# Satellite-based Machine Learning Estimates of Freshwater Cyanobacterial Harmful Algal Blooms and Links to Incident Dementia in California

EDGE Symposium
June 10, 2026

UW EDGE ENVIRONMENTAL & OCCUPATIONAL HEALTH SCIENCES

# Background

- The frequency and geographic extent of cyanobacterial harmful algal blooms (cyanoHABs) is increasing
- cyanoHABs can produce neurotoxins, which may play a role in the risk of dementia
- Existing methods for quantifying the spatiotemporal extent of cyanoHABs are limited by gaps in coverage of common satellites (2012 – 2016 missing)

# Data & Methods

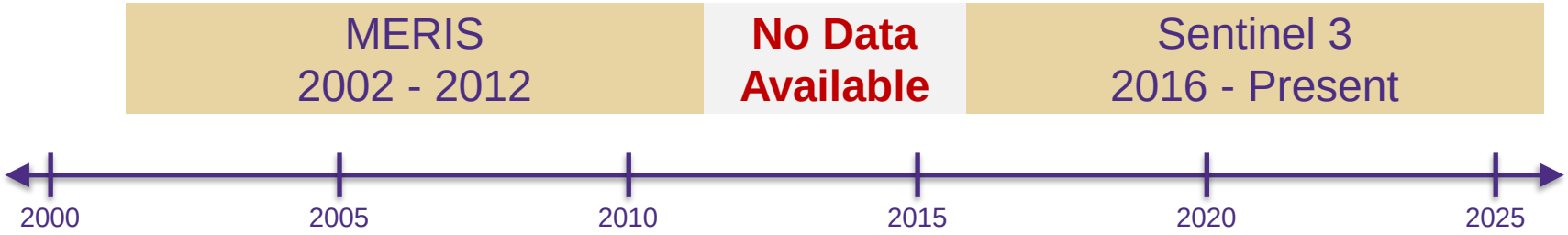
## Outcome data (incident dementia identified with ICD codes

**Table.** Characteristics of KPSC beneficiaries, 2008–2023.

| | |
|---|---|
| **Total** – N (%) | 1,564,087 (100.0%) |
| **Dementia diagnosis** – N (%) | 107,844 (6.9%) |
| **Age at Cohort Entry, Years** – Median (IQR) | 61 (60, 68) |
| **Sex** – N (%) | |
| Male | 737,782 (47%) |
| Female | 826,305 (53%) |
| **Race and Ethnicity** – N (%) | |
| Hispanic | 439,839 (28%) |
| Non-Hispanic | |
| Asian | 168,980 (11%) |
| Black | 143,769 (9.2%) |
| Other[a] | 83,015 (5.3%) |
| White | 728,484 (47%) |
| **Marital Status** – N (%) | |
| Married or partnered | 848,238 (54%) |
| Divorced or separated | 147,297 (9.4%) |
| Other or unknown | 181,736 (12%) |
| Single | 205,066 (13%) |
| Widowed | 181,750 (12%) |

# Satellite Data Availability

Existing exposure assessments (ex: NASA CyAN):

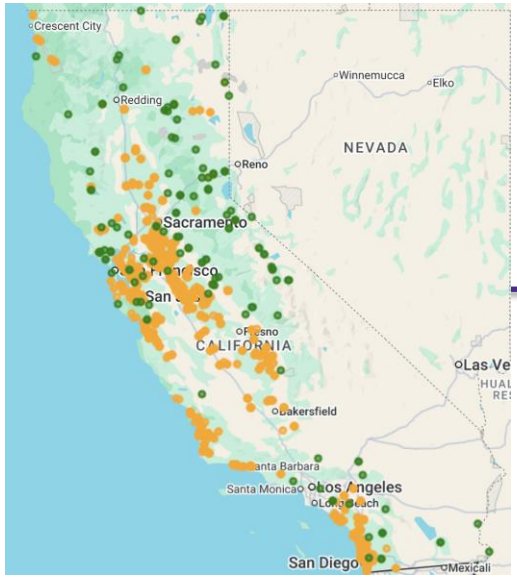



Our exposure assessment:

# Data & Methods

## Exposure assessment (cyanoHABs)

**Training data:**




Cyanobacteria Aggregated Manual Labels (CAML)

EPA National Lakes Assessment

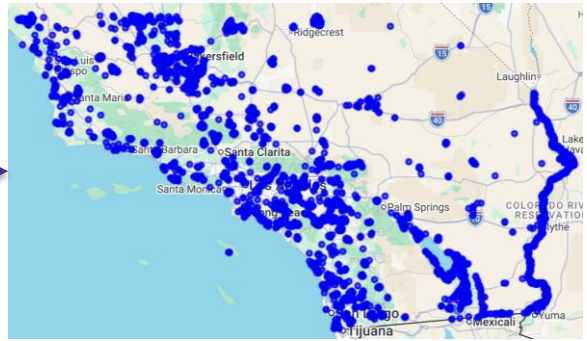
**Machine Learning model:**

Two models trained:

- Primary - binary outcomes in CA
- Secondary - continuous abundance in CONUS

Features from satellite-based water color and meteorology

**Exposure grid:**




**Bi-weekly estimates over 250m grid of freshwater points**

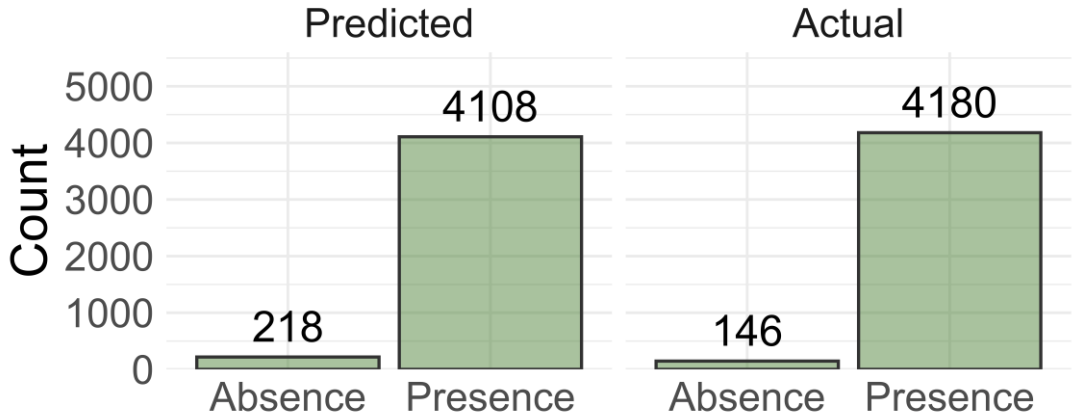
# Model performance

**Primary model**

Binary outcome, trained on Califc
only:

| | |
|---|---|
| Balanced Accuracy* | 0.94 |
| AUC | 0.97 |

*(sensitivity/specificity)



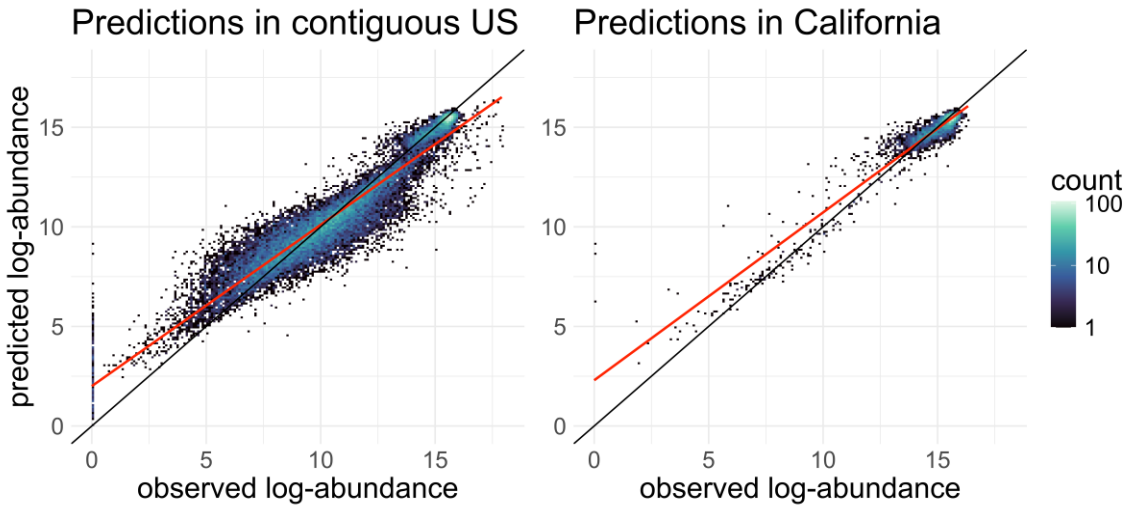
# Model performance

## Secondary model

Continuous outcome, trained on contiguous US:

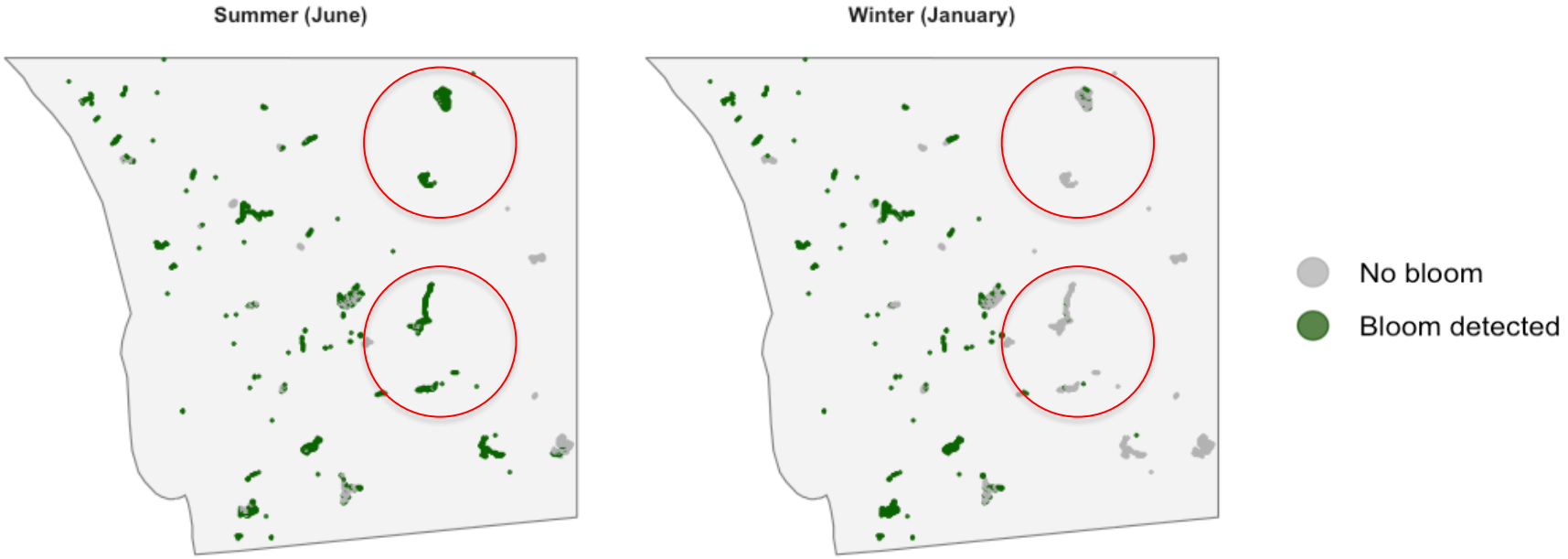
| | |
|---|---|
| $R^2$ | 0.69 |
| RMSE | 0.96 |
| Balanced Accuracy* | 0.88 |
| AUC* | 0.98 |

*abundance converted to binary predictor, where abundance > 20,000 cells/mL = presence of HAB



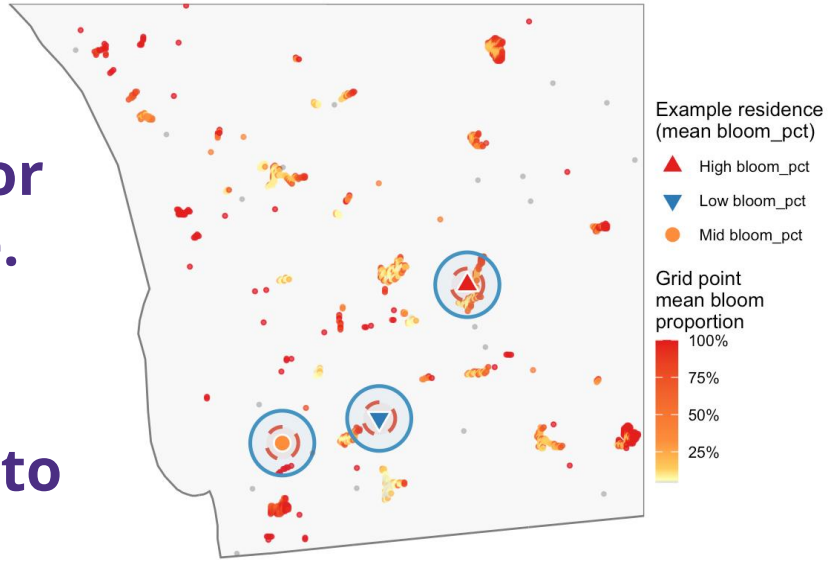
# Exposure Assessment

## Example HAB predictions (San Diego County):



# Conclusions & next steps

- Our exposure assessment approach may provide a scalable, quantitative metric for assessing cyanoHABs exposure.
- Next, we will use the exposure data to assess associations between residential proximity to algal blooms and incident dementia in the KPSC cohort.



# Thank you!


- A huge thank you to the NIEHS EDGE Center for funding this work
- This work was also supported by NIA R01AG071024, the BEBTEH NIEHS T32, and the Robert W. Day Endowed Professorship in Public Health