Seeing the Forest for the (Decision) Trees:

Machine-Learning Enhances PFAS Analytics

Skyler Sorsby, WSP USA

Paul Hurst, WSP Canada

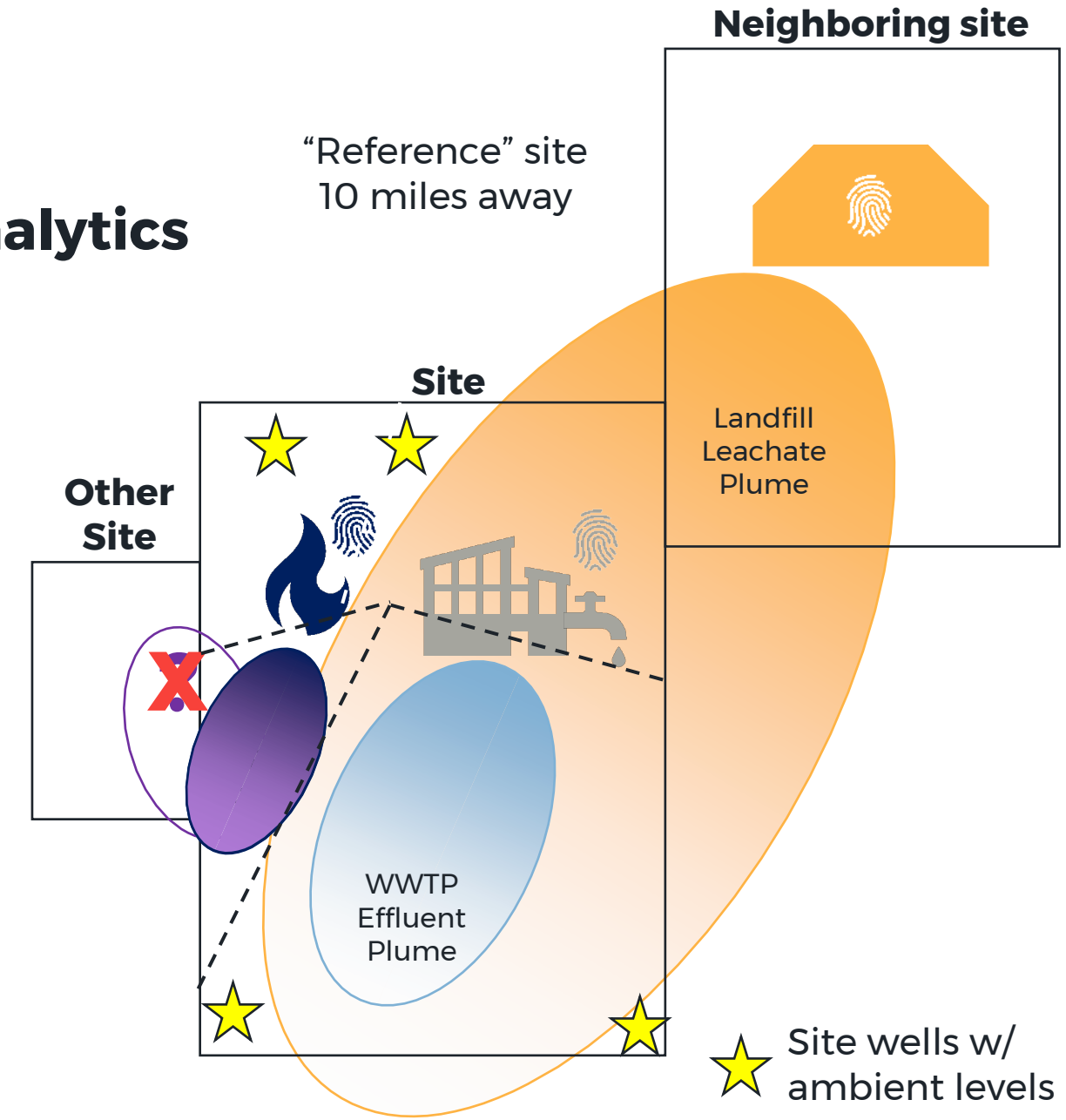
Stefano Marconetto, WSP Canada

2024/06/06

Why do I need advanced analytics for PFAS?

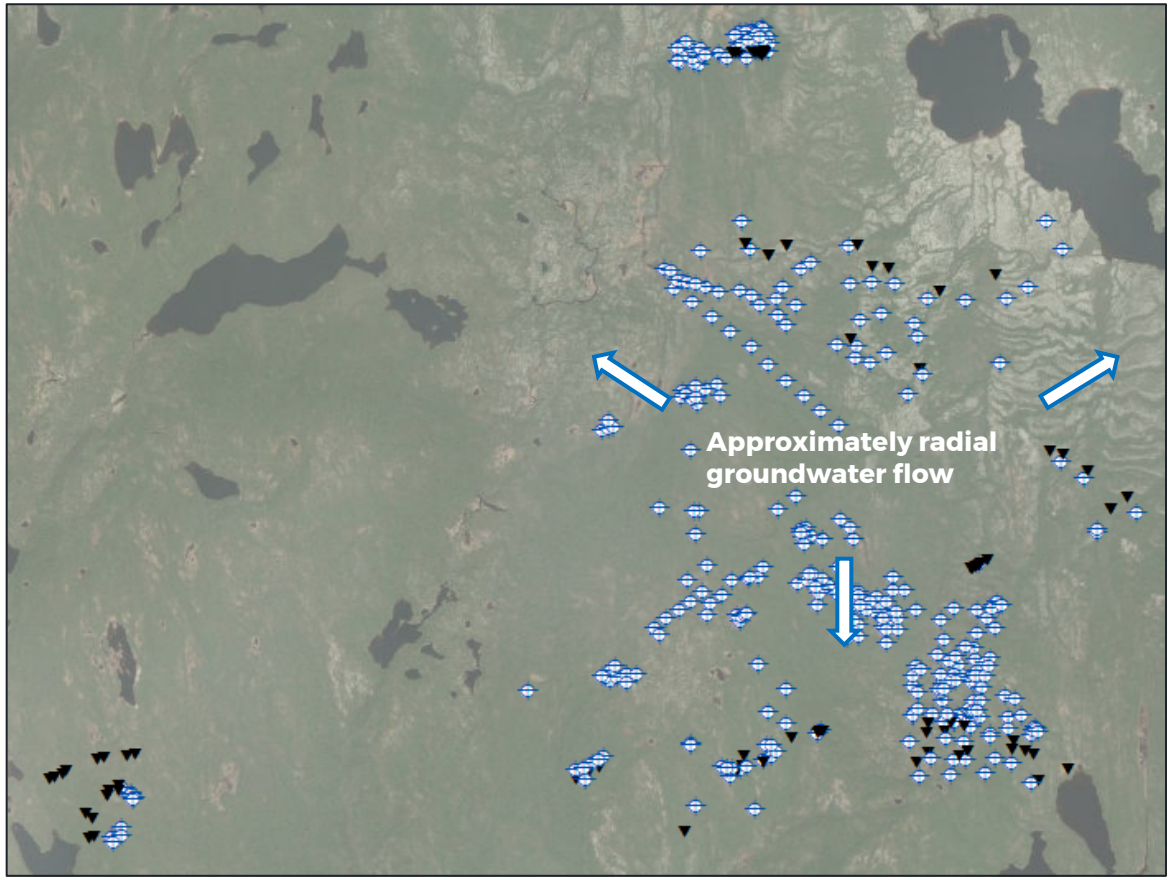
- Locate multiple **source areas**
  - Onsite vs. neighboring site
  - Varying onsite sources (i.e., multiple generations of AFFF)
- Distinguish sources from **transport patterns**
- Identify potential **precursor contributions**
- Define site-specific **ambient levels**
  - Up to 11 ppt (PFOA) and 14 ppt (PFOS) **in precipitation alone**[1]
- **Data-driven site partitioning** for investigation / remediation

[1]*Gewurtz et al. (2019)*



Case Study

- Confidential airport site in Canada (data modified to preserve anonymity).
- Multiple onsite sources.
- Potential non-source fingerprints.
- Potential links between surface water and groundwater.
- Potential for ambient levels.
- Radial groundwater flow away from the site.



Ambient levels

Estimate ambient levels

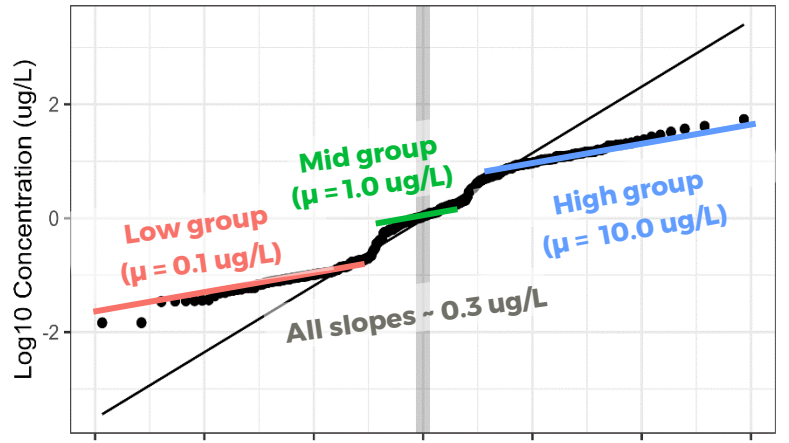
Gaussian Mixture Model (GMM) clustering

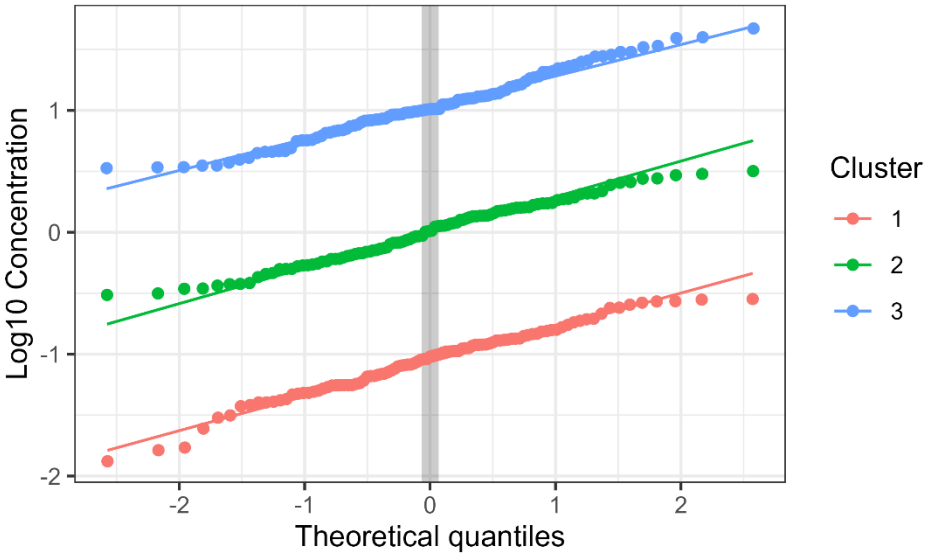
Problem: site-specific ambient levels may **exceed regulatory criteria**. Reference locations may **not adequately represent** full range of site-specific levels.

Solution: GMM uses **familiar statistical practices** to extract low vs high concentrations reflective of ambient levels and impacts.

Benefit: site-specific levels based on comprehensive dataset, **determine feasible cleanup limits**.

Sample	PFOS	PFOS	PFHxS	…
1	↑	↑	↑	↑
2	↕	↕	↕	↕
3	↕	↕	↕	↕
…	↓	↓	↓	↓





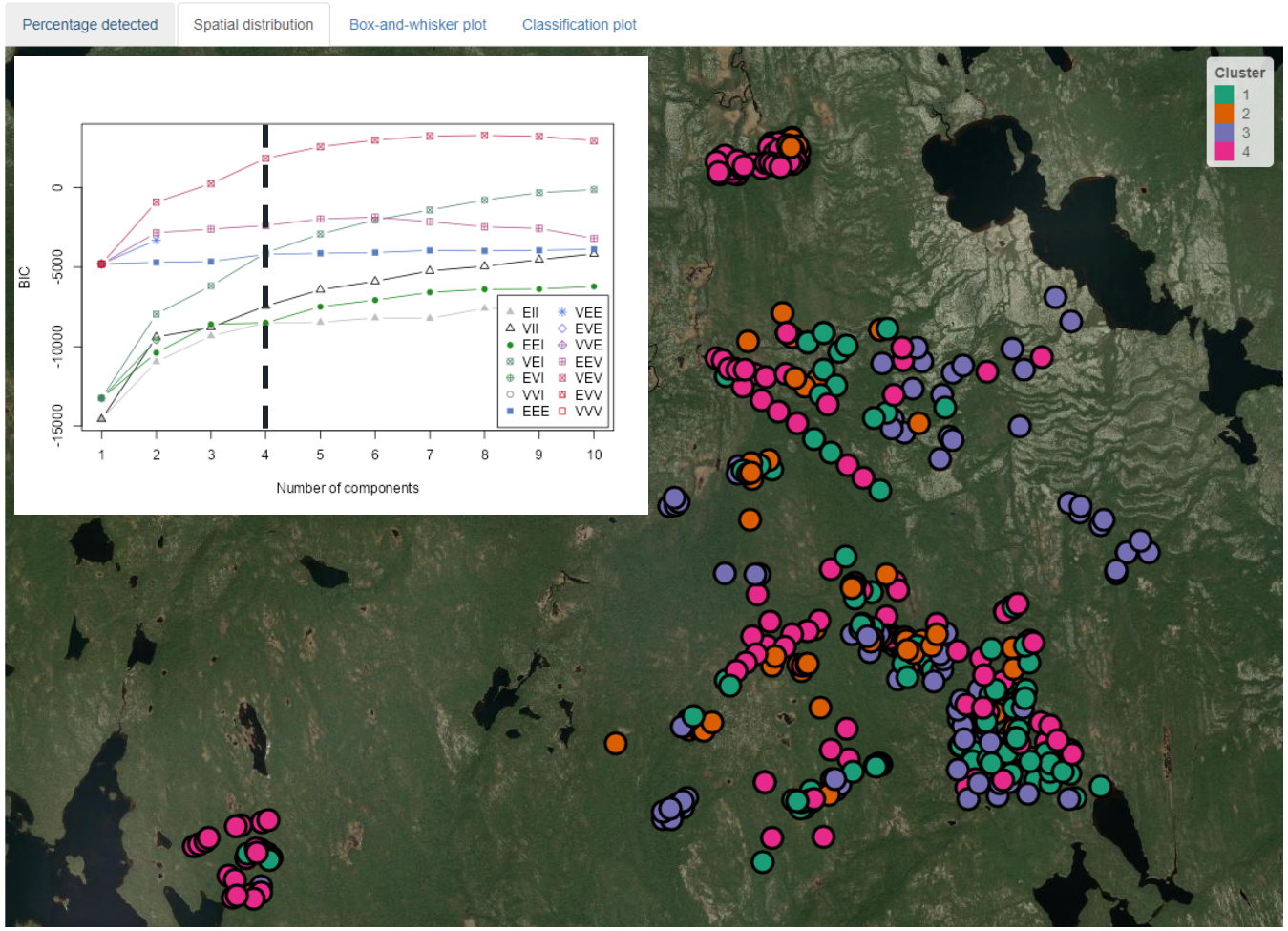
Ambient levels

Estimate ambient levels

Sample	PFOS	PFOS	PFHxS	…
1	↑	↑	↑	↑
2	↕	↕	↕	↕
3	↕	↕	↕	↕
…	↓	↓	↓	↓

Gaussian Mixture Model (GMM) clustering

- Check goodness-of-fit metrics
- Check spatial sensibility.
  - Nested physiography
  - Partitioned between geologic units
  - Upgradient / downgradient
- Four clusters fit well and align with spatial context.
  - Hot-spot locations
  - Downgradient transport/dilution
  - Proximal and distal reference areas

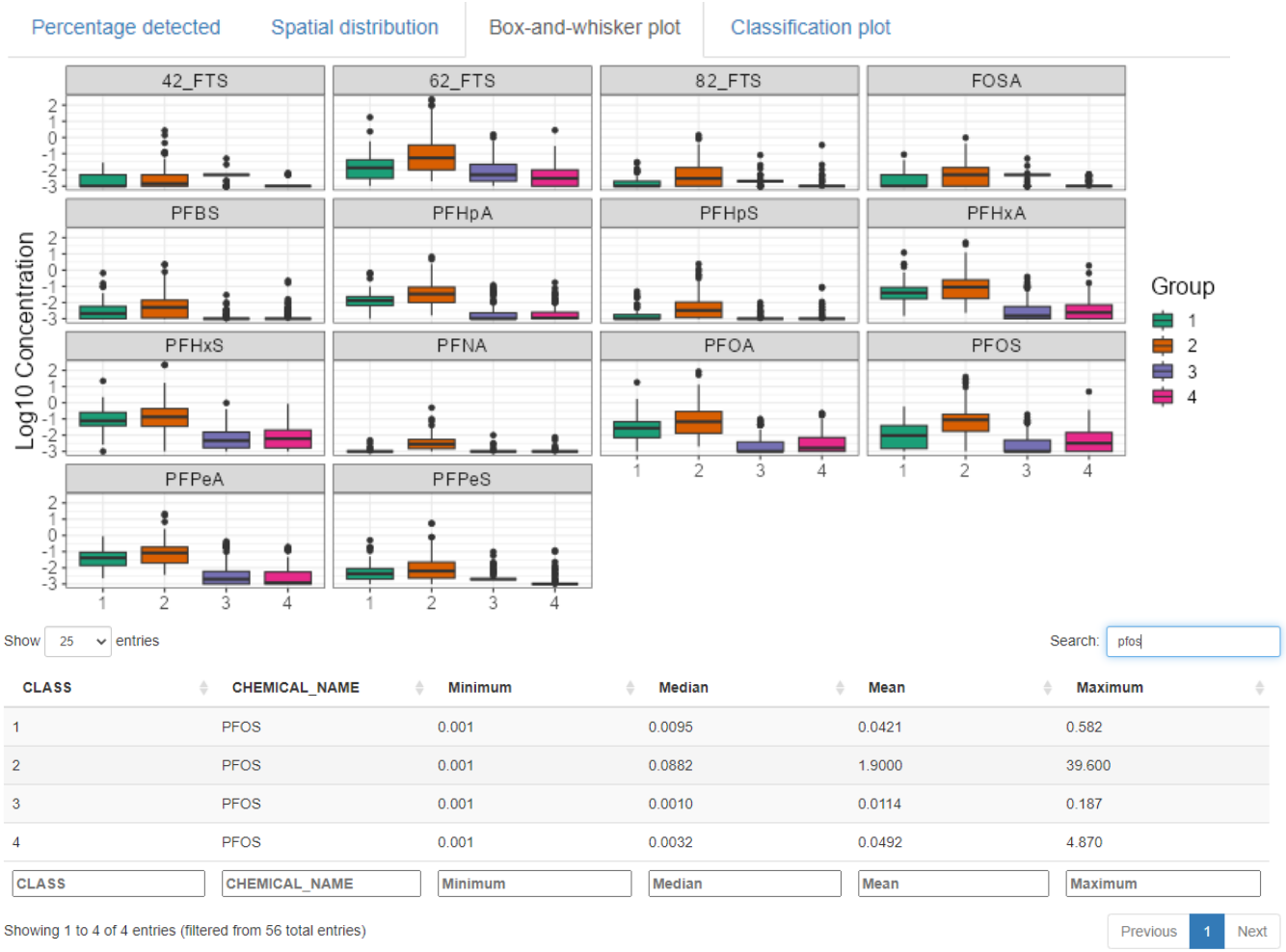


Estimate ambient levels

Gaussian Mixture Model (GMM) clustering

- Clear difference in concentration levels between groups.
- **Clusters 1 & 2** = potential point impacts / sources.
- **Cluster 3** = potential diffuse impacts / transport-related dilution.
- **Cluster 4** = potential ambient levels.

Sample	PFOS	PFOS	PFHxS	…
1	↑	↑	↑	↑
2	↕	↕	↕	↕
3	↕	↕	↕	↕
…	↓	↓	↓	↓



Fingerprinting

Un-mix signatures

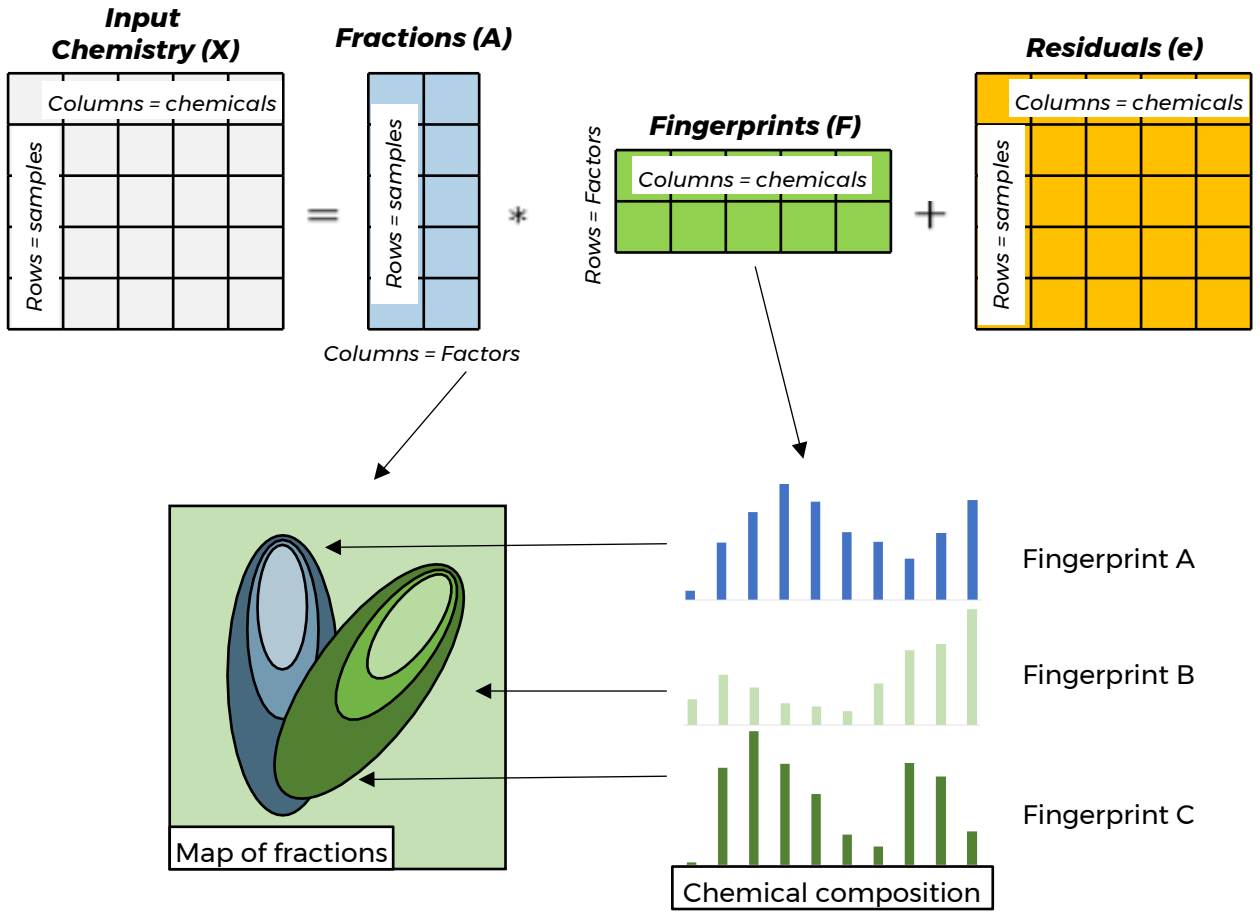
Blind “Source” Separation

Problem: chemical signatures **difficult to resolve** by traditional graphical methods.

Solution: Matrix factorization performs **“blind source separation”** without prior knowledge.

Benefit: **data-driven fingerprints** + spatial distribution to assess onsite / offsite sources & fate-and-transport patterns.

Sample	PFOS	PFOS	PFHxS	…
1	←			→
2	←			→
3	←			→
…	←			→



Fingerprinting

Un-mix signatures

Blind Source Separation

- Resolved fingerprints reflect “pure” versions of the most unique profiles that **bracket the data**.
- Should **align with real samples** / established profiles.
- EMs 2-3 & 5-8 potential source / F&T patterns
- EM4 ~ a profile of detection limits (background)

Sample	PFOS	PFOS	PFHxS	…
1	←			→
2	←			→
3	←			→
…	←			→

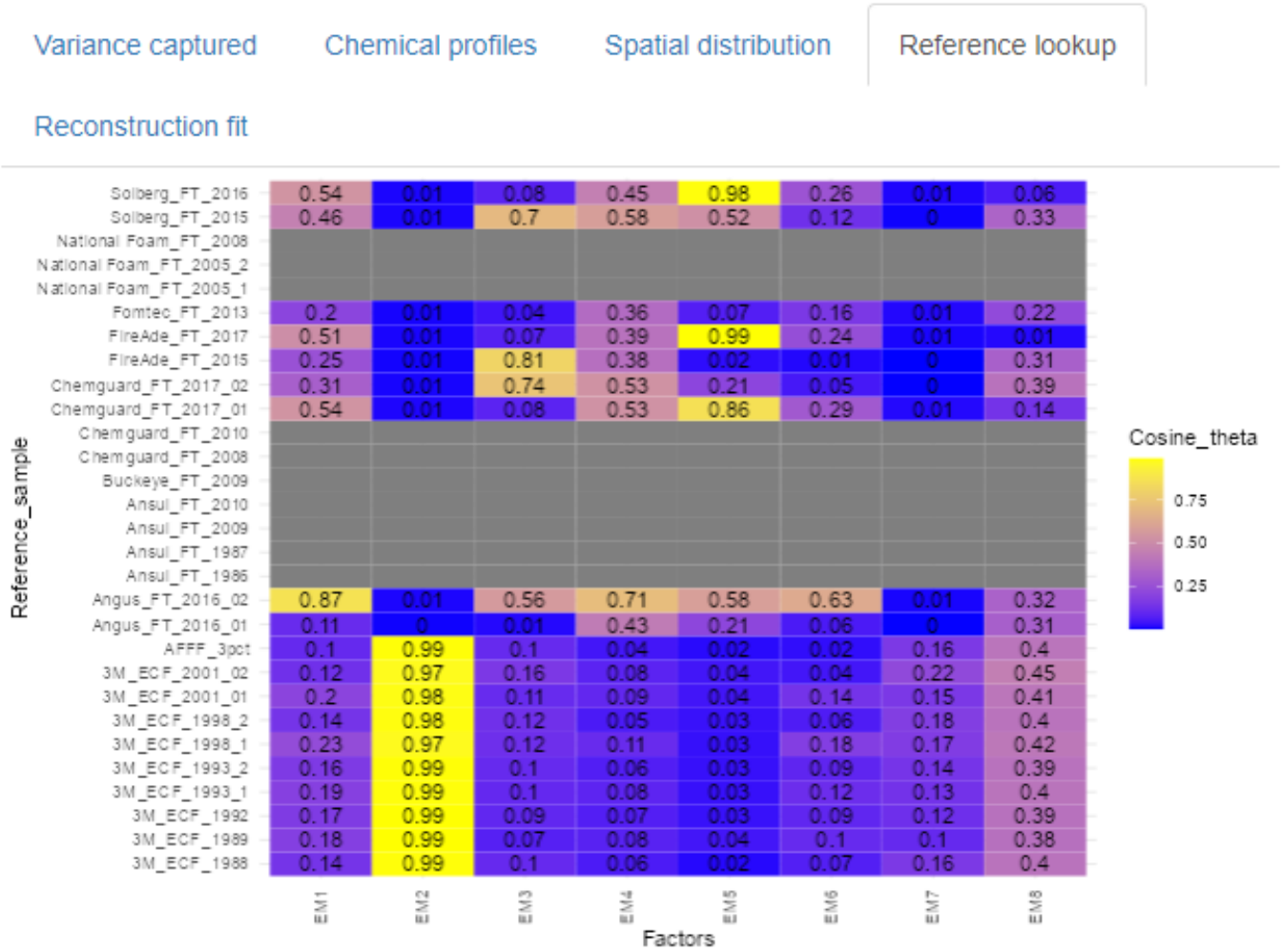
Fingerprinting

Un-mix signatures

Cosine Theta

- Commonly-used metric compares fingerprint profiles (rows) to reference profiles.
- Similar to correlation calculations for features (columns).
- EMs 1, 2, and 5 relate strongly with **different formulations of legacy and modern foam.**

Sample	PFOS	PFOS	PFHxS	…
1	←			→
2	←			→
3	←			→
…	←			→



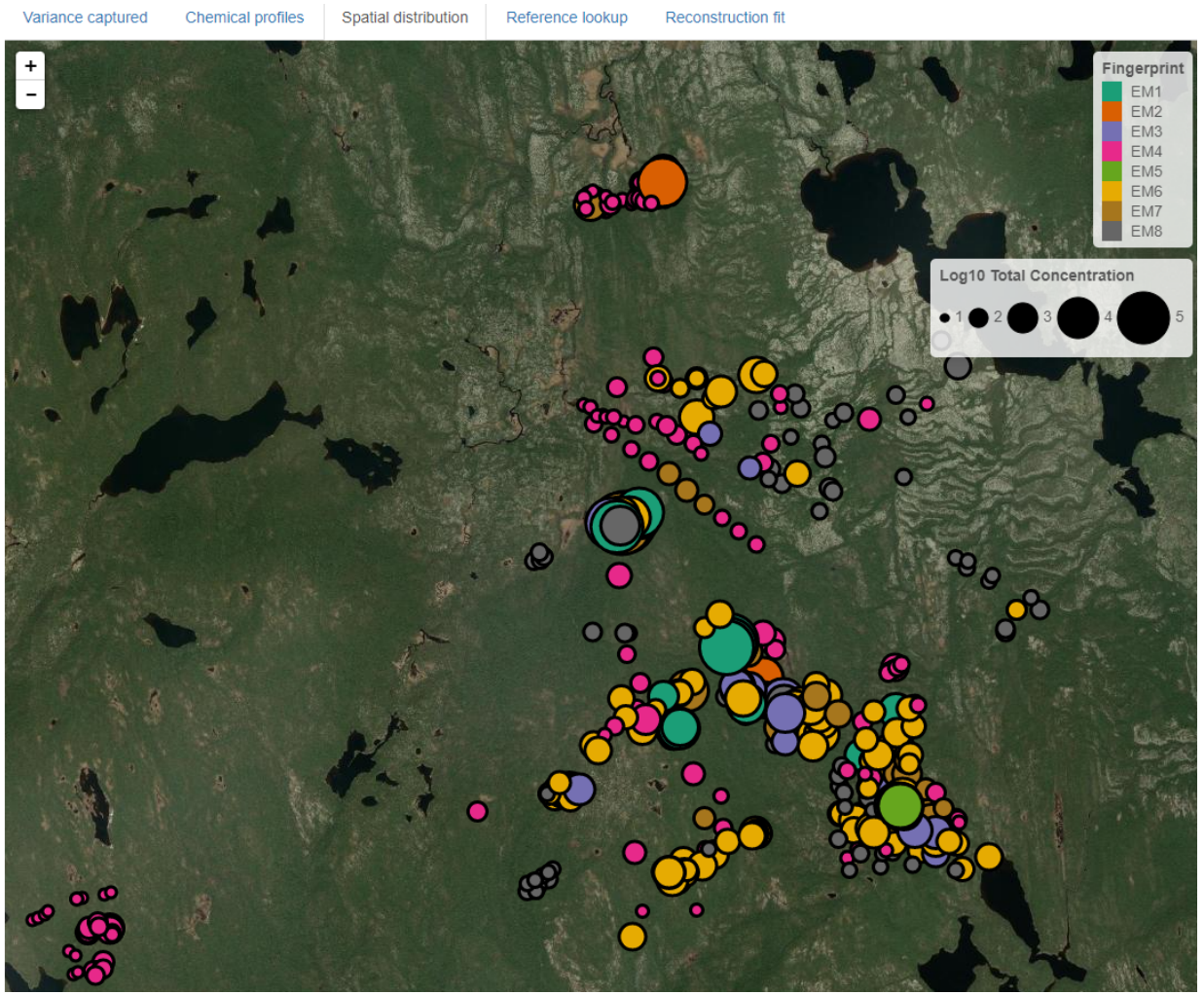
Fingerprinting

Un-mix signatures

Spatial fingerprints

- EM4 relates to low concentrations around the perimeter of the site / offsite (**ambient signature**).
- EMs 1, 2, and 5 indicate three **point sources** onsite, and one offsite north.
- EMs 3, 6, 7, and 8 indicate potential **F&T / precursor transformation patterns**.

Sample	PFOS	PFOS	PFHxS	…
1	←			→
2	←			→
3	←			→
…	←			→



Design

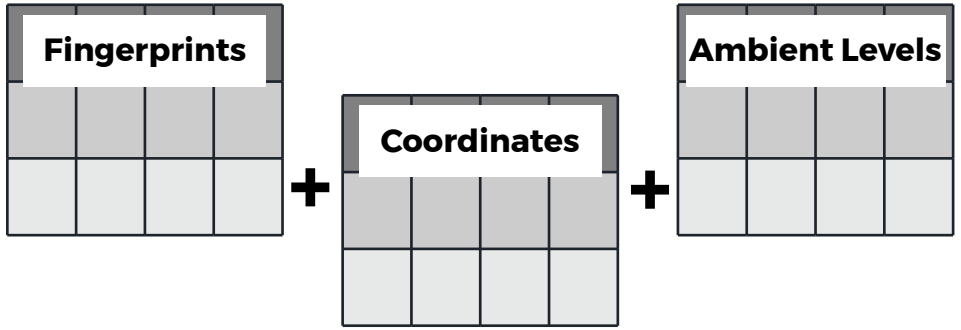
Site partitioning

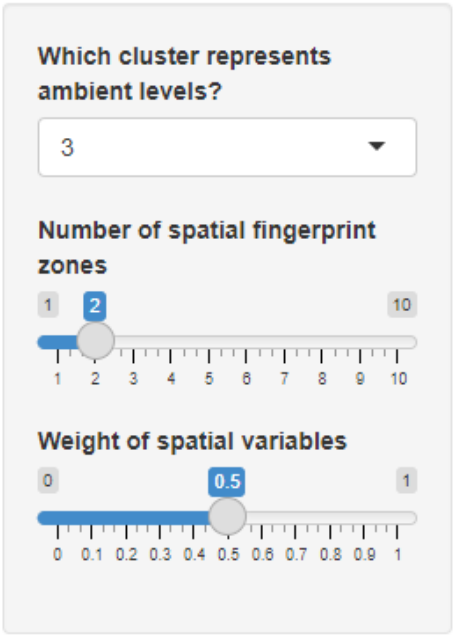
Spatial clustering

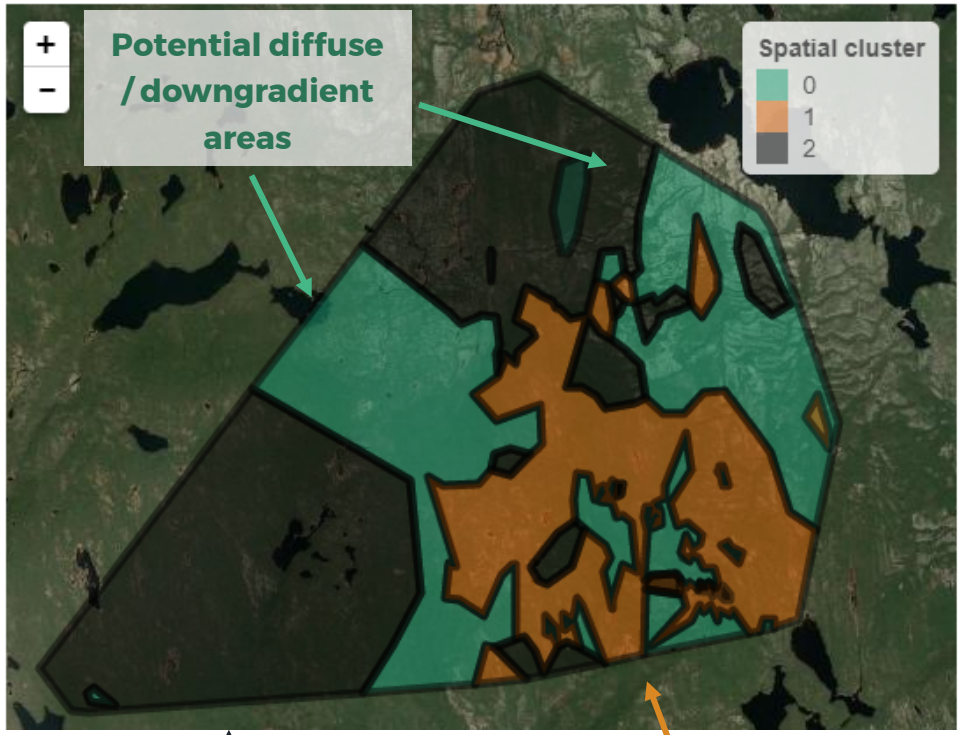
Problem: **Challenging to partition areas** for detailed investigation and management at complex sites.

Solution: Spatial clustering incorporates fingerprints, ambient levels, and spatial relationships to **define chemically similar zones**.

Benefit: **Data-driven site partitioning** for further investigation or remediation, incorporating knowledge of ambient levels.







Complex interactions

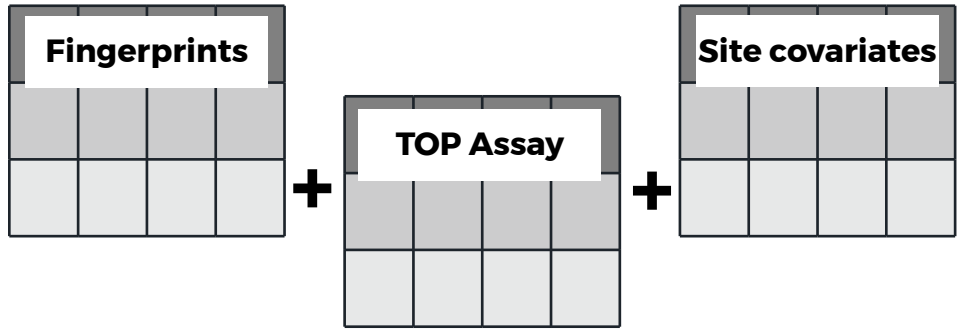
Explainable AI (XAI)

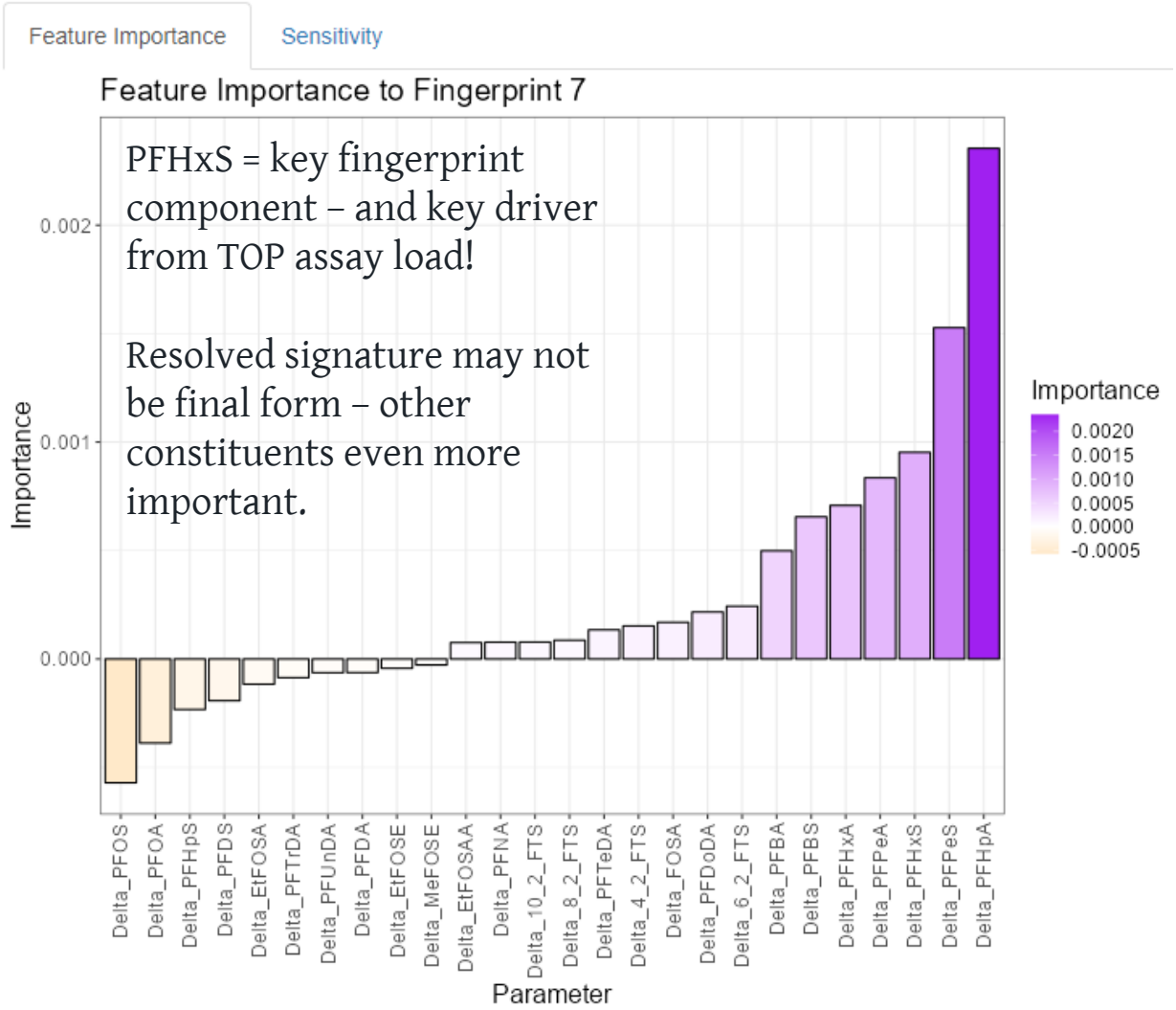
Feature importance

Problem: PFAS precursors may continue to **transform to PFAAs over time.** Other measurements may co-vary with other F&T patterns.

Solution: Feature importance **identifies potential drivers in a single graphic**, using ensemble tree-based nonlinear regression. One of **many XAI tools available!**

Benefit: Simplify complex data relations and **identify key geochemical / precursor drivers**, guide remedial design.





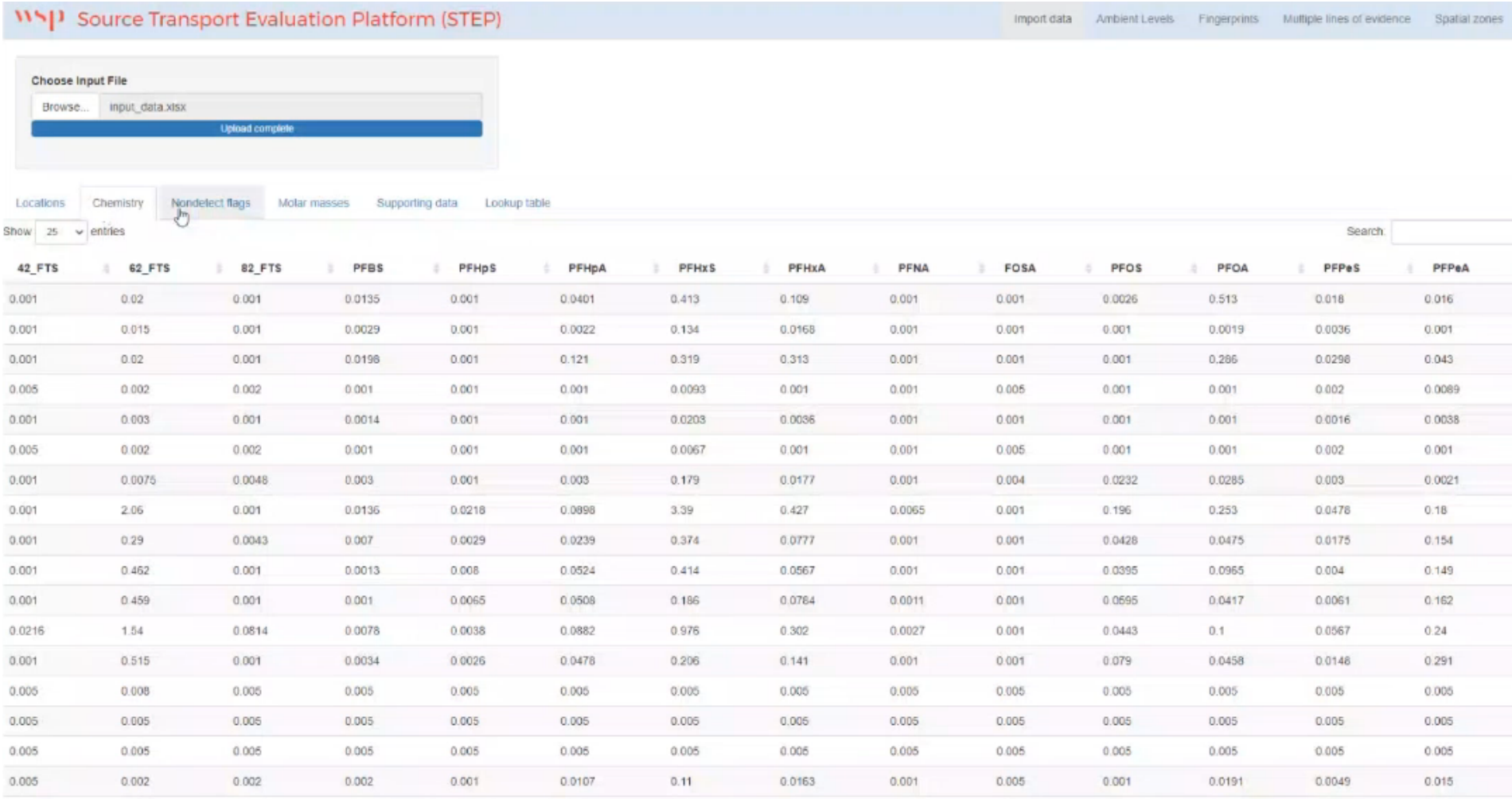
WSP Source & Transport Evaluation Platform (STEP)

Capabilities and outlook

- In-house web application focused on answering strategic site investigation questions:
  - What are **ambient levels**?
  - Is this a **source, or a fate-and-transport pattern**?
  - How should we **divide the site** for further investigation?
- Reduces analysis time-to-insight from years/months to weeks/days.
- Complement/vet Conceptual Site Model derived from traditional tools.
- Early insights promote better planning & more robust decisions.

Q&A

More WSP PFAS Capabilities



42_FTS	62_FTS	82_FTS	PFBS	PFHpS	PFHpA	PFHxS	PFHxA	PFNA	FOSA	PFOS	PFOA	PFPeS	PFPeA
0.001	0.02	0.001	0.0135	0.001	0.0401	0.413	0.109	0.001	0.001	0.0026	0.513	0.018	0.016
0.001	0.015	0.001	0.0029	0.001	0.0022	0.134	0.0168	0.001	0.001	0.001	0.0019	0.0036	0.001
0.001	0.02	0.001	0.0198	0.001	0.121	0.319	0.313	0.001	0.001	0.001	0.286	0.0298	0.043
0.005	0.002	0.002	0.001	0.001	0.001	0.0093	0.001	0.001	0.005	0.001	0.001	0.002	0.0089
0.001	0.003	0.001	0.0014	0.001	0.001	0.0203	0.0036	0.001	0.001	0.001	0.001	0.0016	0.0038
0.005	0.002	0.002	0.001	0.001	0.001	0.0067	0.001	0.001	0.005	0.001	0.001	0.002	0.001
0.001	0.0075	0.0048	0.003	0.001	0.003	0.179	0.0177	0.001	0.004	0.0232	0.0285	0.003	0.0021
0.001	2.06	0.001	0.0136	0.0218	0.0898	3.39	0.427	0.0065	0.001	0.196	0.253	0.0478	0.18
0.001	0.29	0.0043	0.007	0.0029	0.0239	0.374	0.0777	0.001	0.001	0.0428	0.0475	0.0175	0.154
0.001	0.462	0.001	0.0013	0.008	0.0524	0.414	0.0567	0.001	0.001	0.0395	0.0965	0.004	0.149
0.001	0.459	0.001	0.001	0.0065	0.0508	0.186	0.0784	0.0011	0.001	0.0595	0.0417	0.0061	0.162
0.0216	1.54	0.0814	0.0078	0.0038	0.0882	0.976	0.302	0.0027	0.001	0.0443	0.1	0.0567	0.24
0.001	0.515	0.001	0.0034	0.0026	0.0478	0.206	0.141	0.001	0.001	0.079	0.0458	0.0148	0.291
0.005	0.008	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
0.005	0.002	0.002	0.002	0.001	0.0107	0.11	0.0163	0.001	0.005	0.001	0.0191	0.0049	0.015