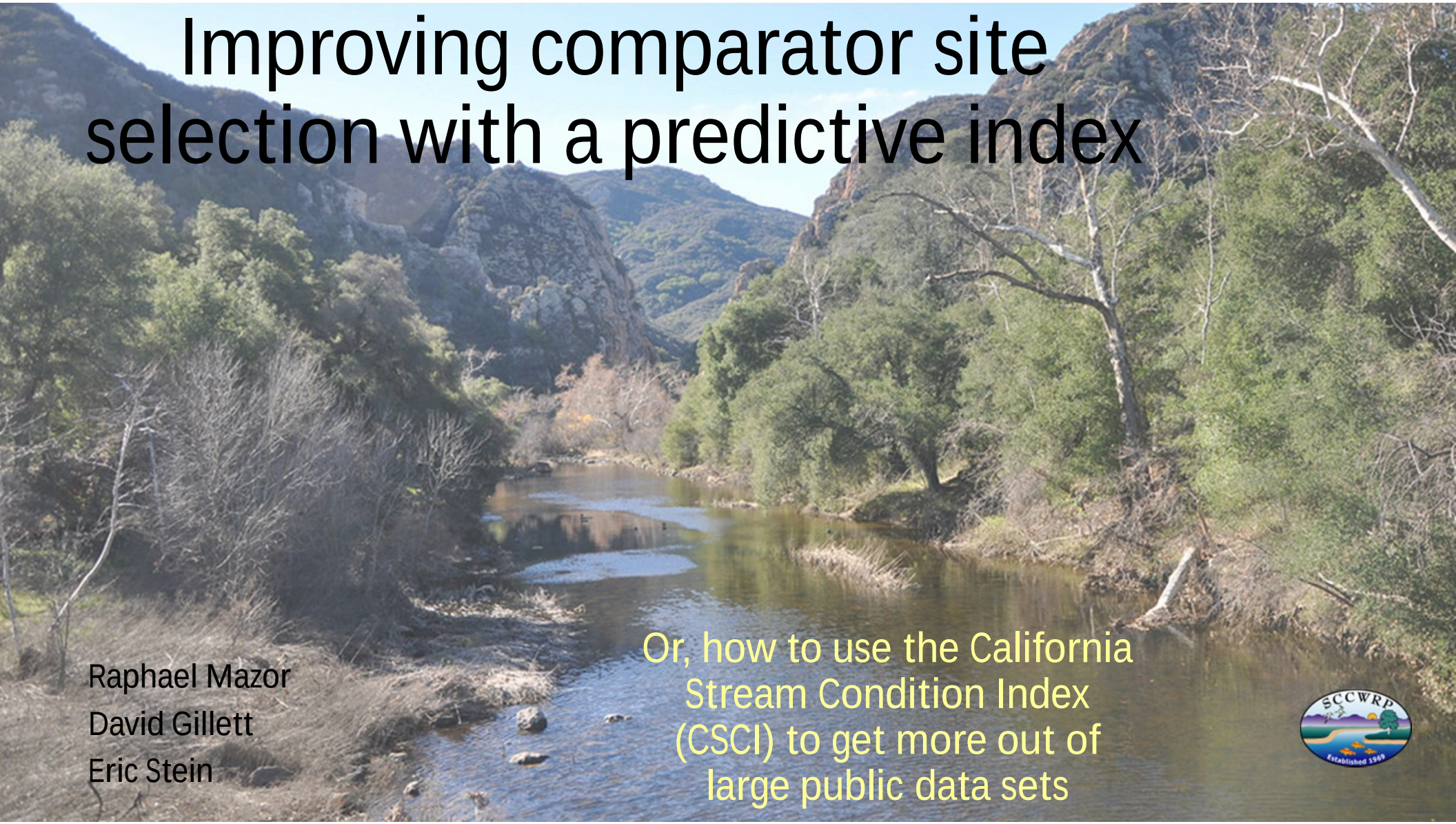
Improving comparator site selection with a predictive index



Raphael Mazor
David Gillett
Eric Stein

Or, how to use the California Stream Condition Index (CSCI) to get more out of large public data sets

A vision for causal assessment

- Routine coarse-scale implementation at many sites
  - Low cost
  - Easy and automated
  - Produce coarse-scale results about stressor classes (e.g., pesticides vs. metals)
  - Determines if more work is necessary
- Detailed, fine-scale implementation at strategically selected sites
  - Similar to EPA's "CADDIS"
  - Larger investment
  - Analyses more specific to each site
  - Produce results about specific stressors (e.g., bifenthrin vs. DDT)

Causal assessment needs comparator sites

Finding useful comparator sites is one of the biggest challenges for causal assessment.

We compare the biological conditions and stressors observed at a degraded “test” site to similar comparator sites that:

1. Could support similar biology under unstressed conditions
2. Represent a range of biological conditions (ideally, includes “healthy” or reference-condition sites)
3. Represent a range of stress levels

The California Stream Condition Index (CSCI) can make the task a lot easier

The CSCI is a predictive index that measures stream health as deviations between observed biology, and biology expected under natural conditions.

CSCI directly addresses **two** of the three site-selection criteria:

✔ 1. Could support similar biology under unstressed conditions

✔ 2. Represent a range of biological conditions

✘ 3. Represent a range of stress levels

(You need non-biological data for #3)

How are comparator sites used?

Site	Biological Condition	Stressor A (relative to test)	Stressor B (relative to test)
Test site	Poor	<>	<>
Comparator 1	Fair	-4	-3
Comparator 2	Poor	-3	+5

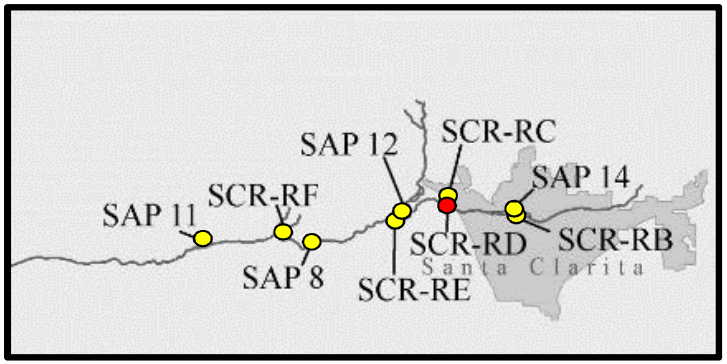
What do the comparators provide?

- Strengthened support for Stressor B
- Weakened support for Stressor A

How to pick comparator sites?

Traditionally, a BPJ exercise (e.g., upstream/downstream). Local experts hand-pick the sites they think are best suited to the analyses.

- Often oriented around point-source problems
- EPA's approach assumed very little data available
- Assume no access to data outside immediate area/watershed
- Sites selected because of proximity, presumed environmental similarity

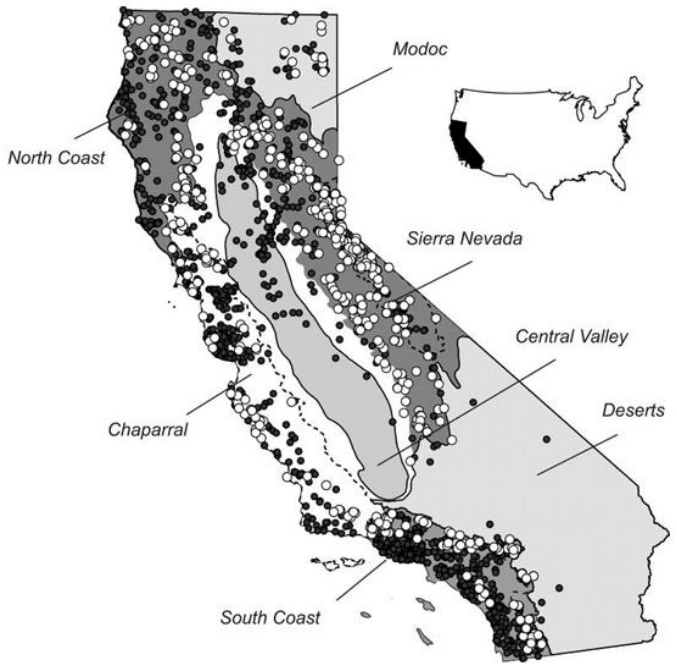


Permit-required monitoring sites for sanitary discharge into the Santa Clara river

California has a wealth of data from potential comparators

Thanks to many large-scale programs….

- Perennial Stream Assessment
- SMC Stream Survey
- Bay Area RMC
- RCMP



Which are the best comparators for your test site?
Hundreds to choose from….

The CSCI calculates expected biology

<u>Environmental variables</u>

Latitude

Elevation

Precipitation

Temperature

Watershed area

<u>Capture probabilities</u>

Acari	0.89
Baetis	0.94
Calineuria	0.35
Epeorus	0.72
Hydroptila	0.66
Chironominae	0.12

You can then calculate expected similarity

Bug	Site1	Site2	Site3
Acari	0.88	0.92	0.89
Baeties	0.91	0.25	0.83
Calineuria	0.53	0.56	0.62
Epeorus	0.10	0.97	0.48
Hydroptila	0.28	0.65	0.07
Chironominae	0.40	0.32	0.98

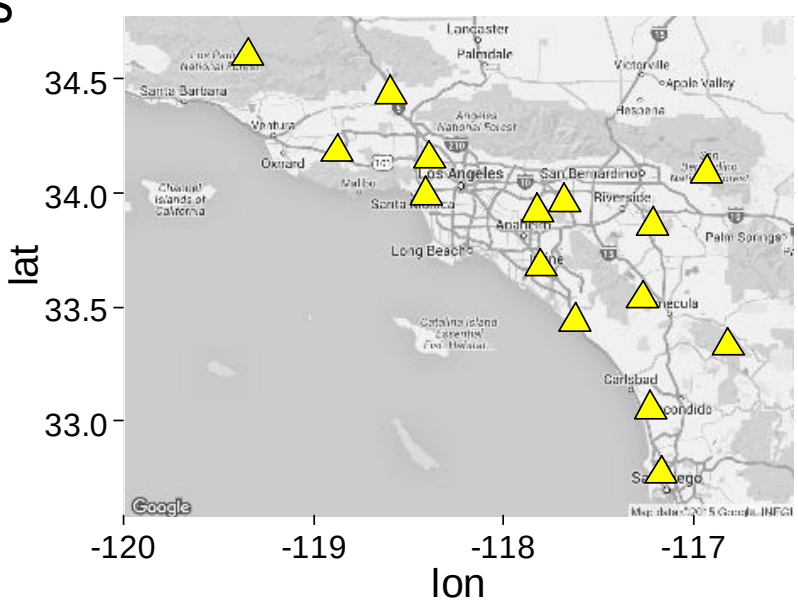
Euclidean

Distance

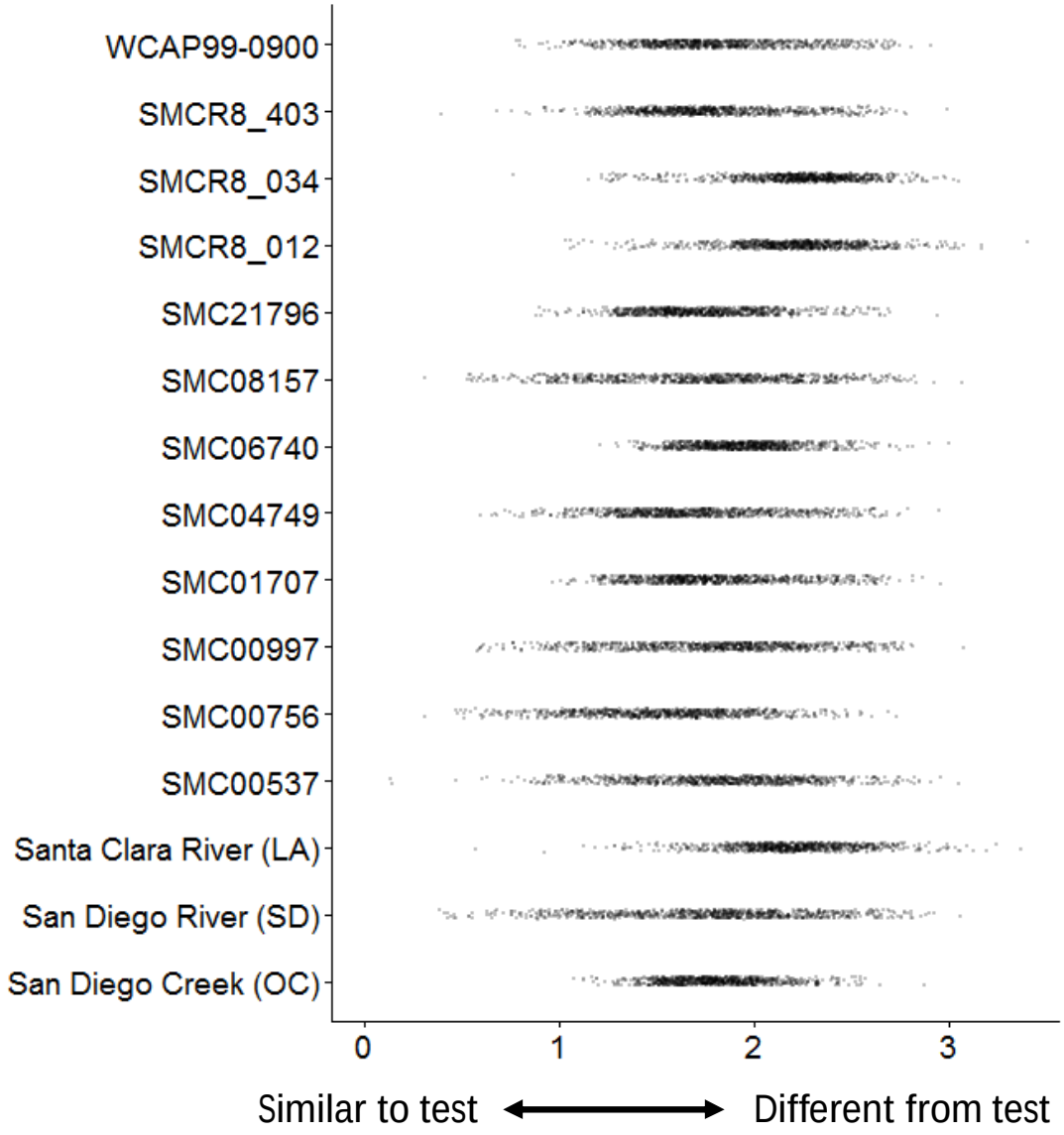
	Site1	Site2	Site3
Site1	0	2.16	1.81
Site2	2.16	0	1.97
Site3	1.81	1.97	0

Demonstration: Selecting comparators from SoCal Stream Survey data

- SMC, EMAP, PSA: 700+ potential comparators
- 3 test sites from previous case studies
  - Santa Clara River (LA Co)
  - San Diego River (SD Co)
  - San Diego Creek (Orange Co)
- Also hand-pick 12 low-scoring test sites from across the region
- Calculate similarity to all potential comparators

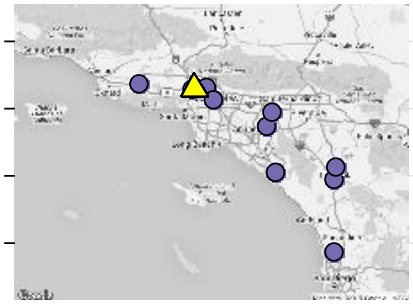


Each test site poses a unique set of challenges

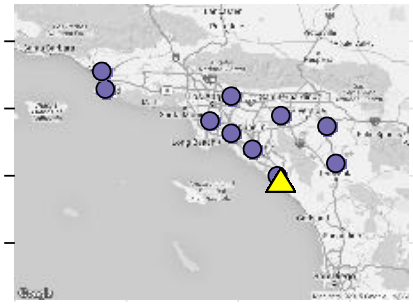


The best comparators aren’t always right next door

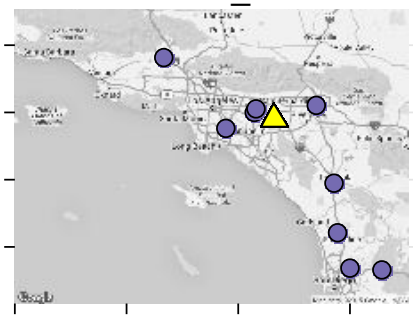
SMC00756



SMC00997



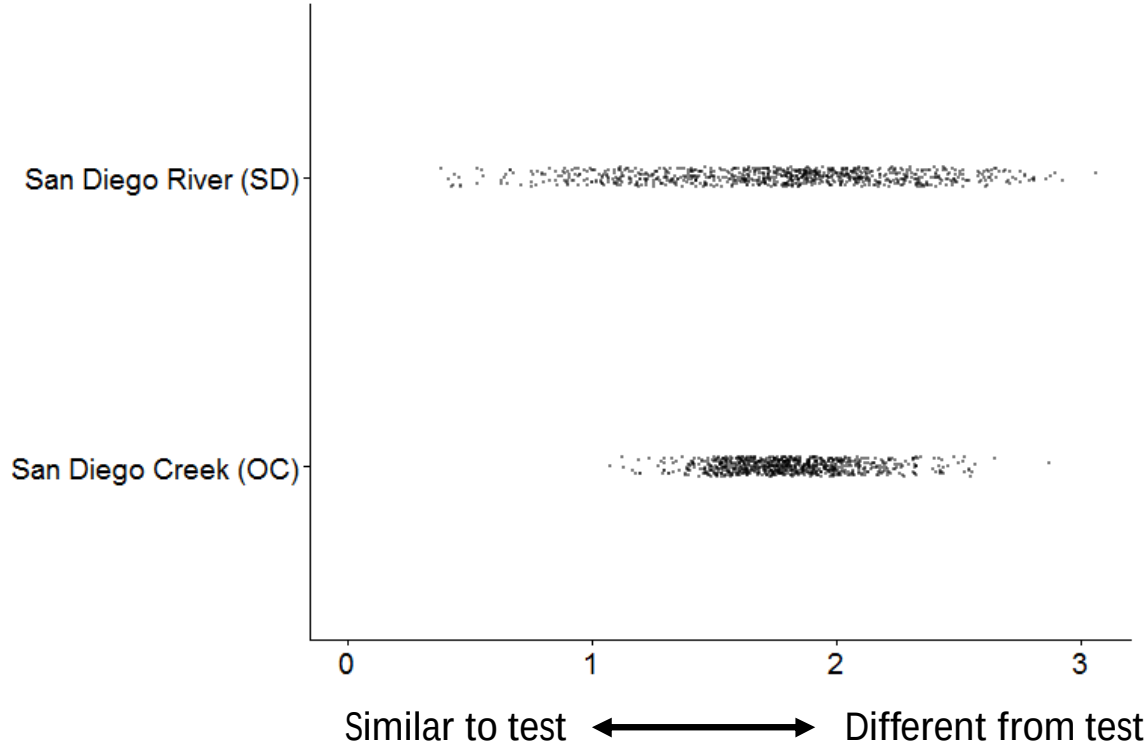
SMCR8_403



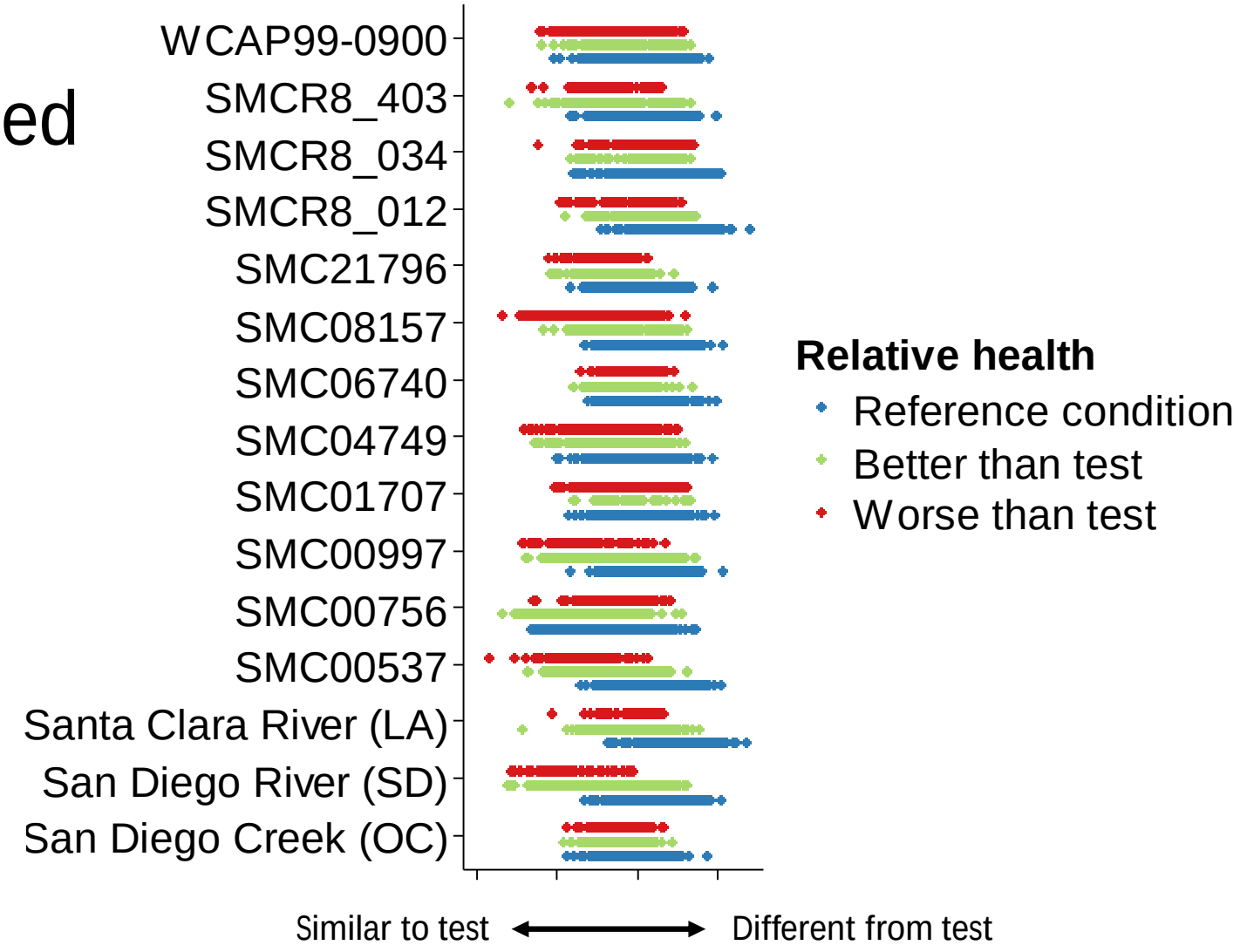
Some sites had few similar comparators

Many similar comparators

Few similar comparators

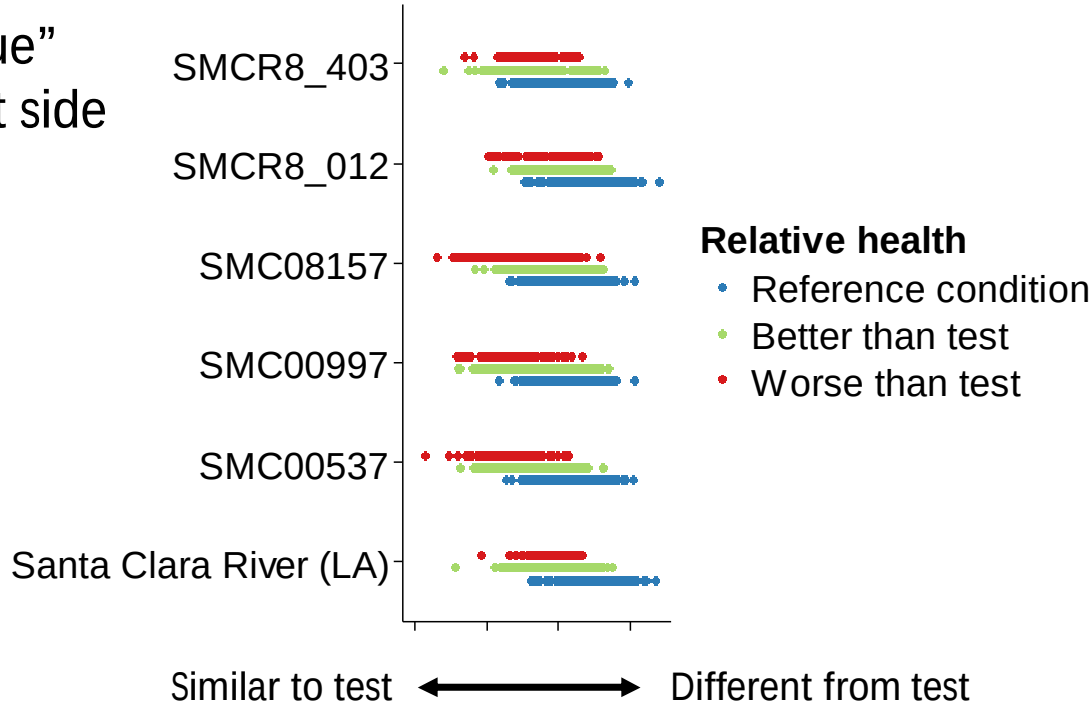


Comparators need to represent a range of conditions



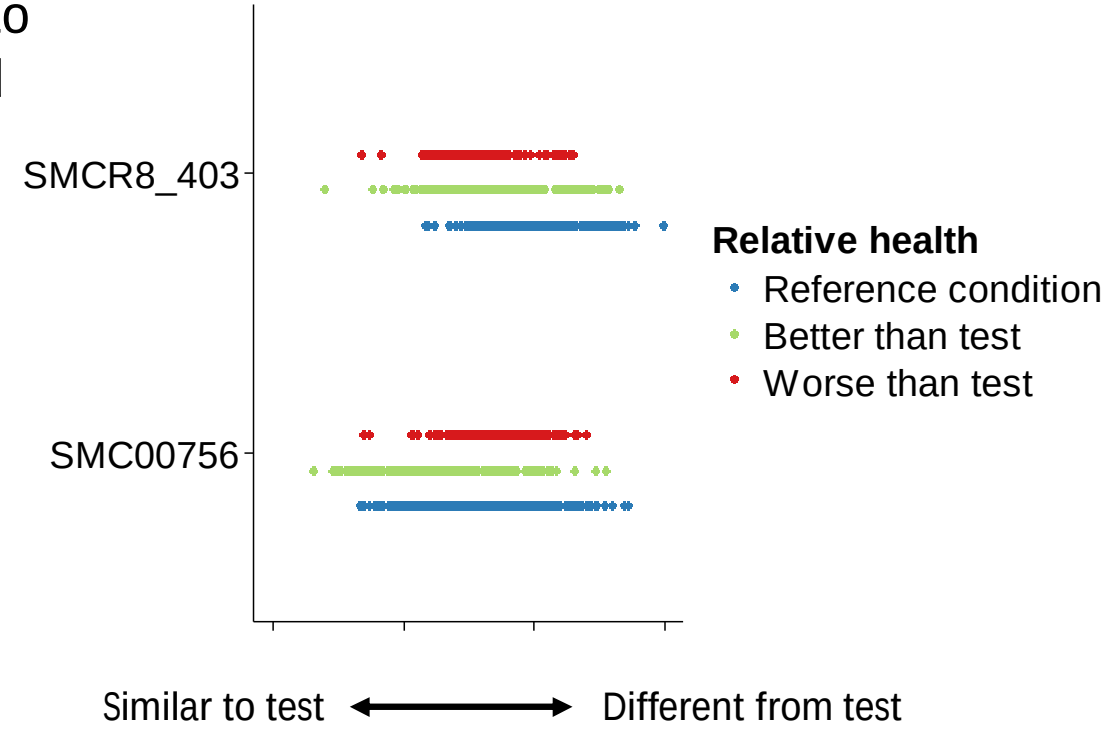
Few comparators meet objective for some sites

Hardly any “blue” dots on the left side of plot.



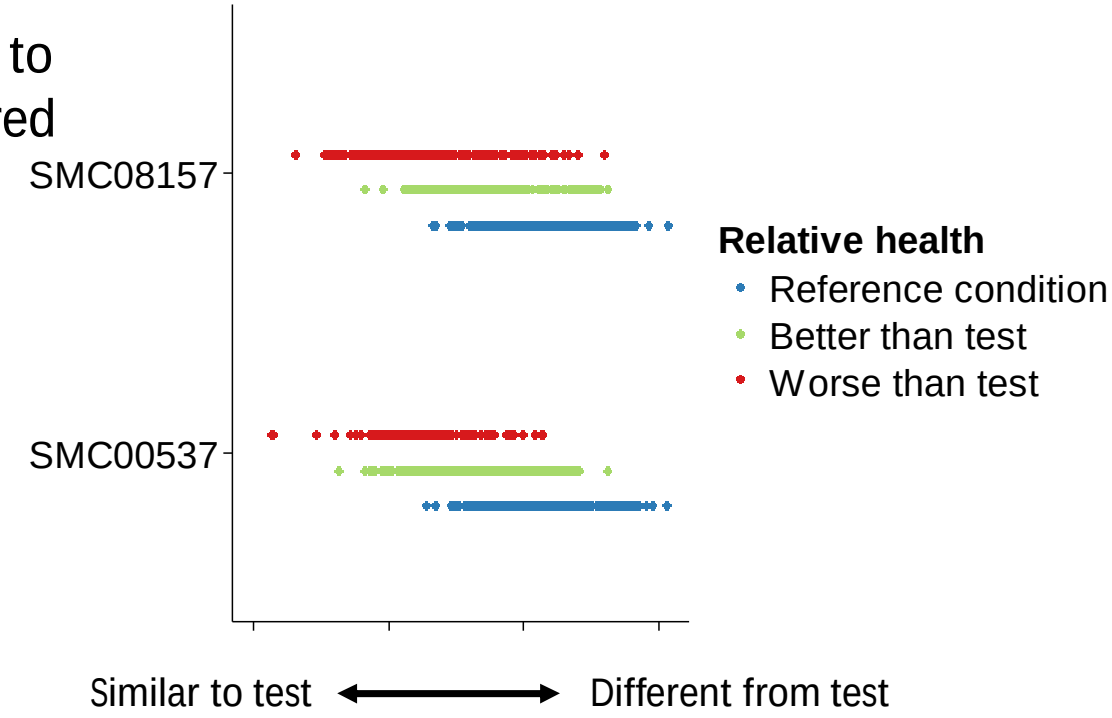
Some test sites are the worst of their closest peers

Green dots start to the left of the red dots



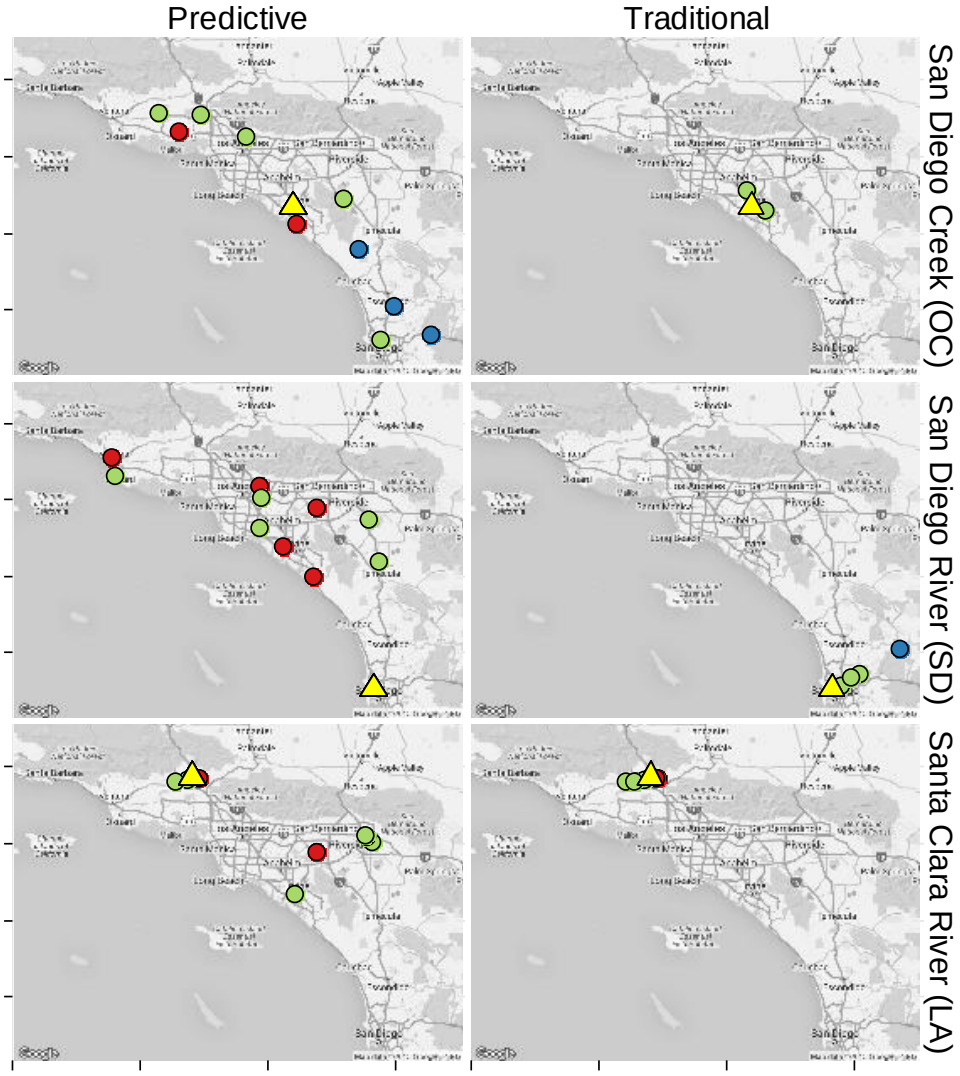
Test sites may actually in better condition than their closest peers

Green dots start to the right of the red dots



Predictive versus traditional approach

- Very little overlap between methods.
- Traditional approach is geographically constrained
- Predictive approach not a substitute, but a better starting point.
- Local expertise, familiarity with sites is still necessary



Benefits of this approach

- Helps you look outside your watershed
- Direct link to biological endpoints with management relevance (e.g., bio-objectives)
- Based on quantitative measures of biological similarity, not on assumptions or proximity to test site
- Support more data-intensive analyses (e.g., relative risk) than traditionally used in causal assessment
- Other applications:
  - Supports study design by pre-screening of comparators (e.g., permit requirements)
  - Evaluate similarity of mitigation sites

Next steps

- Explore tradeoffs between more sites versus degree of similarity
- Develop interface for CEDEN/SWAMP databases
- Create coarse screening tools for routine causal assessments
- Create standard data queries that support detailed casual assessments


Fin

How might causal assessment change with a predictive approach?

- Easier access to bigger data sets.
- More sophisticated analyses, with stronger, more nuanced results.