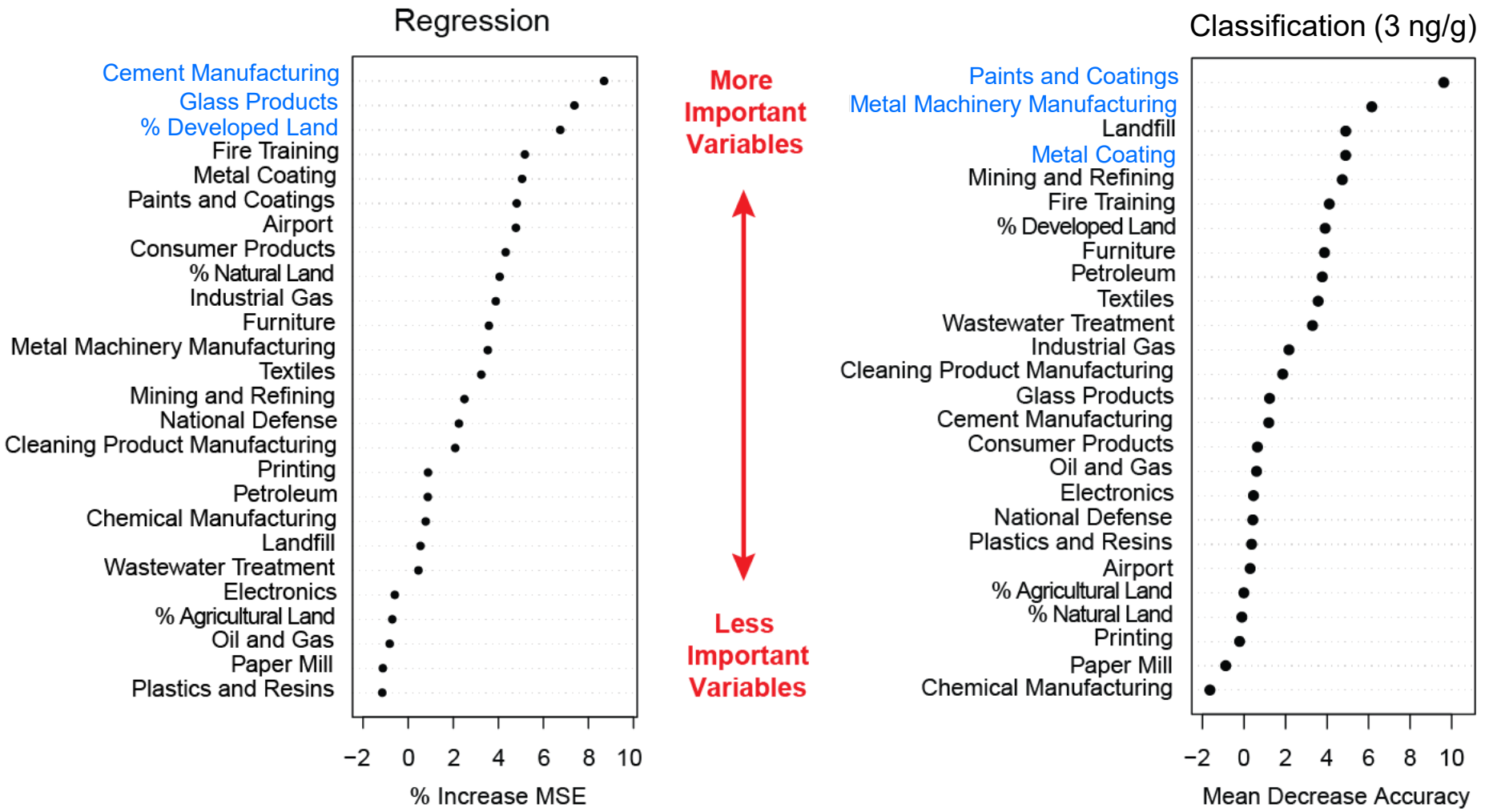
Threshold = 1.5 ng/g



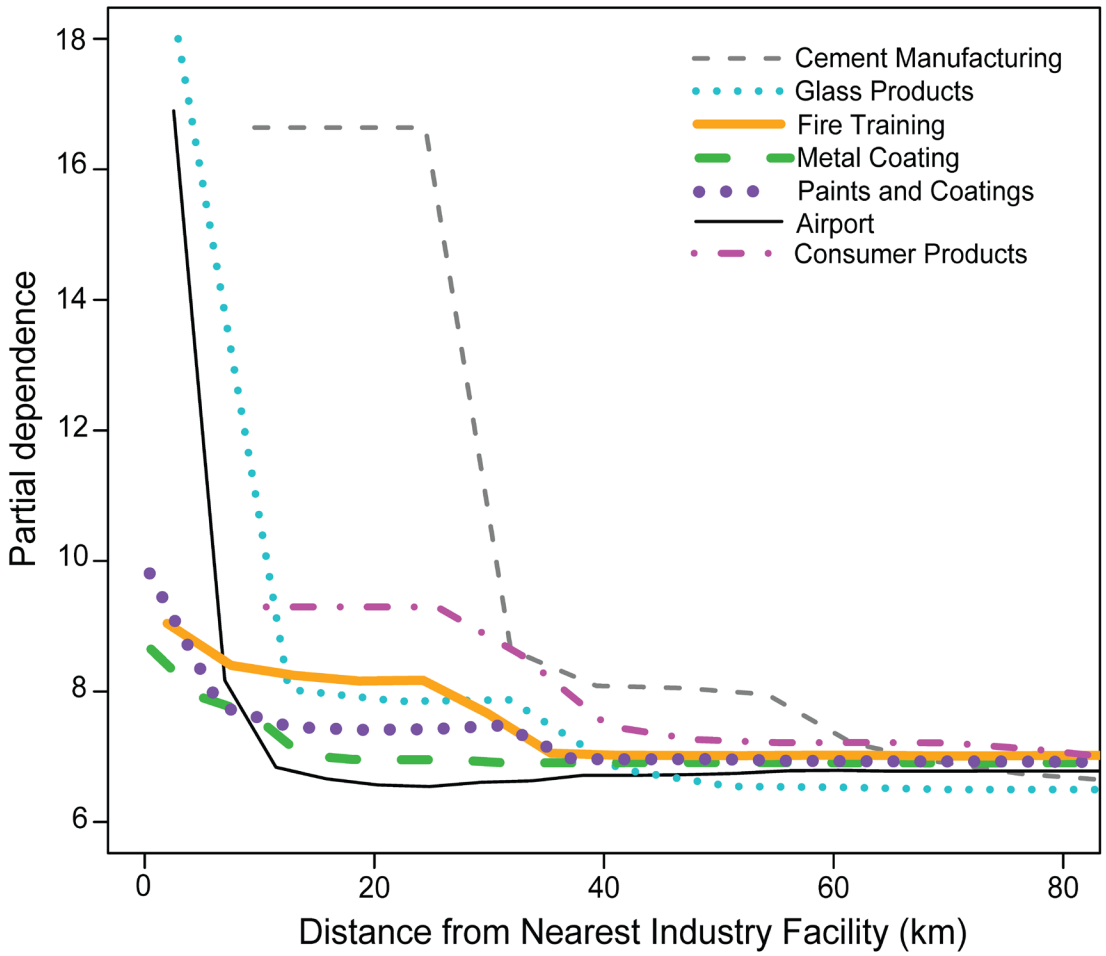
Threshold = 3 ng/g

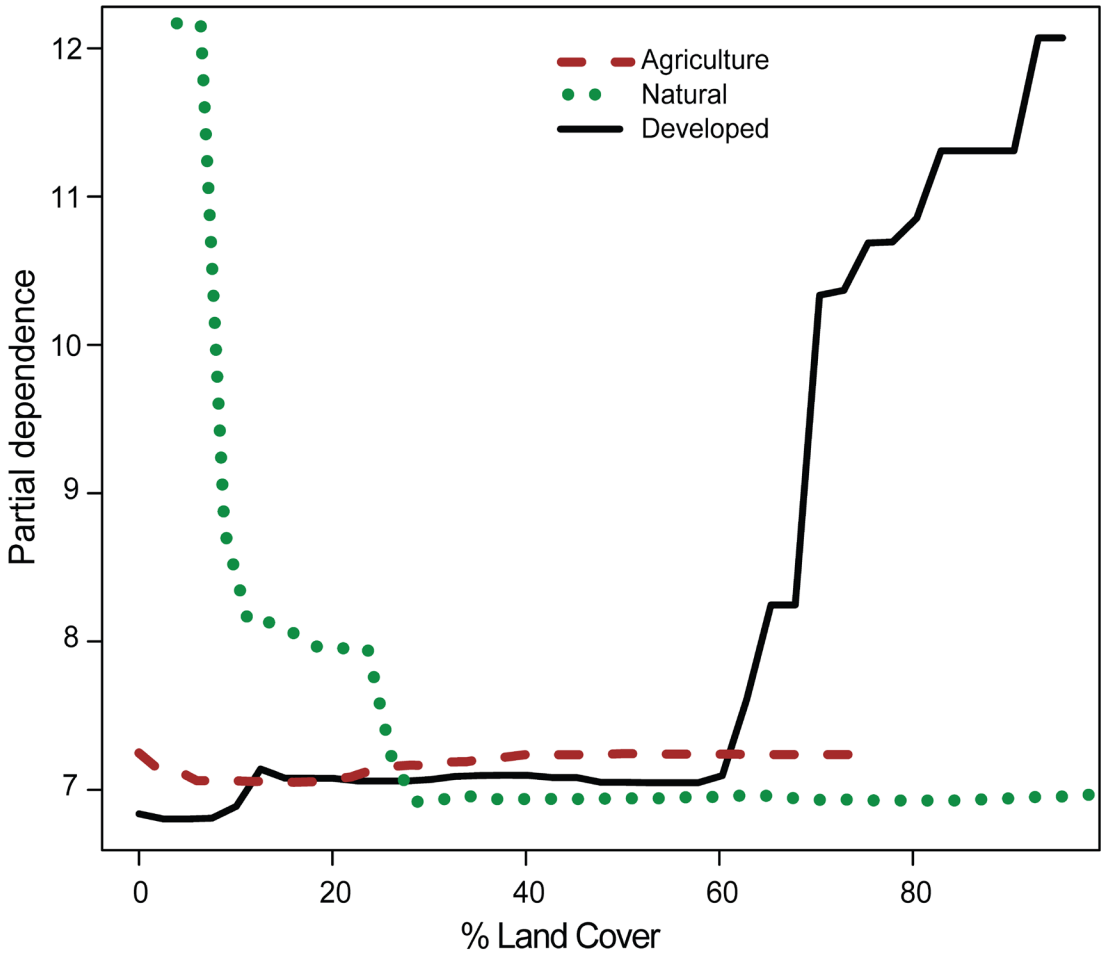


Results – Variable Importance



Results – Regression, Partial Dependence Plot





Summary

Contact:
Nikki DeLuca
deluca.nikki@epa.gov

- Piloted a modeling approach for PFAS contamination in fish tissue that can be used for screening and site prioritization in the Columbia River Basin
- Spatial data for potential sources (EPA's PFAS Analytic Tools) and land cover used as predictor variables in Random Forest models
- Found existing PFAS measurements in environmental media (EPA's PFAS Analytic Tools) to develop and evaluate models
- Regression models could be used to identify hotspots of PFAS in fish tissue
- Regulators could utilize classification models by choosing a threshold concentration value meaningful to their jurisdiction's health advisory, chemical, or vulnerable populations
- **More details can be found in publication**:

DeLuca, N. M., Mullikin, A., Brumm, P., Rappold, A. G., & Cohen Hubal, E. (2023). Using Geospatial Data and Random Forest To Predict PFAS Contamination in Fish Tissue in the Columbia River Basin, United States. *Environmental Science & Technology*, *57*(37), 14024-14035.

https://pubs.acs.org/doi/full/10.1021/acs.est.3c03670