



# CyFi

*Machine learning for harmful algal bloom detection*

**DRIVEN** DATA

Harmful algal blooms (HABs) are a **common threat to marine and human health**.

Existing automated detection tools focus on ocean and coastal areas. But blooms in smaller inland water bodies are still monitored manually, **which is very time intensive**.



**CyFi** (Cyanobacteria Finder) is an open-source Python package that uses satellite imagery and machine learning to detect cyanobacteria levels, one type of HAB.

CyFi can help decision makers protect the public by **flagging the highest-risk areas in lakes, reservoirs, and rivers quickly and easily.**

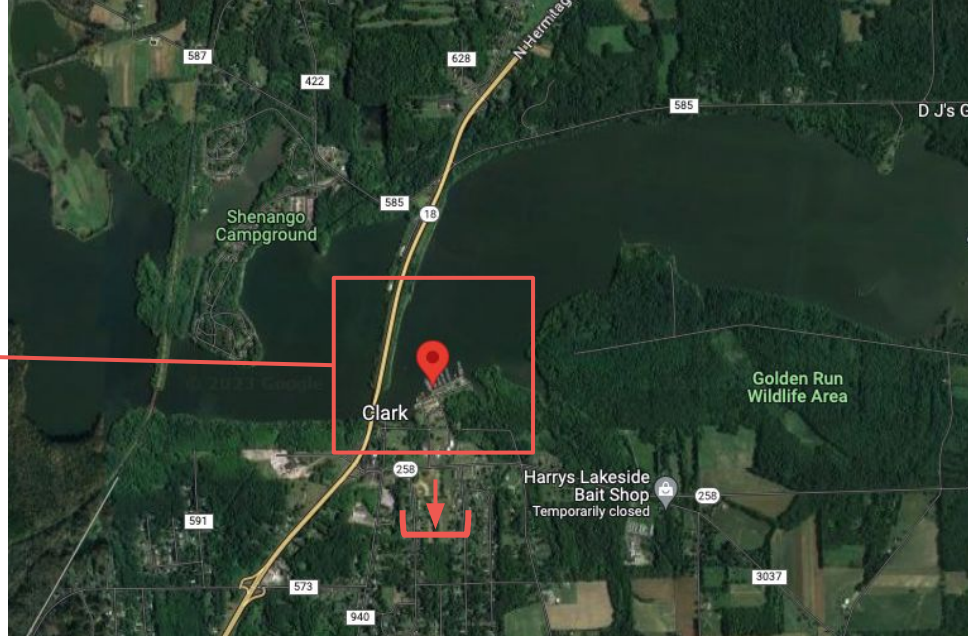




CyFi uses **high-resolution Sentinel-2 satellite imagery** (10-30m) to focus on smaller water bodies with rapidly changing blooms.

**Sentinel-3 is used by most existing tools**, but its resolution of 300-500m is often too coarse for small, inland water bodies.

	date	latitude	longitude
0	2023-06-27	41.287577	-80.424543
1	2023-07-10	35.650000	-78.682816
2	2023-08-31	35.705416	-79.164659
3	2023-09-22	37.564318	-101.335575



Generate estimates for many points at once with a simple CSV of **dates and locations!**

CyFi searches for and downloads publicly available **satellite imagery** around each point, which is passed into a machine learning model.

sample_id	Sample information			Predicted cyanobacteria density (cells/ml)	Severity level
	date	latitude	longitude	density_cells_per_ml	severity
<b>89e12c14b5a131b82e9738932a7fa9c8</b>	2023-06-27	41.287577	-80.424543	57,433	moderate
<b>087d604d9d8568761513d26a47c94bc8</b>	2023-07-10	35.650000	-78.682816	83,609	moderate
<b>a0517780fa24874ebf166aefa17a0c1b</b>	2023-08-31	35.705416	-79.164659	5,733	low
<b>cde656c081bfe8fa99c7c8b20ff547f7</b>	2023-09-22	37.564318	-101.335575	3,684,003	high

Cyanobacteria estimates are saved out as a CSV that can be plugged into any existing decision-making process.

For each point, the model provides a severity level based on World Health Organization (WHO) guidelines and an estimated density in cells per mL for detailed analysis.

Severity	Density range (cells/mL)
Low	0 - 20,000
Moderate	20,000 - 100,000
High	Over 100,000

Simply run one  
line of code to  
generate  
predictions

```
$ cyfi predict list_of_points.csv
```

```
SUCCESS | Loaded 5 sample points (unique combinations of date, latitude, and longitude) for prediction  
SUCCESS | Downloaded satellite imagery  
SUCCESS | Cyanobacteria estimates for 4 sample points saved to preds.csv
```

Or estimate  
cyanobacteria for  
a single point  
rather than  
providing a file

```
$ cyfi predict-point --lat 35.6 --lon -78.7 --date 2023-09-25
```

```
SUCCESS | Estimate generated:  
date          2023-09-25  
latitude      35.6  
longitude     -78.7  
density_cells_per_ml  22,836  
severity      moderate
```



Launch the CyFi Explorer to view cyanobacteria estimates alongside Sentinel-2 imagery!

### CyFi estimates

sample_id ▲	date ▲	latitude ▲	longitude ▲	density_cells_per_ml ▲	severity ▲
6be1f8ed407e0ec7ab0c9a42394d9d44	2023-08-24	38.32629	-119.21121	7957	low
c485b9c41484d4d0b82b8580a215a43c	2023-08-23	34.24757	-117.2664	9234	low
3935648294a71be0197814c37de2f9a8	2023-08-23	38.466885	-123.01219	16141	low
389fee8dbca6759f0588dc842396c6b6	2023-08-22	37.7726963	-119.08373	17313	low
8b31451562d5ebd26cfa2cc5eb7357cc	2023-08-22	37.822007	-119.11876	17113	low

### Sentinel-2 Imagery



### Details on the selected sample



Estimated cyanobacteria density (cells/ml)

16141

Estimated severity level

low

Location

(-123.01219, 38.466885)

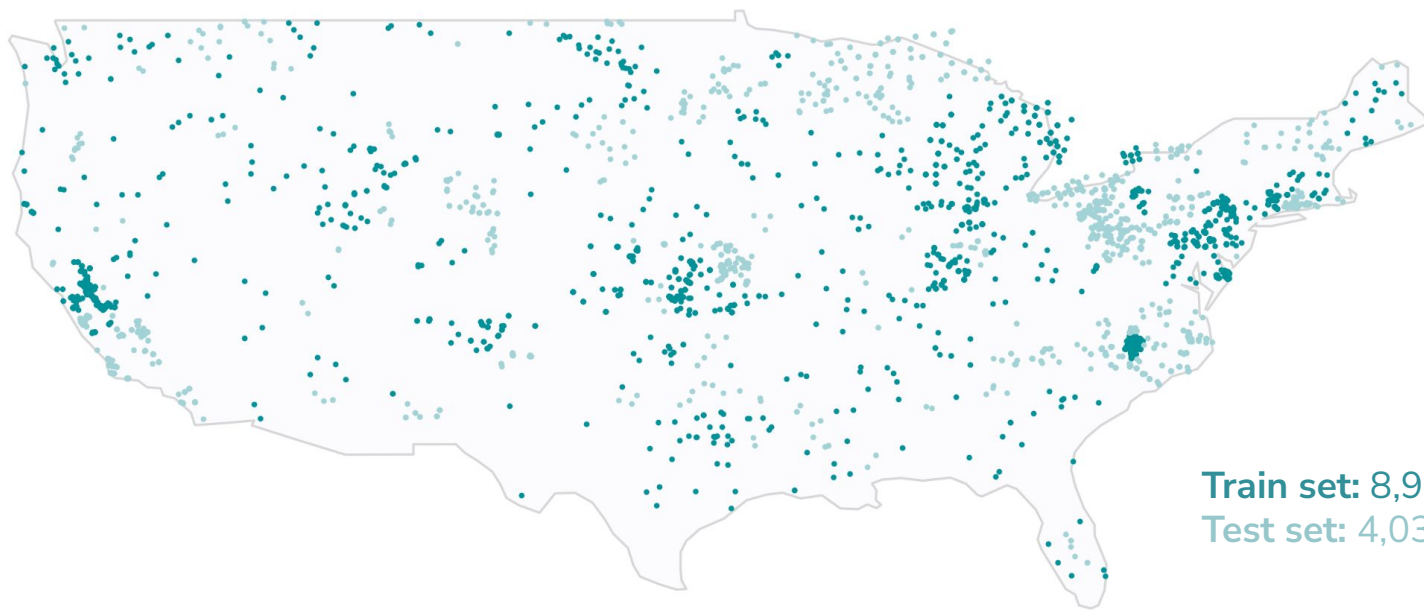
Sampling date

2023-08-23

Satellite imagery date

2023-08-12

CyFi was trained and evaluated using in-situ measurements of cyanobacteria density from across the U.S.



**Train set:** 8,979 observations  
**Test set:** 4,035 observations

## CyFi is most accurate at low and high cyanobacteria densities



### **Low severity**

Better allocate ground sampling resources by deprioritizing water bodies where blooms are likely absent



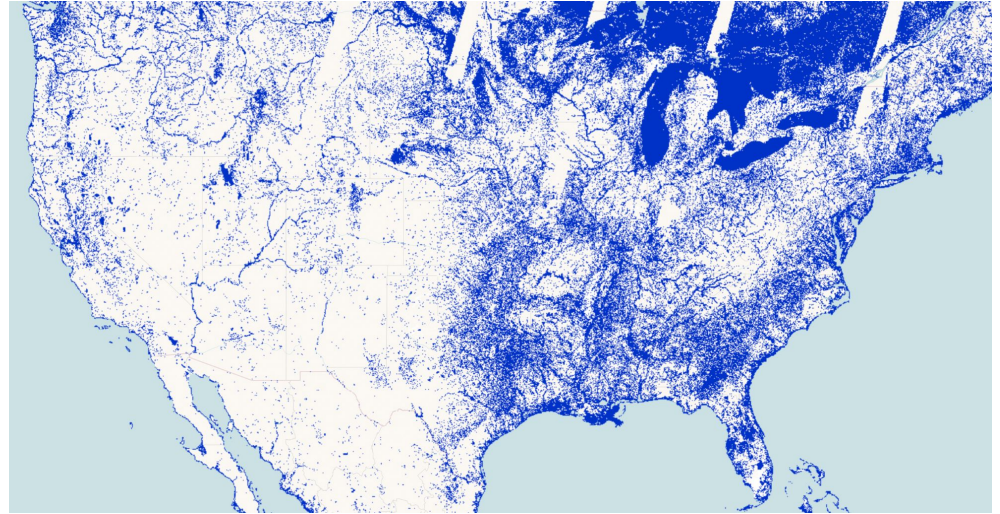
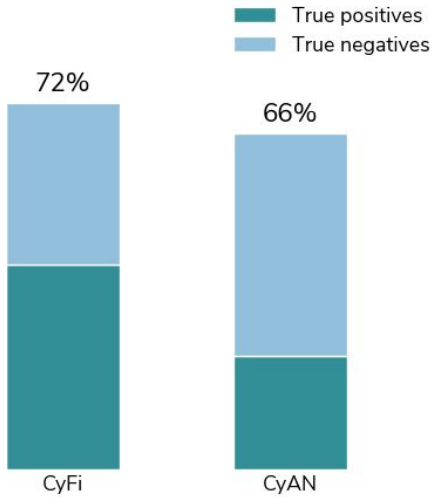
### **High severity**

Support public health interventions by flagging water bodies where severe blooms are likely present

CyFi performs at least as well as Sentinel-3 based tools

And has **10x greater coverage of lakes** across the U.S. thanks to Sentinel-2 imagery!

### Bloom detection accuracy



A true positive (bloom presence) is where cyanobacteria density > 10,000 cells/mL. Uses a dataset of 756 ground measurement observations from across the U.S.

Water bodies detected by Sentinel-2 across the U.S. Source: [Global Water Bodies Product](#)

CyFi makes it simple for water quality managers to take advantage of state-of-the-art machine learning.

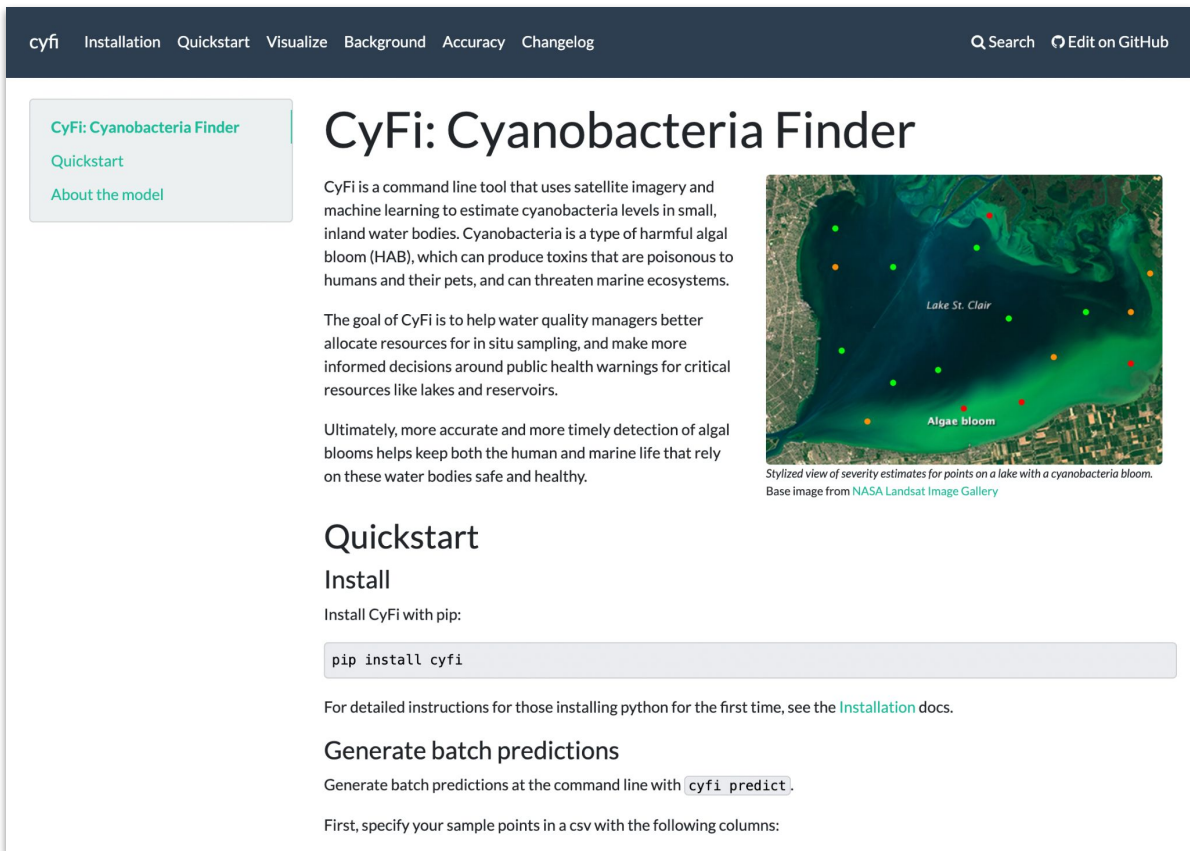
Plus, the algorithm is open source so anyone can reuse, update, or contribute.

The screenshot shows the GitHub repository page for 'drivendataorg / cyfi'. The repository is public and has 1 star, 0 forks, and 1 watch. The main branch is 'main'. The repository description is 'Estimate cyanobacteria density based on Sentinel-2 satellite imagery'. The repository URL is 'cyfi.drivendata.org/'. The repository has 1 release, 'v1.0.0', which is the latest version, released 7 hours ago. The repository contains several files and folders, including '.github/workflows', 'cyfi', 'docs', 'tests', '.gitignore', 'CHANGELOG.md', 'LICENSE', 'Makefile', 'README.md', 'pyproject.toml', and 'requirements\_docs.txt'. The repository also has 66 issues, with the most recent one being 'fix broken link' by 'ejm714' 7 hours ago.

File/Folder	Commit Message	Time Ago
.github/workflows	Prepare for release (#111)	17 hours ago
cyfi	release v1.0.0	7 hours ago
docs	release v1.0.0	7 hours ago
tests	Expand bbox to 2000 meters, filter to water are...	last week
.gitignore	Add docs (#107)	4 days ago
CHANGELOG.md	release v1.0.0	7 hours ago
LICENSE	initial commit	3 months ago
Makefile	Prepare for release (#111)	17 hours ago
README.md	fix broken link	7 hours ago
pyproject.toml	release v1.0.0	7 hours ago
requirements_docs.txt	Add docs (#107)	4 days ago

To learn more and start using CyFi today, go to:

[cyfi.drivendata.org](https://cyfi.drivendata.org)



The screenshot shows the homepage of the CyFi project. At the top, there is a navigation bar with links for 'cyfi', 'Installation', 'Quickstart', 'Visualize', 'Background', 'Accuracy', and 'Changelog'. On the right side of the navigation bar, there are search and GitHub links. Below the navigation bar, on the left, is a sidebar menu with three items: 'CyFi: Cyanobacteria Finder' (highlighted), 'Quickstart', and 'About the model'. The main content area features a large heading 'CyFi: Cyanobacteria Finder'. Below the heading is a paragraph describing CyFi as a command-line tool that uses satellite imagery and machine learning to estimate cyanobacteria levels in small, inland water bodies. To the right of this text is a satellite image of Lake St. Clair with a color-coded overlay representing cyanobacteria severity, with a label 'Algae bloom' pointing to a red area. Below the image is a caption: 'Stylized view of severity estimates for points on a lake with a cyanobacteria bloom. Base image from NASA Landsat Image Gallery'. Further down, there is a 'Quickstart' section with a sub-section 'Install' and the instruction 'Install CyFi with pip:'. Below this is a code block containing the command 'pip install cyfi'. The next section is 'Generate batch predictions', which includes the instruction 'Generate batch predictions at the command line with cyfi predict.' and a note that the user should first specify their sample points in a csv with the following columns:

CyFi: Cyanobacteria Finder

Quickstart

About the model

## CyFi: Cyanobacteria Finder

CyFi is a command line tool that uses satellite imagery and machine learning to estimate cyanobacteria levels in small, inland water bodies. Cyanobacteria is a type of harmful algal bloom (HAB), which can produce toxins that are poisonous to humans and their pets, and can threaten marine ecosystems.

The goal of CyFi is to help water quality managers better allocate resources for in situ sampling, and make more informed decisions around public health warnings for critical resources like lakes and reservoirs.

Ultimately, more accurate and more timely detection of algal blooms helps keep both the human and marine life that rely on these water bodies safe and healthy.



Stylized view of severity estimates for points on a lake with a cyanobacteria bloom. Base image from NASA Landsat Image Gallery

### Quickstart

#### Install

Install CyFi with pip:

```
pip install cyfi
```

For detailed instructions for those installing python for the first time, see the [Installation docs](#).

#### Generate batch predictions

Generate batch predictions at the command line with `cyfi predict`.

First, specify your sample points in a csv with the following columns: