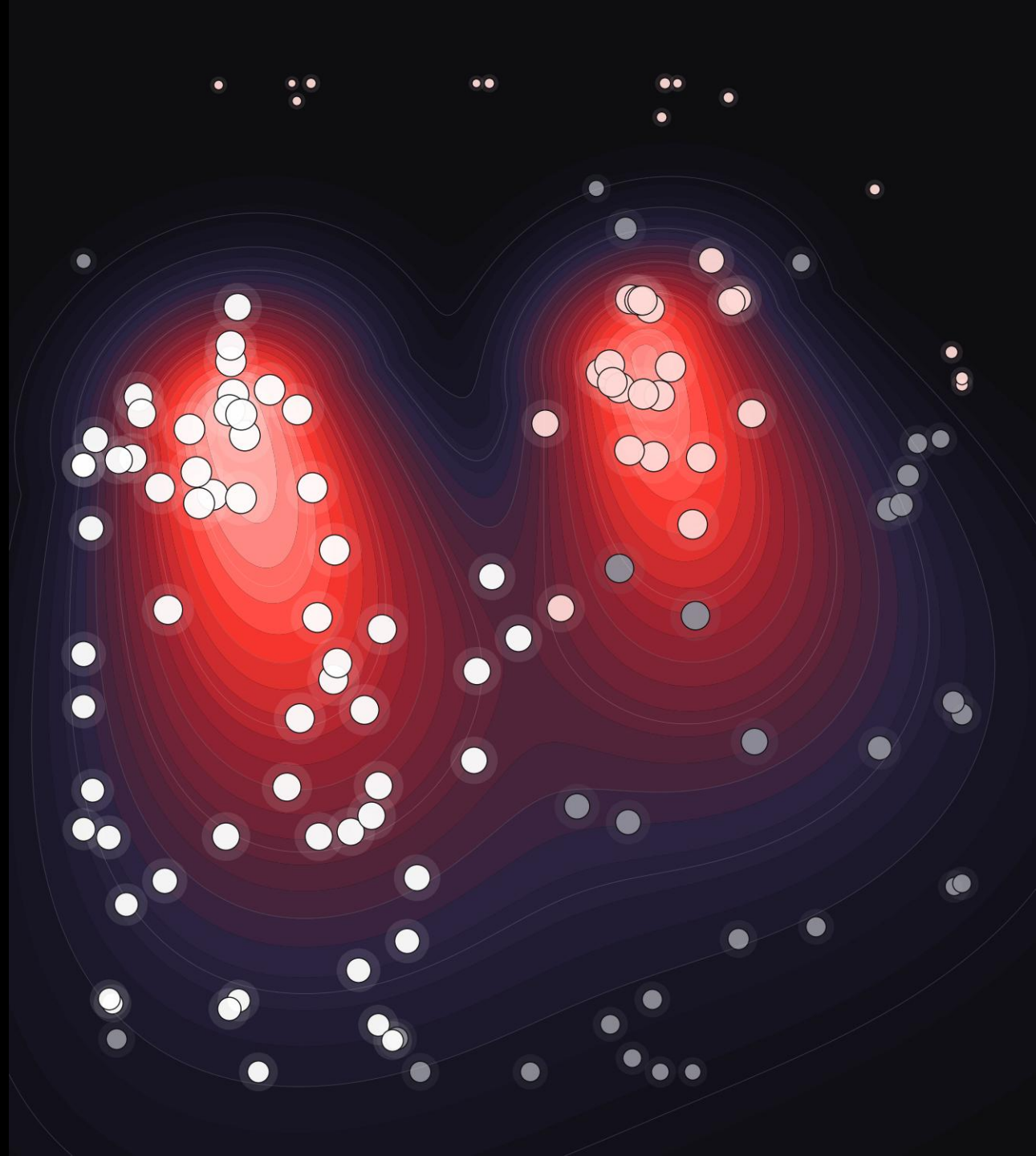
FROM DATA TO DECISIONS

Transforming Contaminated Site Management with AI and Machine Learning



JADEN FORD, P.Eng., PMP | WSP CANADA INC.

RemTech 2026 | Kelowna, BC

Itinerary

01 **Why now**
Bigger datasets, lower standards, new questions

02 **Four ideas**
The principles under every result

03 **Two tools**
Clustering and matrix factorization

04 **Three results**
Fingerprints, background and zones, then real projects

05 **Keeping it defensible**
No answer key means no accuracy score

06 **Getting started**
Where to begin, and questions

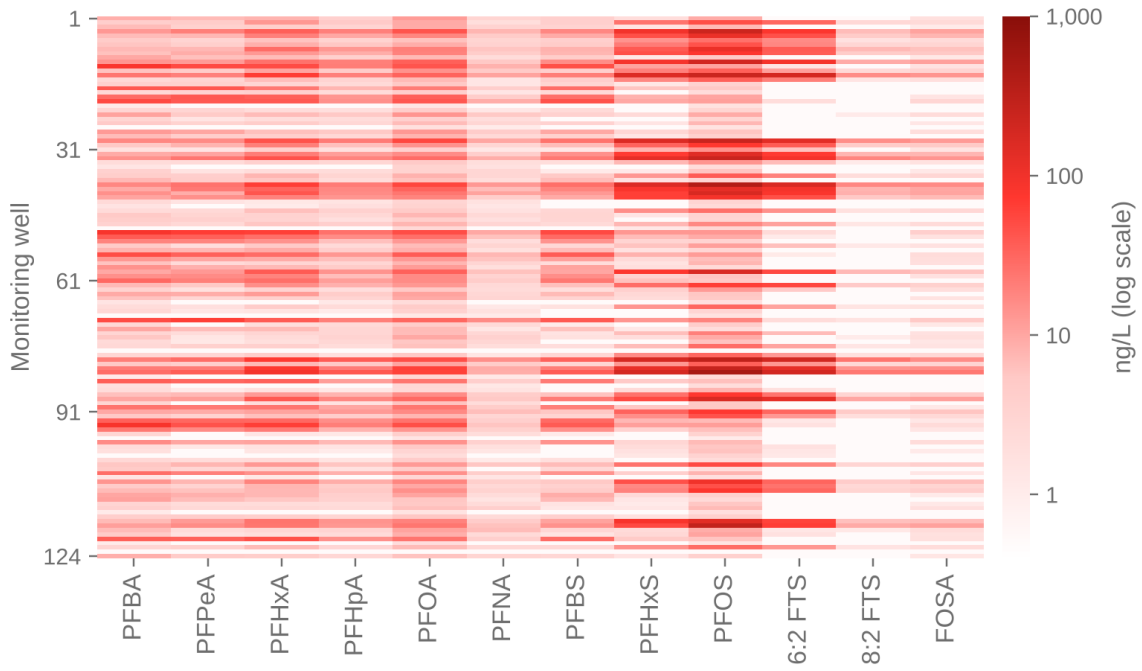
Built on the work of
Paul Hurst
WSP Canada
Skyler Sorsby
WSP USA, developer of STEP

PART 01 OF 06

WHY NOW?

Bigger datasets, lower standards, new questions

This is one sampling round



1,488

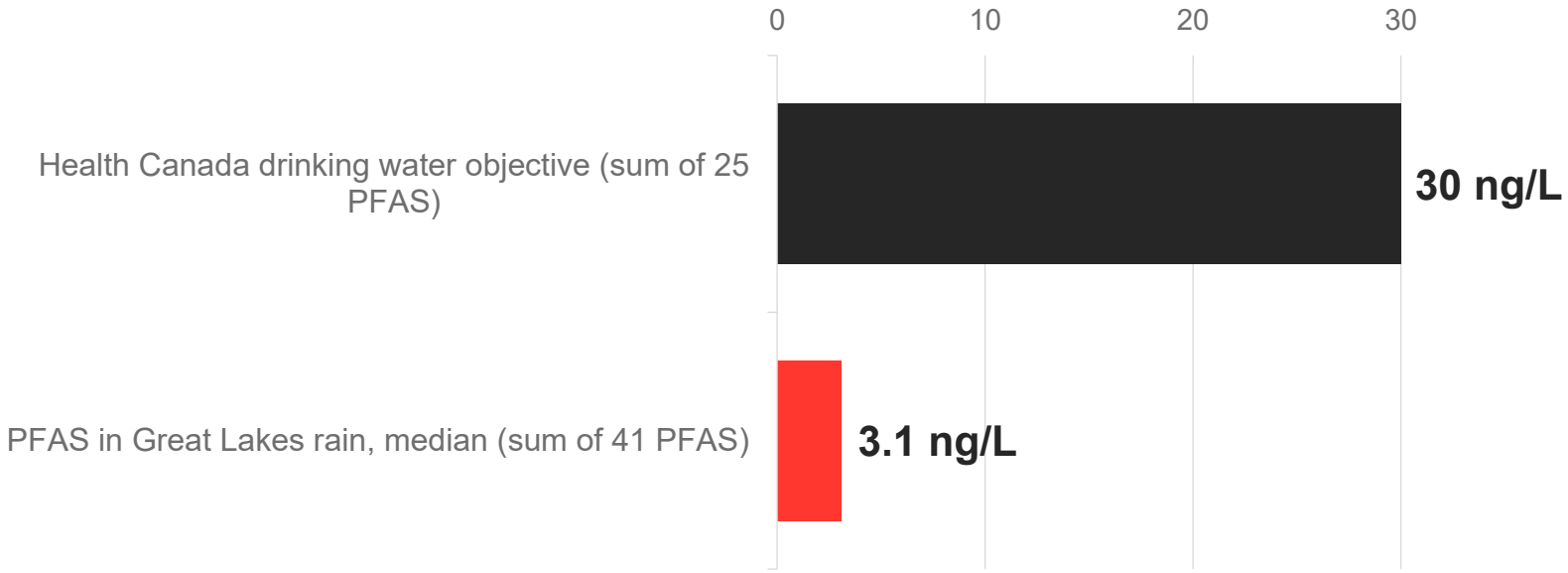
results from one sampling round: 124 wells × 12 PFAS analytes

≈ 18,000

results after three years of quarterly sampling

Typical practice contours two or three analytes. The rest goes to an appendix.

The standard, and what falls with the rain



≈ 1/10

of the drinking water objective can arrive as rain, before any spill

Site medians in rain ranged 2.4 to 4.5 ng/L. Analyte lists differ, so compare order of magnitude only.

The question used to be "is there contamination?" Now it is "is it ours?"

Sources: Health Canada (2024), Objective for Canadian Drinking Water Quality: Per- and Polyfluoroalkyl Substances. Xia, C., et al. (2024), Environ. Sci. Technol. 58(21):9303–9313.

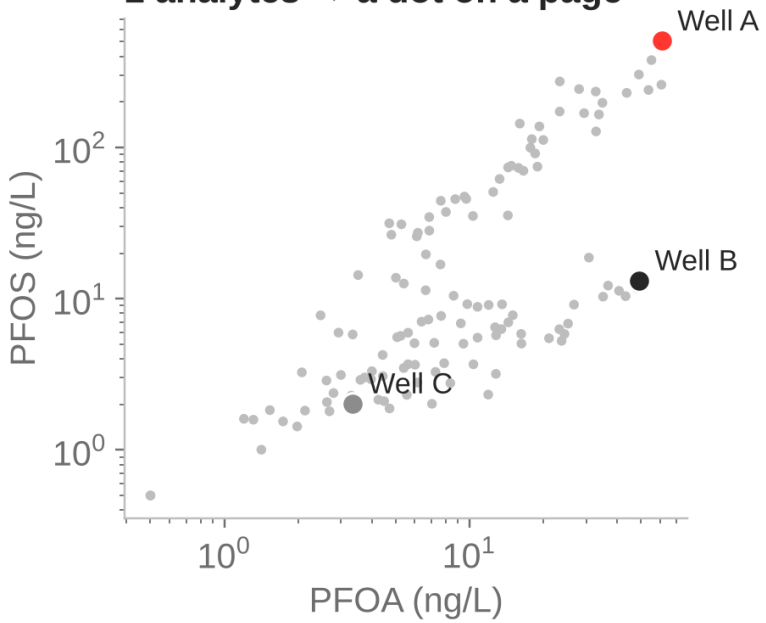
PART 02 OF 06

FOUR IDEAS

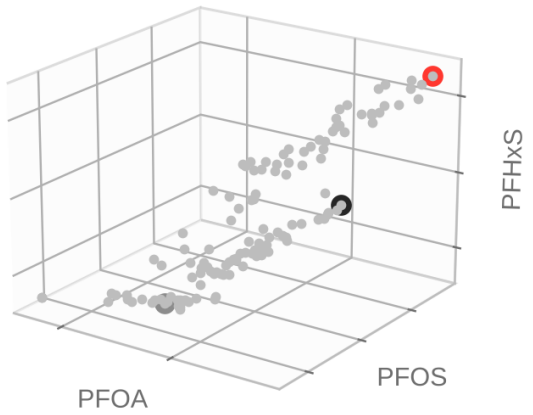
1. A sample is a point
2. Composition, not concentration
3. Similarity is an angle
4. There is no answer key

A sample is a point

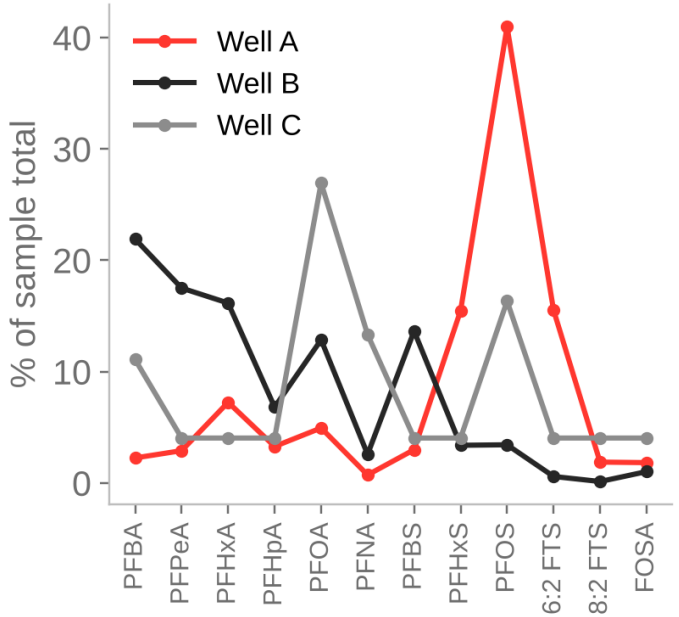
2 analytes → a dot on a page



3 analytes → a dot in a room

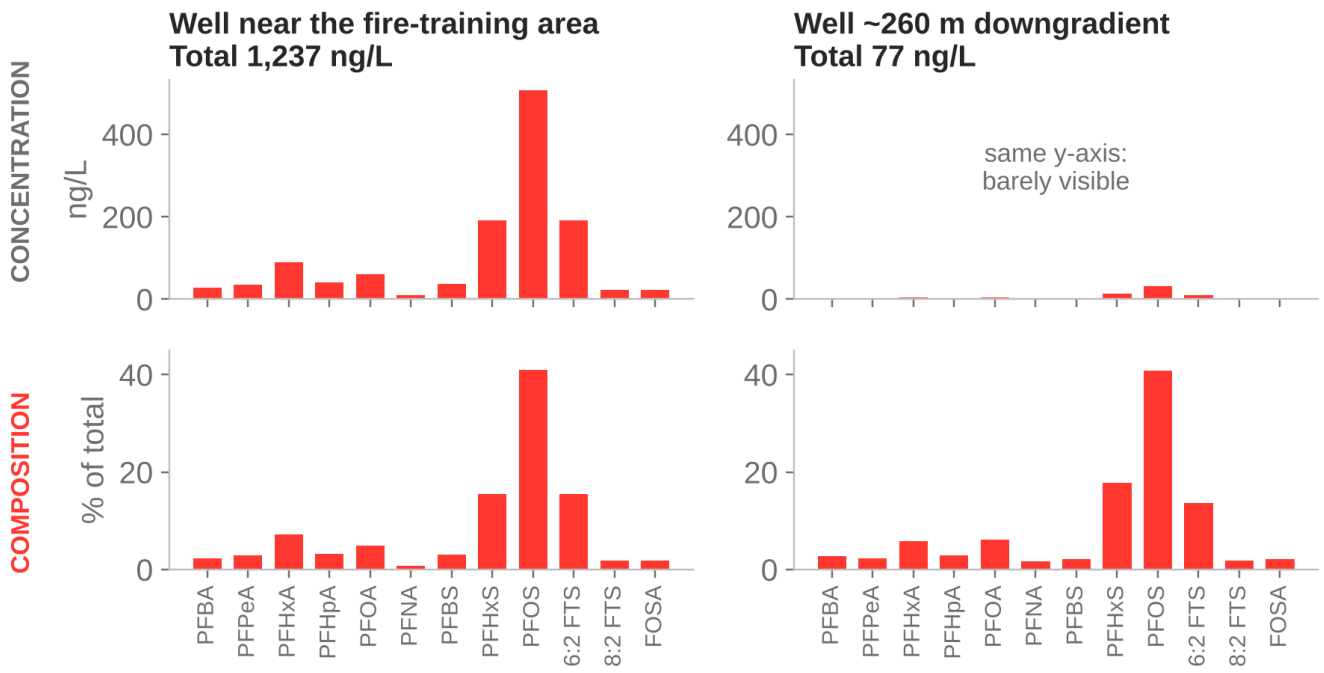


12 analytes → a dot we cannot draw



Twelve analytes are twelve coordinates. Distance works the same in twelve dimensions.

Composition, not concentration



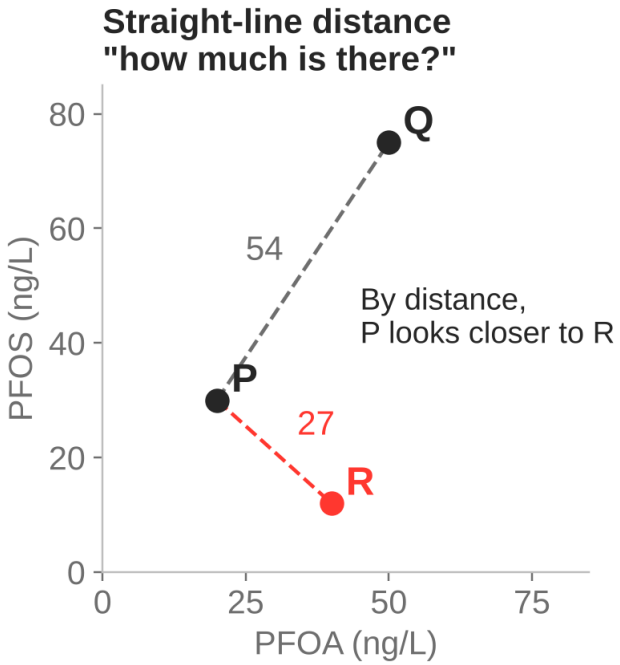
16×

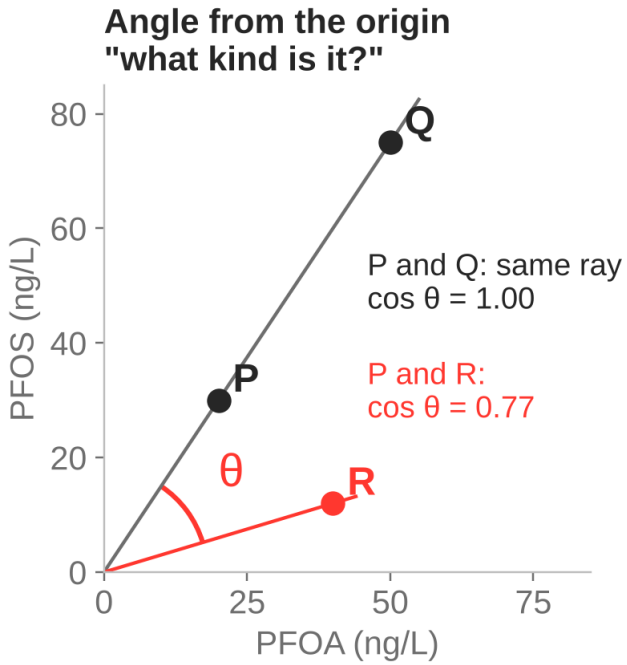
apart in concentration: same source, about 260 m downgradient

between their compositions: the same recipe

Concentration tells you how much. Composition tells you what it is.

Similarity is an angle, not a distance





Distance

Measures how much is there

P to R = 27, P to Q = 54

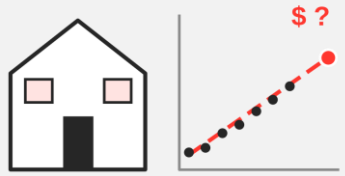
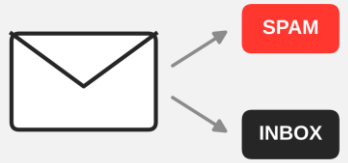
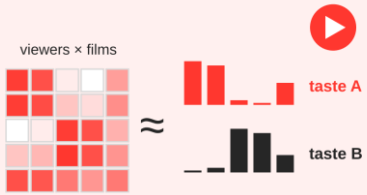
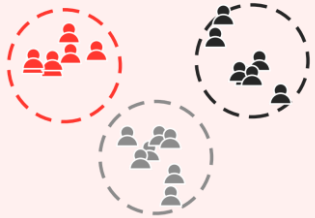
Angle

Measures what kind it is

P to Q: cos θ = 1.00; P to R: 0.77

cos θ = 1.00 means identical composition, whatever the concentration.

Four core machine learning tasks

	PREDICT AN AMOUNT	SORT INTO GROUPS
Answer key supervised	 **REGRESSION** Estimate a home's price from past sales	 **CLASSIFICATION** Filter spam using millions of labelled emails
No answer key unsupervised	 **MATRIX FACTORIZATION** Explain viewing habits as a mix of tastes	 **CLUSTERING** Group customers who behave alike

Contaminated sites sit in the bottom row: there is no answer key.

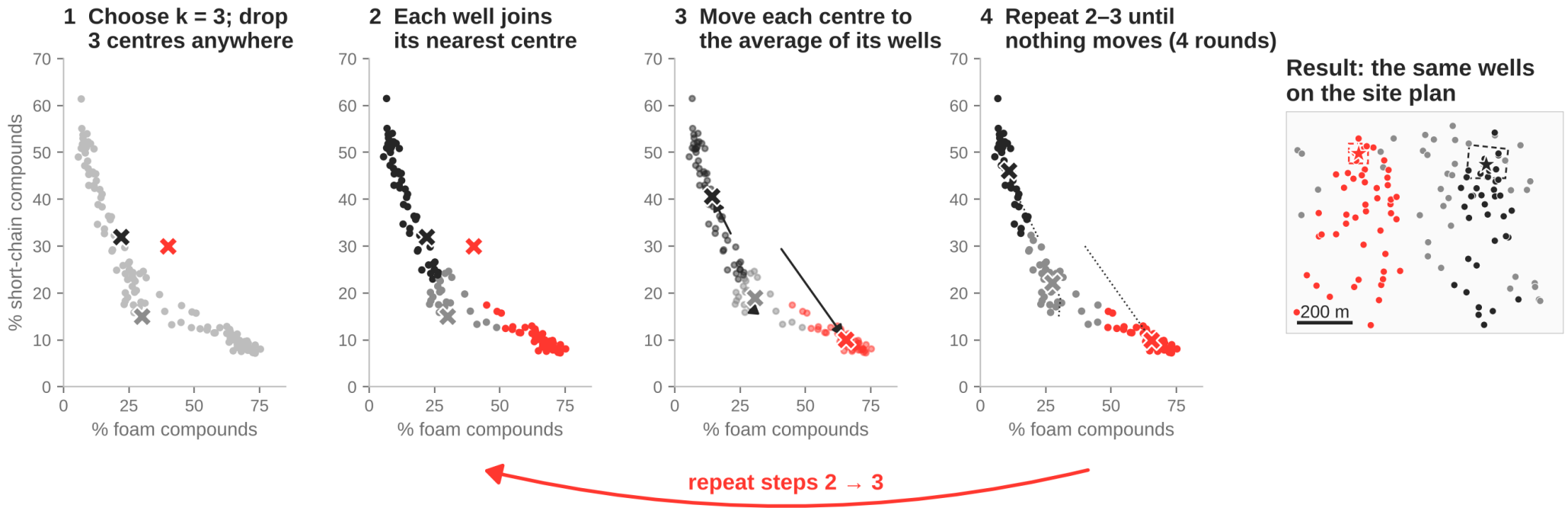
PART 03 OF 06

TWO TOOLS

Clustering asks: Which group does this sample belong to?

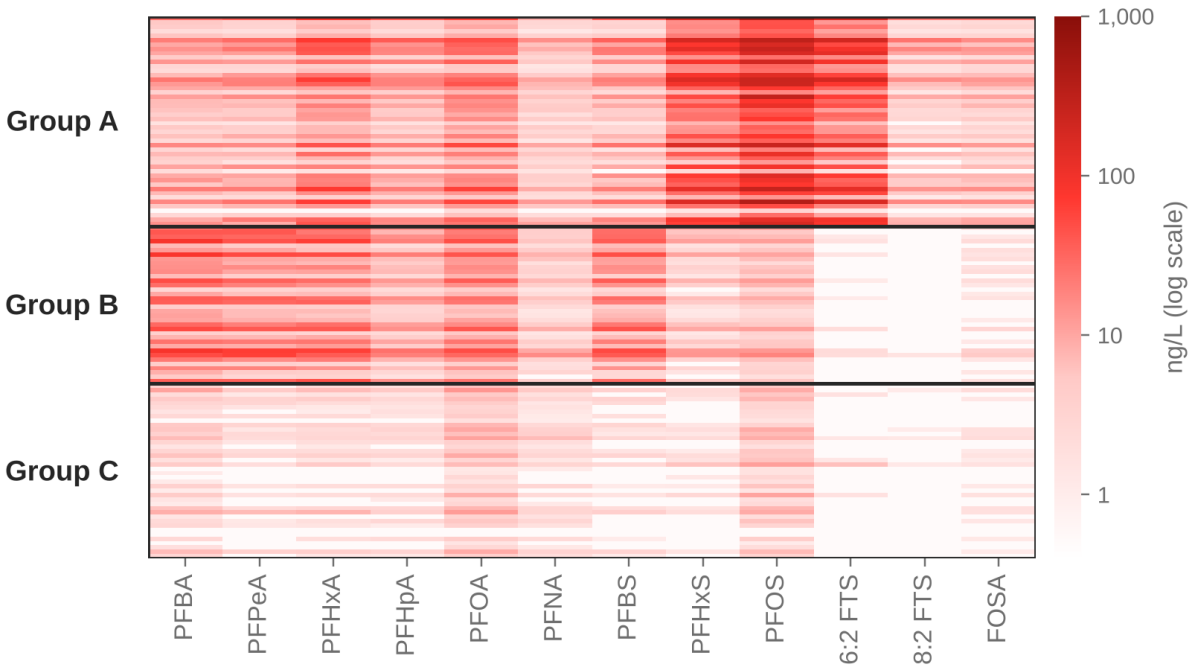
Factorization asks: What is this sample made of?

Clustering: the whole algorithm in four steps



This is k-means, which dates from the 1950s: simple arithmetic, repeated until it settles.

The same numbers, sorted by chemistry



Group A 48 wells

PFOS, PFHxS and 6:2 FTS dominate. Consistent with firefighting foam.

Group B 36 wells

Short-chain acids, almost no fluorotelomers. Consistent with landfill leachate.

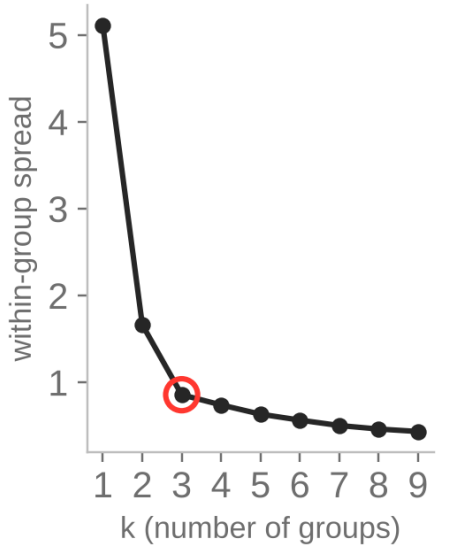
Group C 40 wells

Low everywhere, no dominant compound. The background candidate.

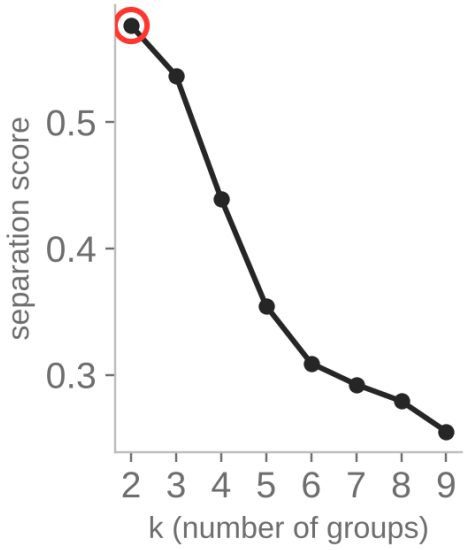
The structure was in the data all along, hidden by the row order.

Choosing k: diagnostics inform, the CSM decides

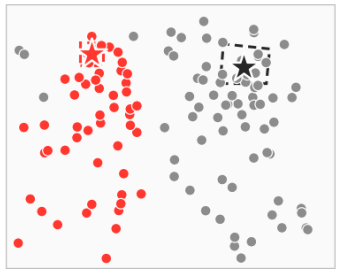
Elbow method
look for the bend → 3



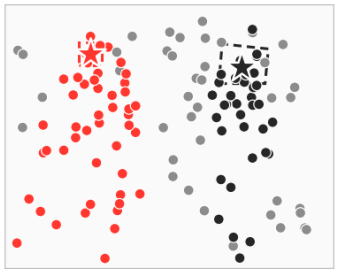
Silhouette score
highest = best → 2



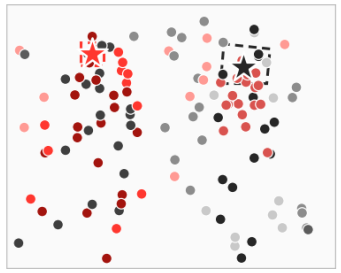
k = 2
two halves



k = 3
two plumes + background



k = 9
confetti



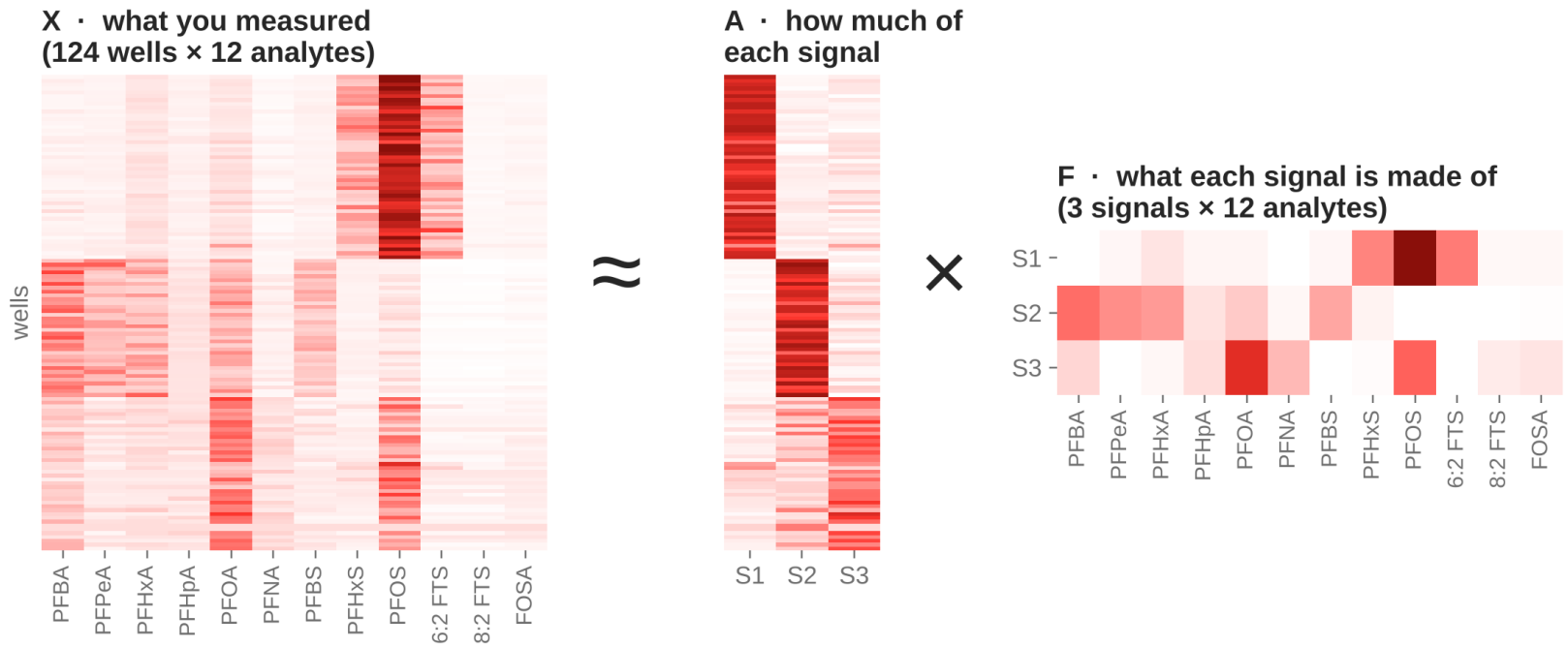
Set k from the diagnostics and the conceptual site model, and record why, before interpreting the maps.

Matrix factorization: a recipe, not a label



Clustering gives each sample one label. Factorization gives each sample a recipe.

One equation: X ≈ A × F



Non-negative

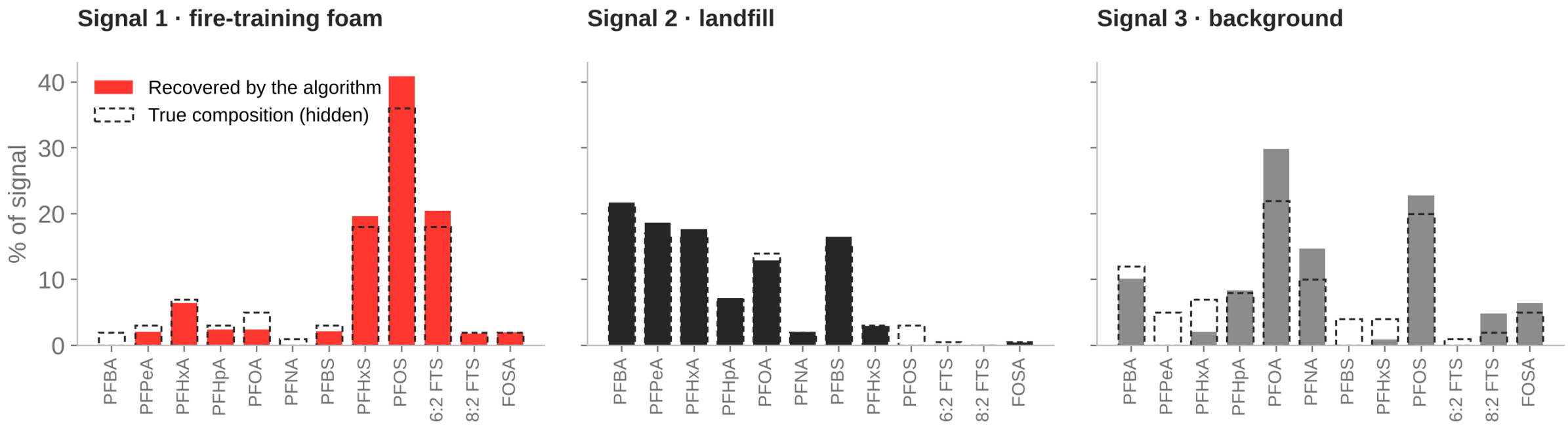
No negative chemistry, and no "source" that removes contamination. Every answer is a mixture that could physically exist.

How many signals?

Chosen the same way as k: unexplained data, chemical sense and the site model, decided before interpreting.

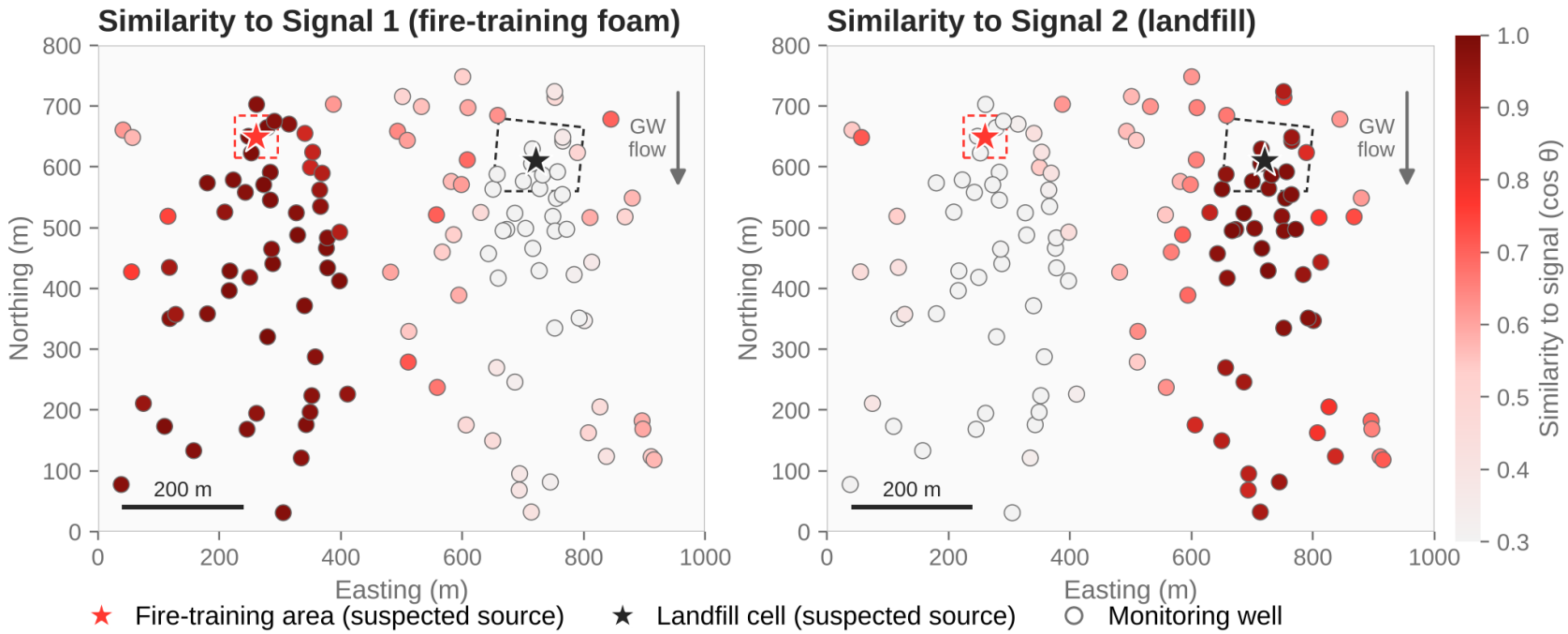
Twelve columns of chemistry become three columns of "how much of each source".

Does it work? A test where the answer is known



Background is the hardest signal to recover because it overlaps with everything.

Score every well against a fingerprint



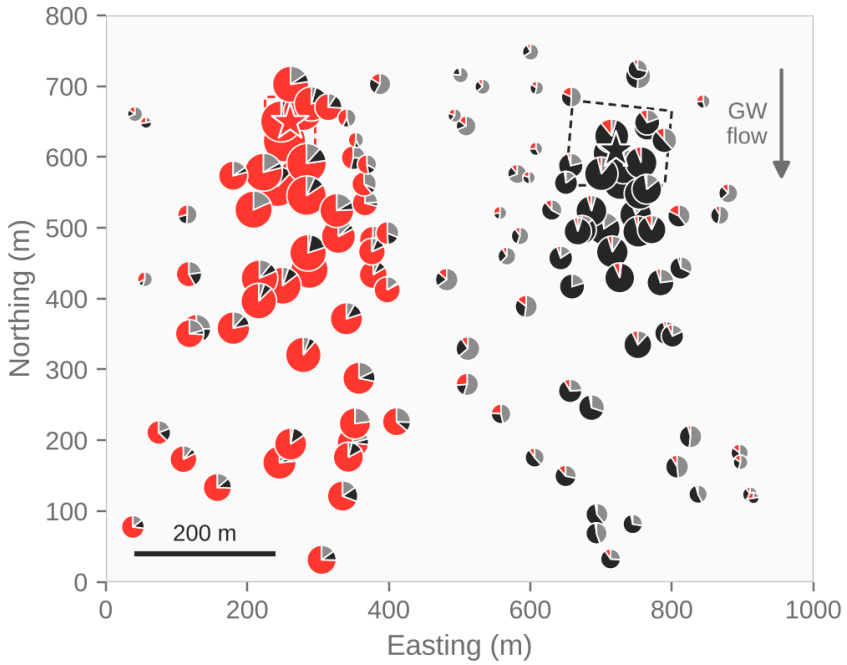
Each circle is a monitoring well. Its colour is cos θ similarity to the signal.

PART 04 OF 06

THREE RESULTS

1. Source fingerprints
2. Data-driven background
3. Management zones

Where each signal went

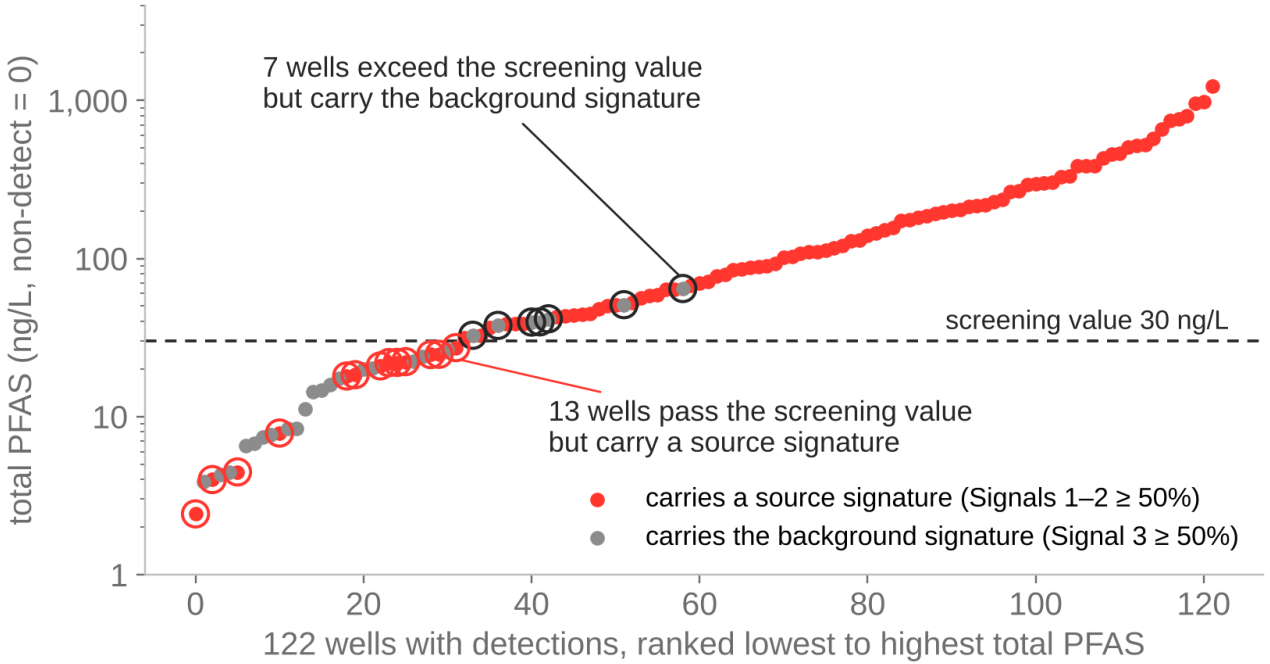


- Signal 1 · fire-training foam
- Signal 2 · landfill
- Signal 3 · background
- small circle = low total PFAS
- large circle = high total PFAS
- Fire-training area
- Landfill cell

- Slices: the mix of signals in each well (composition)
- Size: total PFAS (concentration)
- Where plumes meet, wells show as blends, with no forced either/or

Every slice traces back to measured concentrations. No interpolation, no kriging.

Concentration is not identity



7

wells exceed the screening value but carry the background signature

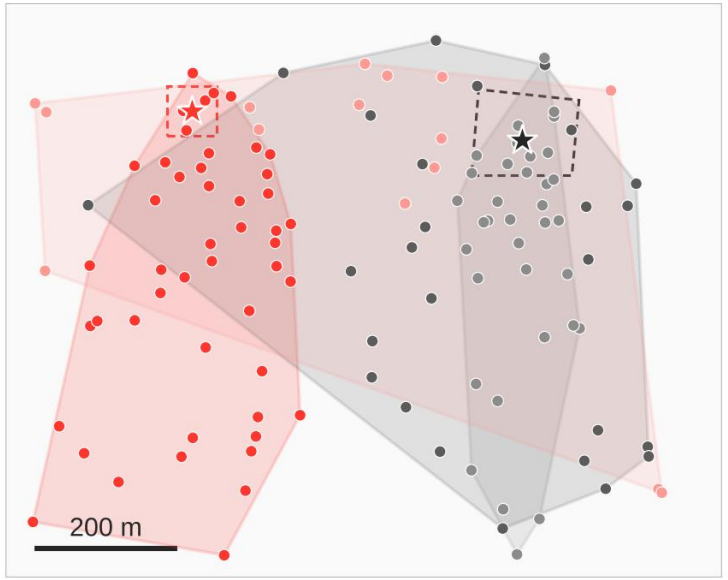
13

wells pass the screening value but carry a source signature

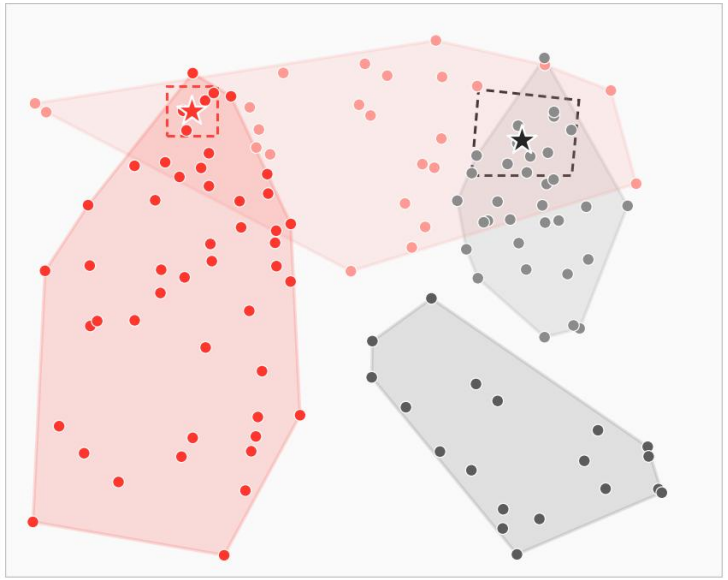
An exceedance is a trigger, not a finding. Composition turns one into the other.

From chemical groups to management zones

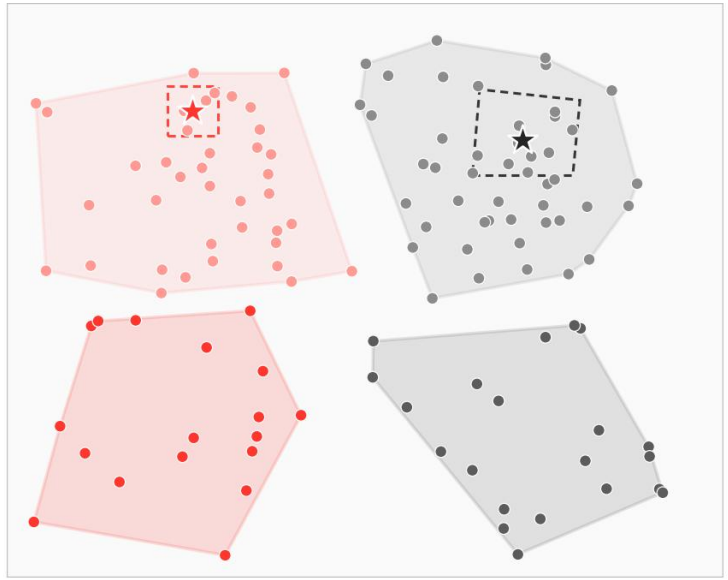
Location weight 0%
chemistry only: zones overlap



Location weight 55%
workable, contiguous zones

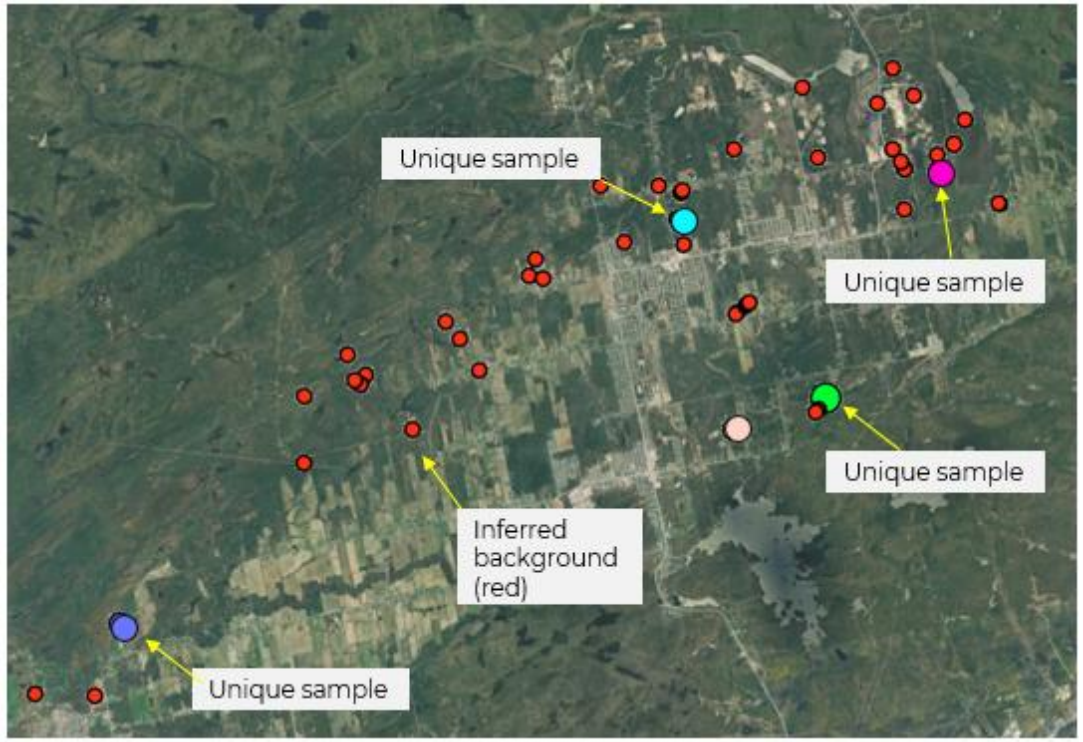


Location weight 85%
geography wins: quadrants



The location weight is a judgement call. Report it, and show what changes when it moves.

On a real site: finding the samples that do not fit

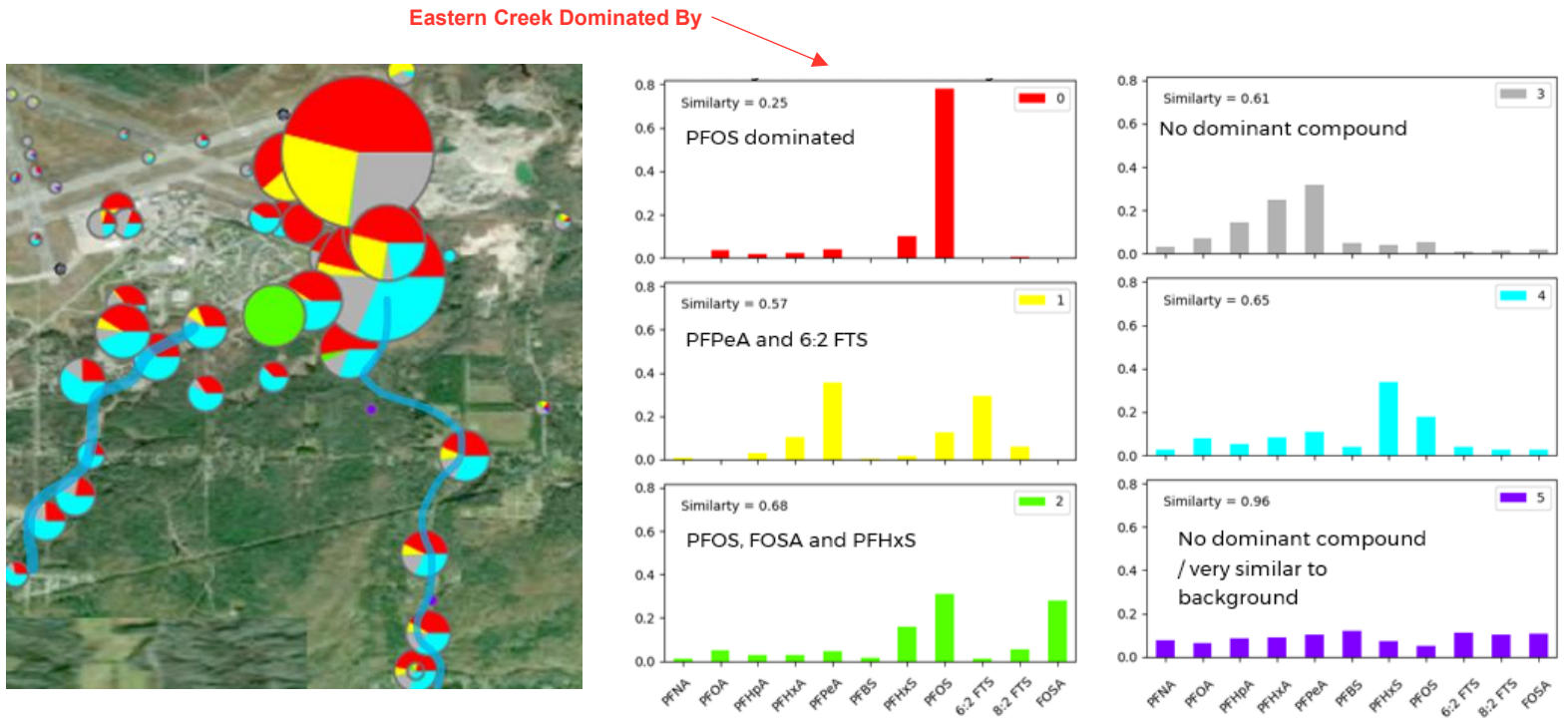


- One round of groundwater sampling
- Most of the site fell into one low-concentration group, interpreted as ambient background
- Five locations would not join any group: each chemically unique
- Triage of 100+ wells took minutes

The tool says "look here". Explaining why is still our job.

WSP project data on an anonymized aerial. Figure courtesy of Paul Hurst (WSP Canada).

On a real site: two creeks, two sources



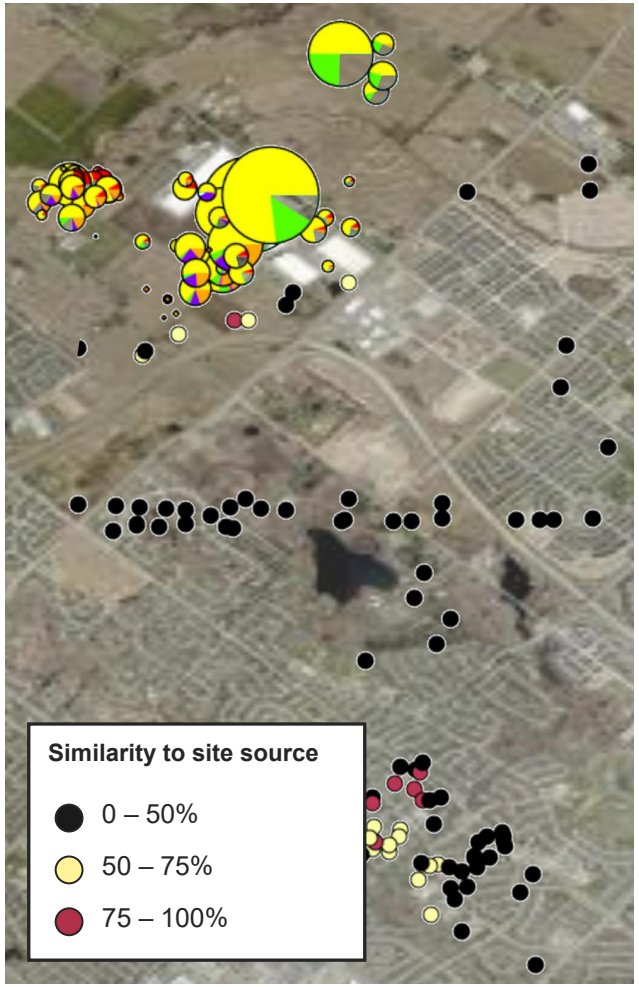
- Surface water in two creeks
- Each creek carried a different signal: two sources, not one plume that split
- Low similarity scores flag samples no signal explains
- Combined with flux, it set which creek to address first

The decision it drove was prioritization, which is what clients pay for.

WSP project data on an anonymized aerial. Figures courtesy of Paul Hurst (WSP Canada) and Skyler Sorsby (WSP USA).

On a real site: ruling a source out

- Drinking water wells, coloured by similarity to one on-site groundwater signal (yellow in the pie chart)
- Every well had PFAS; on concentration alone, all looked like the site's problem
- On composition, the intermediate-depth wells did not match
- The site model explained why



Showing what is not yours is worth as much as showing what is.

WSP project data on an anonymized aerial. Figure courtesy of Paul Hurst (WSP Canada).

None of this is specific to PFAS

PAHs

Combustion (pyrogenic) versus petroleum (petrogenic) patterns, across every compound at once

Metals

Natural weathering of bedrock versus deposition from industry

PCBs, dioxins and furans

Congener and homologue patterns are exactly this problem

Where it does not work

Single-compound sites, and mixtures that transform quickly as they move

Any contaminant that arrives as a mixture and holds its pattern is a candidate.

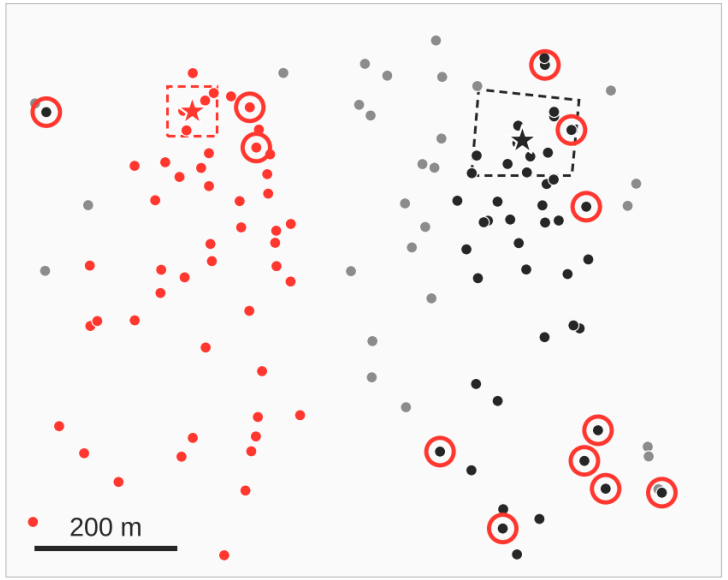
PART 05 OF 06

KEEPING IT DEFENSIBLE

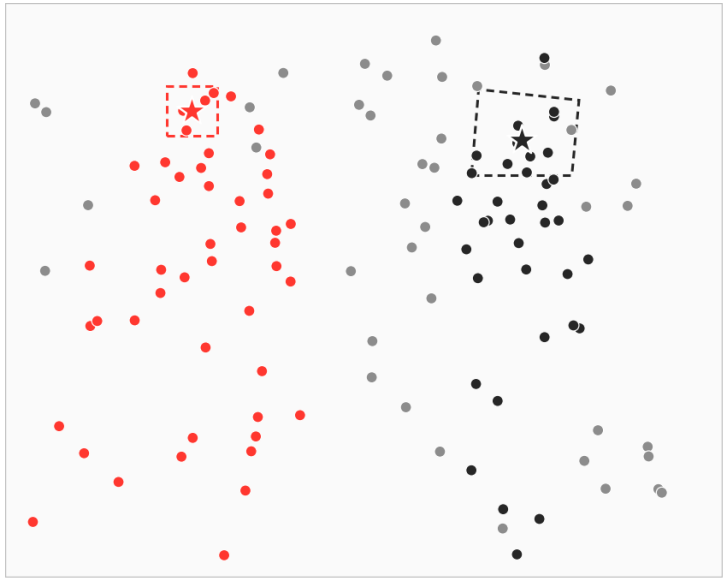
No answer key means no accuracy score

One reporting choice, different answers

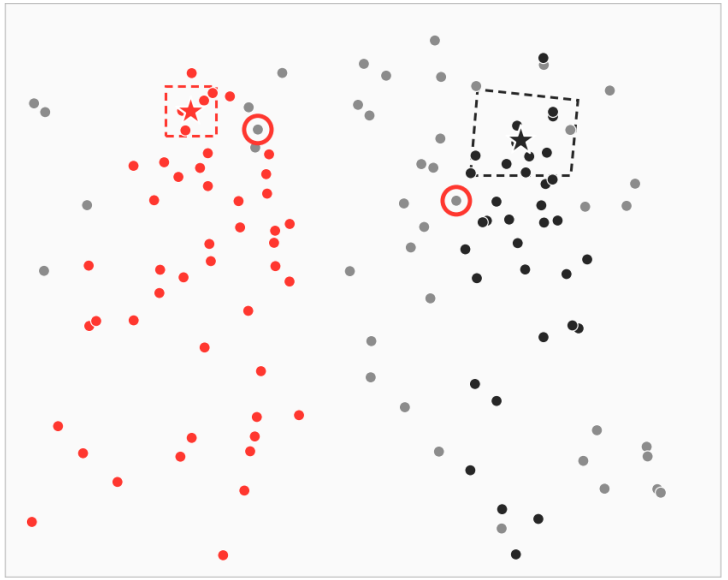
non-detect = 0
12 of 124 wells change group



non-detect = ½ reporting limit (base case)



non-detect = reporting limit
2 of 124 wells change group



There is no universally correct choice. State the one you made, and test the alternative.

Five habits that keep it defensible

1	**Decide before you interpret**	Set k, the number of signals and the non-detect rule from the diagnostics and the site model.
2	**Show the alternative**	Re-run at k ± 1 and with another non-detect rule. If the story holds, say so. If it does not, that is the finding.
3	**The site model has the final say**	The algorithm has never seen a borehole log. Hydrogeology and site history can overrule the output, never the reverse.
4	**A signal is a hypothesis, not a source**	Calling Signal 2 "the landfill" is an interpretation you support with other evidence.
5	**Garbage in, confident garbage out**	These methods never return an error message. QA/QC the dataset before you analyze it.

PART 06 OF 06

GETTING STARTED

Where to begin, and questions

What this looks like in practice

1

You already have the input

A standard laboratory or database export, samples down and analytes across, is exactly what these methods use. No new field program.

2

The methods are open

k-means and non-negative matrix factorization are published methods in free tools such as Python (scikit-learn) and R. The U.S. EPA PMF model uses the same family for source apportionment.

3

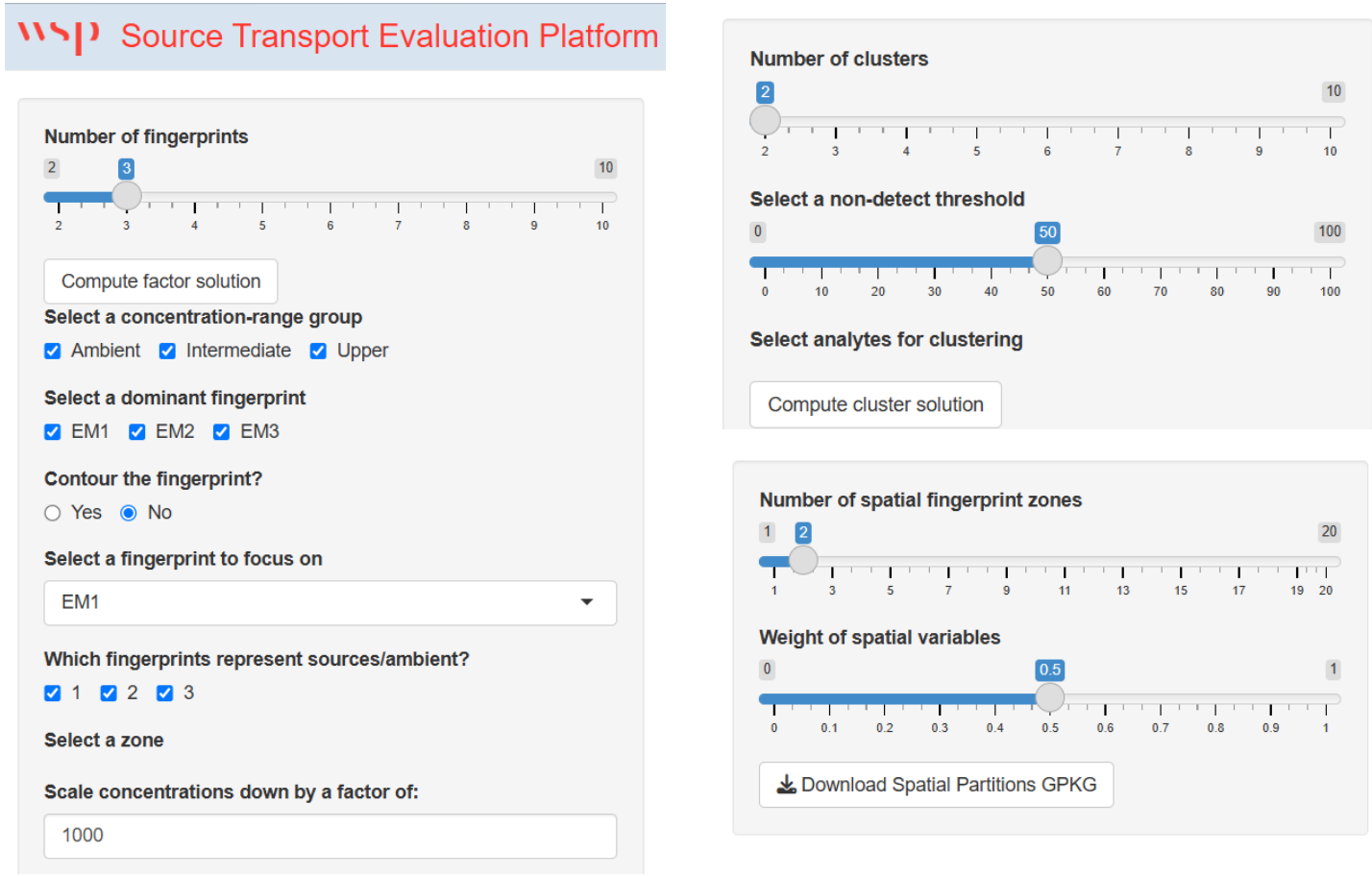
Calibrate on a site you know

Run it where you have already interpreted the data. Agreement builds trust; disagreement shows you where to look.

At WSP, our implementation is STEP (Source Transport Evaluation Platform), a web-based R Shiny application developed by Skyler Sorsby (WSP USA), alongside Python notebooks used by Paul Hurst (WSP Canada).

None of this is proprietary. The principles are available to everyone in this room.

WSP Source Transport Evaluation Platform (STEP)



- Tool allows instant scaling, sorting, and visualization of existing data sets
- Hypotheses can be tested and scaled as appropriate for applicable site and practitioner
- Allows clear validation of inputs and results can be re-evaluated with minimal computation time

These tools do not replace judgement.

They change which questions you can afford to ask.

THANK YOU

Jaden Ford, P.Eng., PMP
WSP Canada · Kelowna, BC
jaden.ford@wsp.com

Paul Hurst, M.Sc., P.Eng.
WSP Canada · methods and project examples
paul.hurst@wsp.com

Skyler Sorsby
WSP USA · developer of STEP
skyler.sorsby@wsp.com

