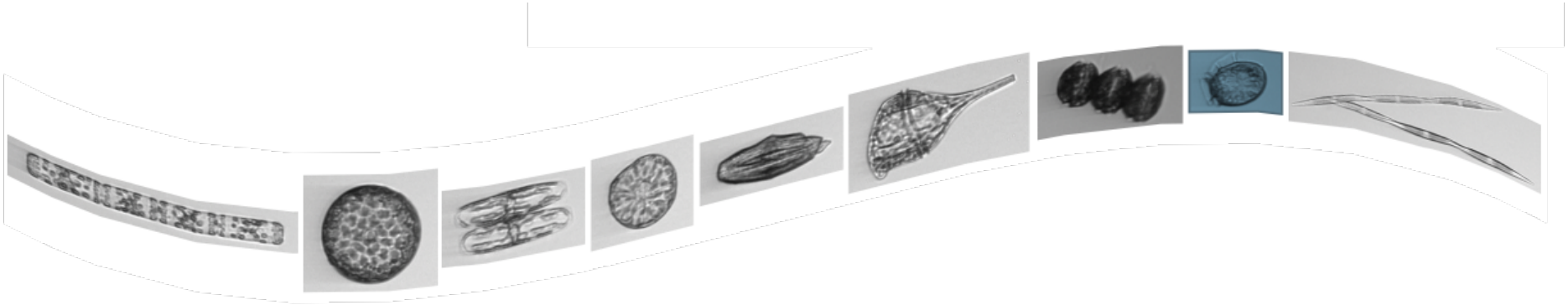
ICHA 19th INTERNATIONAL CONFERENCE ON HARMFUL ALGAE MEXICO 2021

Development and comparison of IFCB classifiers in coastal California

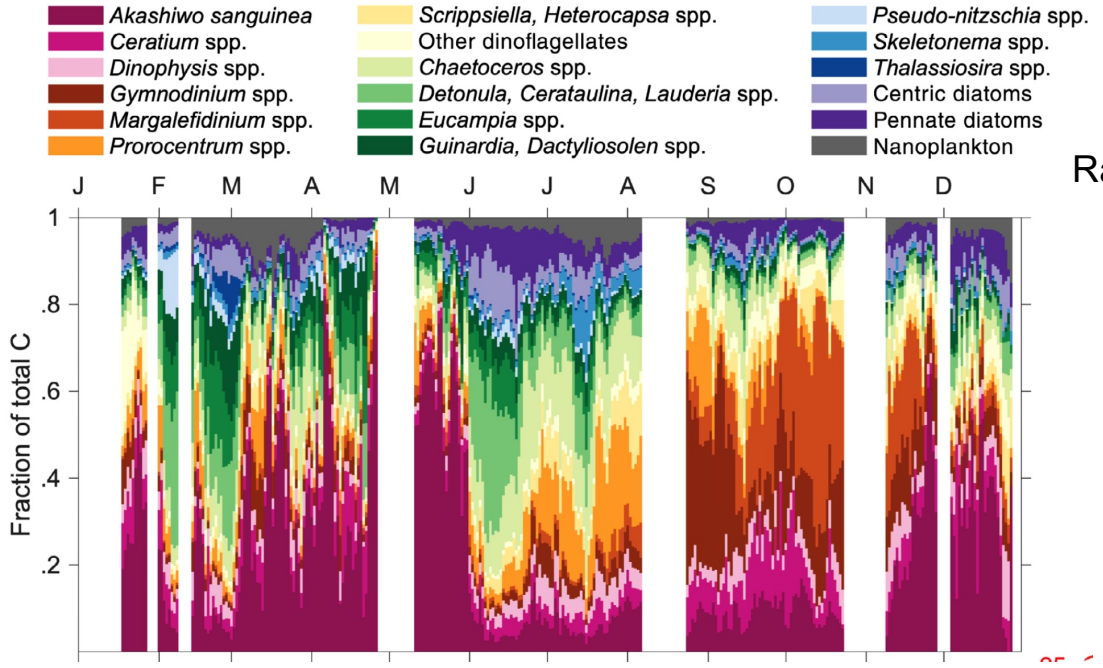
Anna R. McGaraghan, Kendra Hayashi, Raphael M. Kudela
Ocean Sciences, University of California Santa Cruz, 1156 High St, Santa Cruz, California, USA

NT-P-4

Alexis Fischer (UCSC, NOAA), Patrick Daniel (UCSC), Jamie Walton (UC Berkeley)



UCSC Image Database



Random Forest Model built with ~112,000 images

- 24 group/genus/species classes
- > 90% of total biomass
- Includes major HAB groups

Total of 1,003,334 annotations:

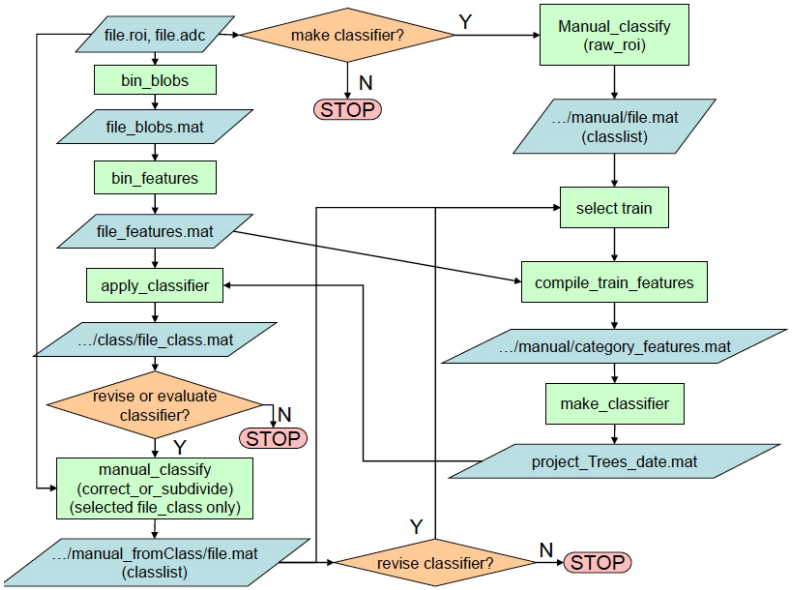
➔ 97 genus/species classes

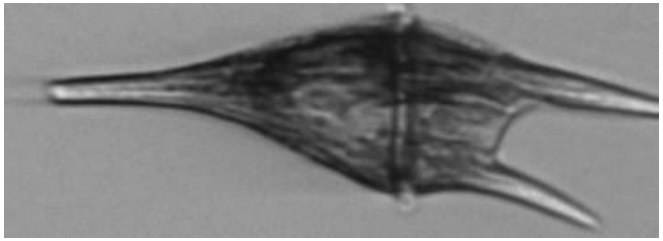
➔ Several “super groups”

➔ Ash, cysts, detritus, etc.

Fischer et al. L&O 65: 2125-2141, 2020

Supervised Classification





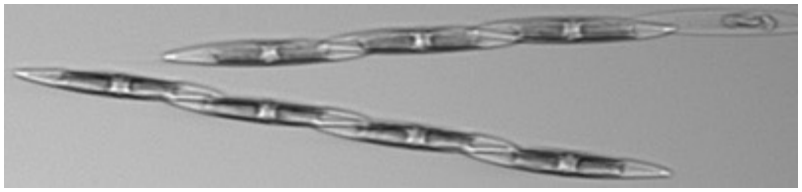
Santa Cruz Wharf IFCB

- 78 Classes (46 routinely used)
- Manually classified ~210,000 images
- ~100,000 images both manual and auto

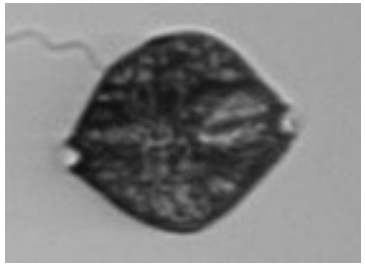
- 38 classes have at least 50 manual IDs, or ~70% of all the images

- 88% accuracy for this subset, 80% accuracy for the full dataset
- Most accurate (>90%) are most abundant or easy to identify: *Akashiwo, Ceratium, Prorocentrum, Pseudo-nitzschia, Lingulodinium*

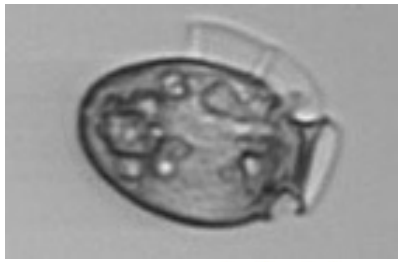
Examples of Auto-Classification



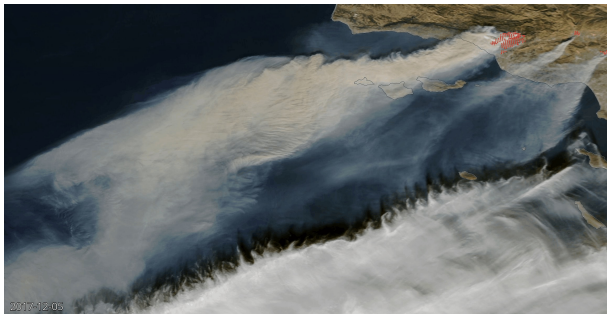
Examples of Auto-Classification



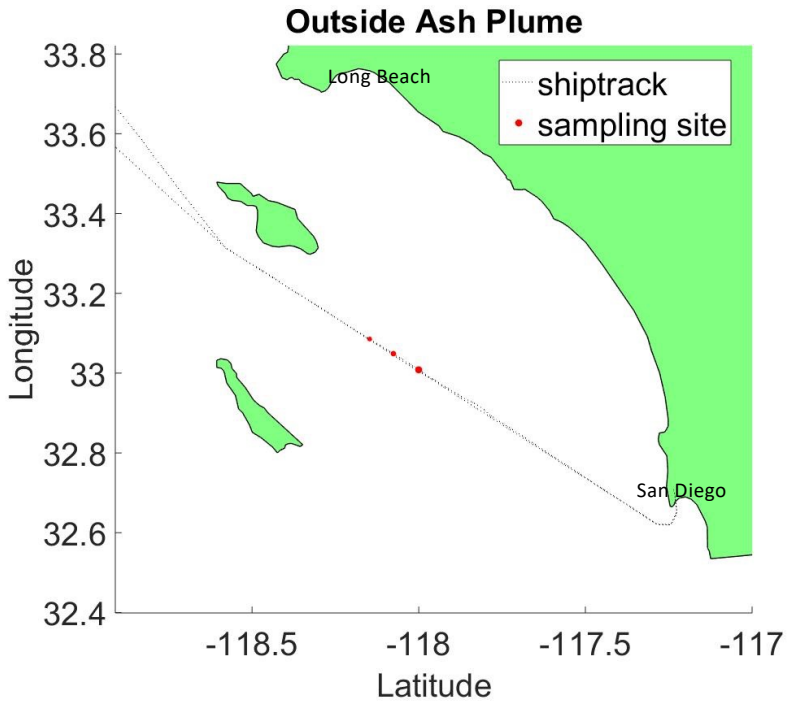
Examples of Auto-Classification

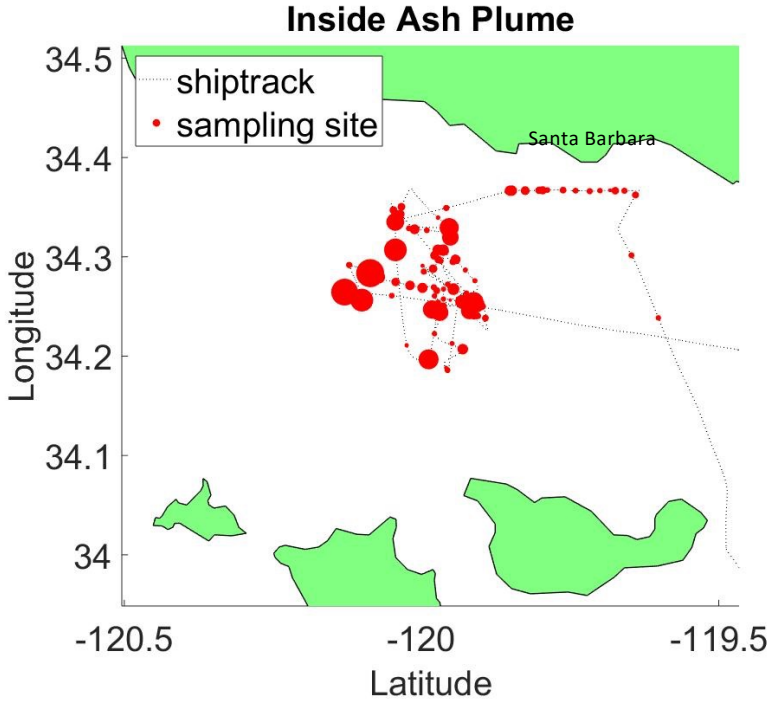


ACIDD Cruise (UC Santa Barbara)

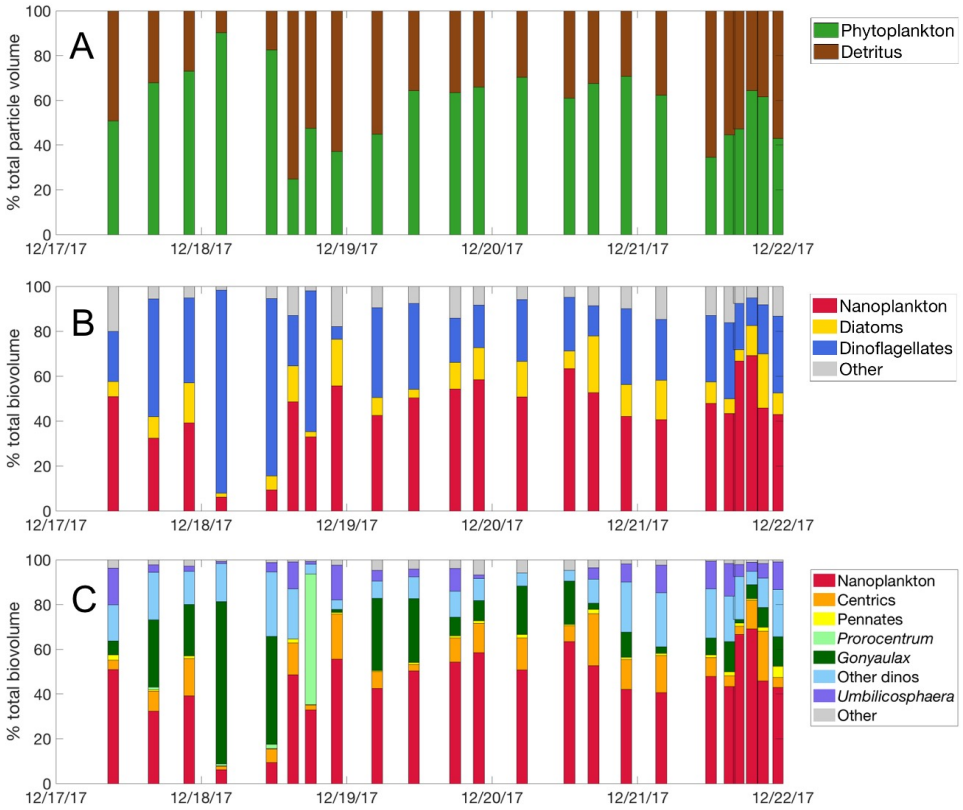


Alexis Fischer installed an IFCB as part of the cruise—cruise map with average ESD

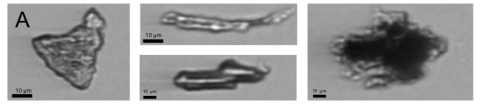




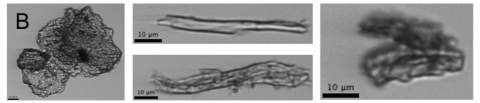
UCSC Image Database



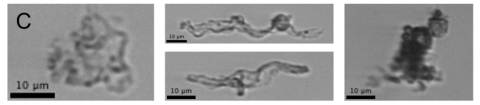
ACIDD "pure ash in filtered seawater"



NAAMES 3 microlayer "microplastics"



Damariscotta River Estuary "detritus"



Total of 1,003,334 annotations:

➔ 97 genus/species classes

➔ Several "super groups"

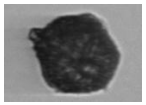
➔ Ash cysts, detritus, etc.

Kramer et al. JGR Oceans, 2020

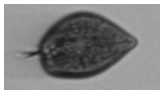
How Portable are Classifiers?

- Applied SCW Random Forest to Santa Barbara
- Worked OK for some organisms, not others….

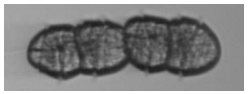
- ***Lingulodinium polyedrum***
  - $1.4x10^4$ cells/Liter
  - >1 SD from mean

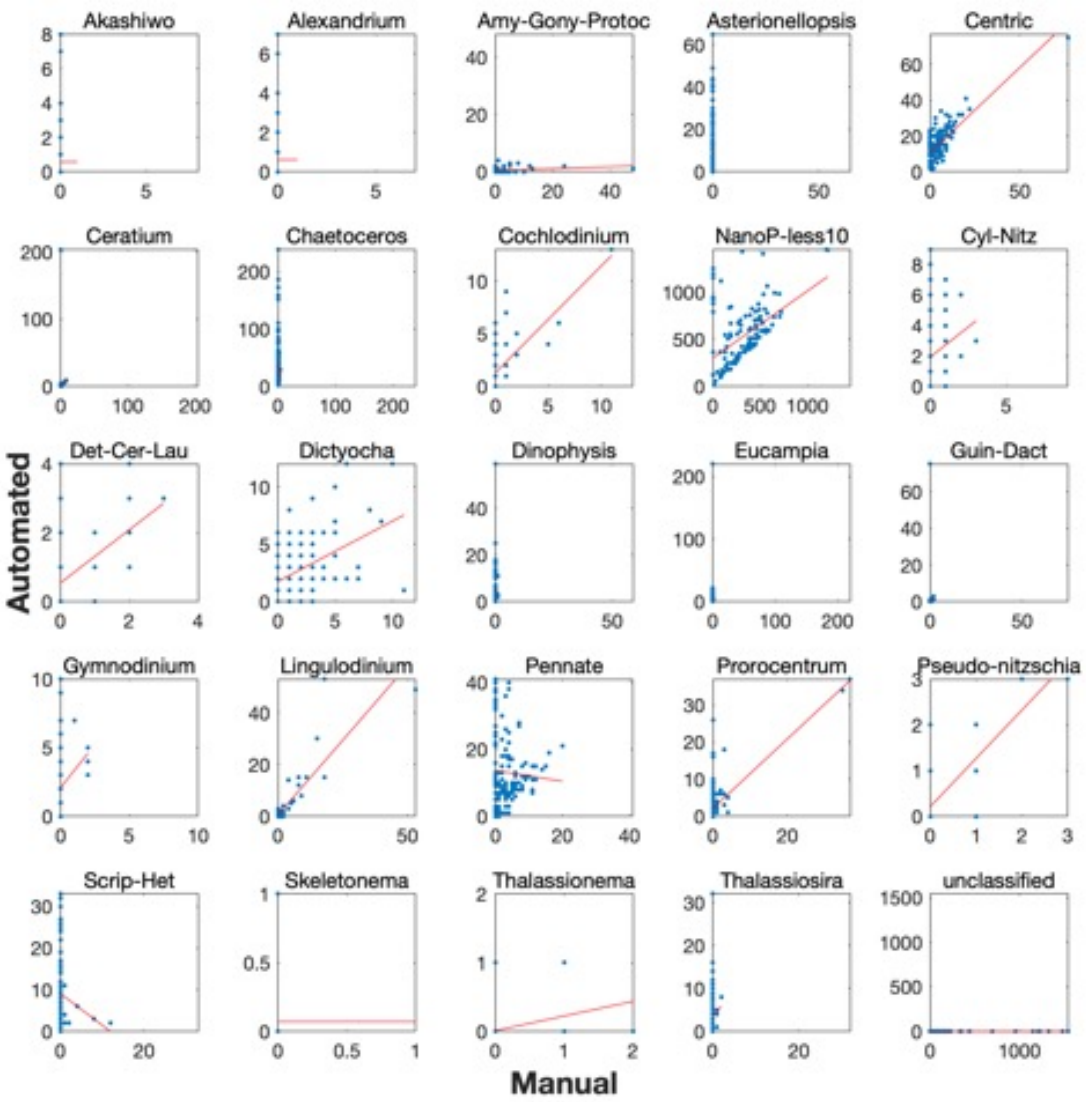


- ***Prorocentrum***
  - $0.9x10^4$ cells/Liter
  - >8 SD from mean



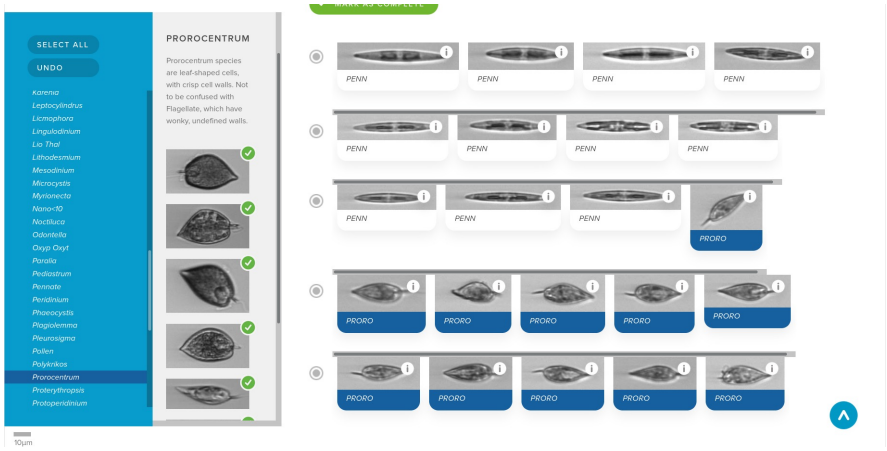
- ***Cochlodinium***
  - ~4000 cells/Liter, 4x more during bloom conditions





Moving from Random Forest to CNN

Data

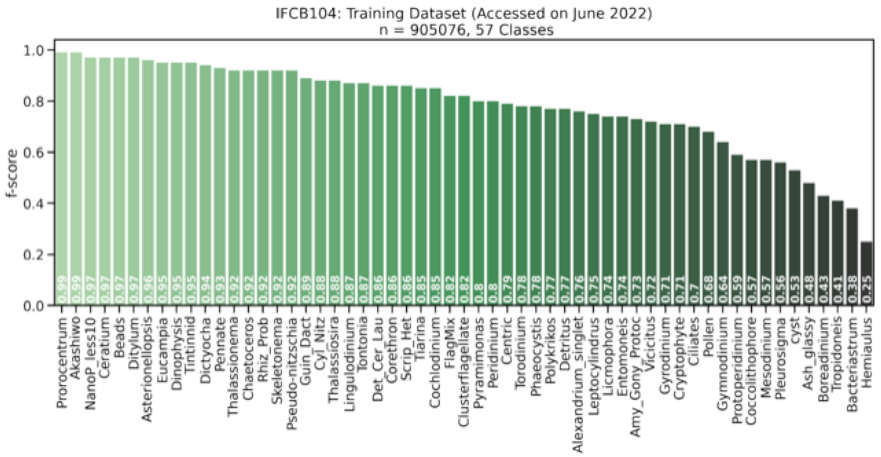


export_png_manual_fromROI.m

Training: 80%
Testing: 10%
Validation: 10%

Tesla P100 GPU

[train-xception-hb.py]

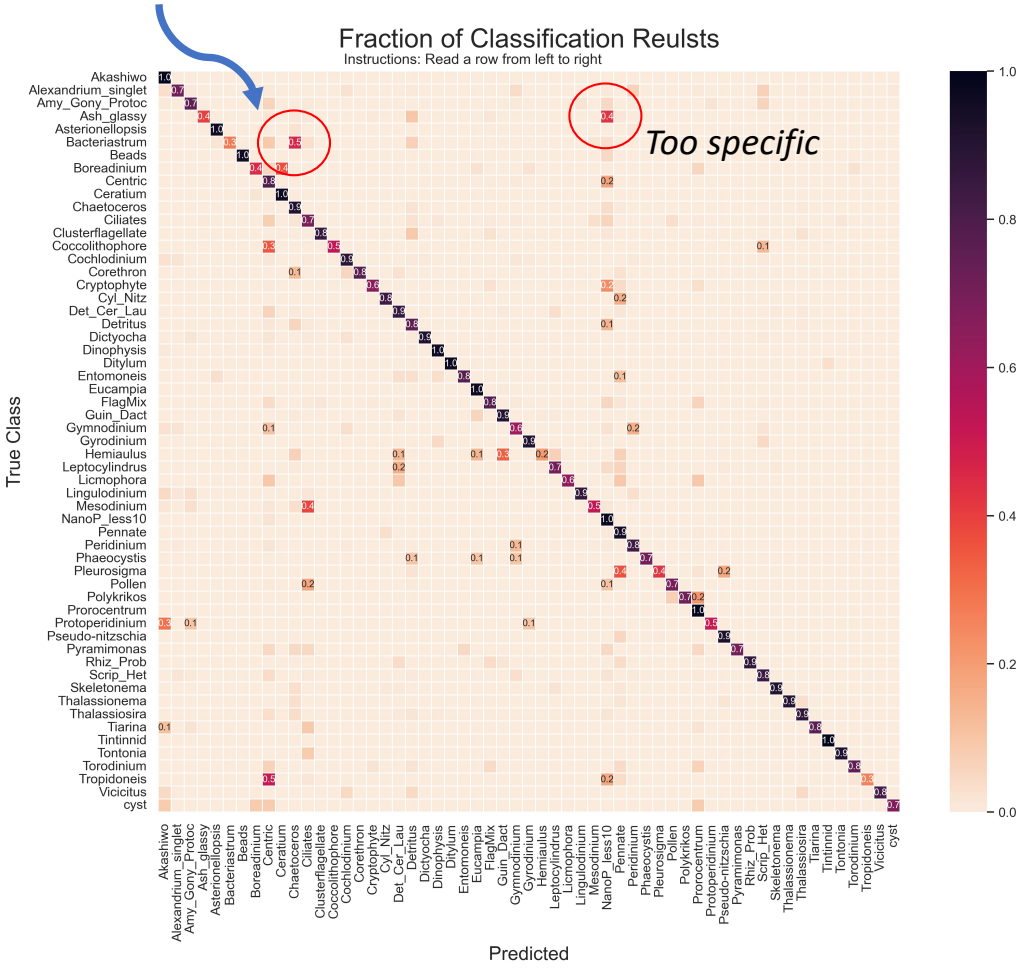


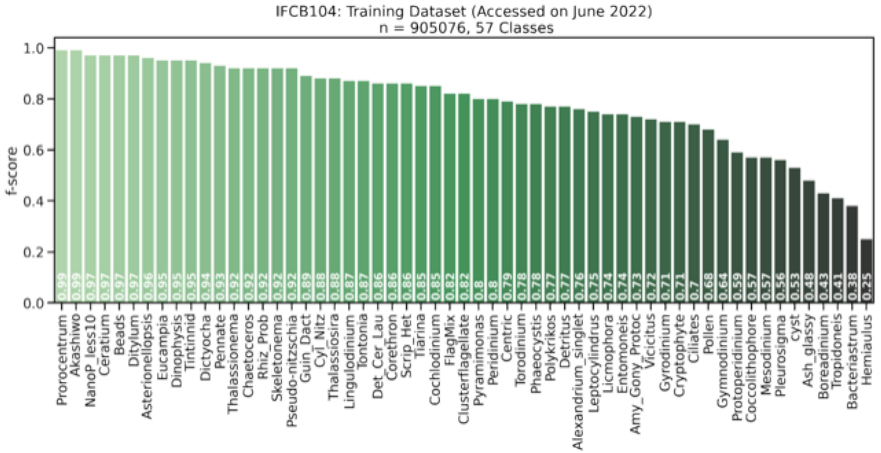
905,076 annotations:

- ➔ 57 genus/species classes
- ➔ CNN (Xception)
- ➔ $F_1 = 0.93$

UCSC Image Database

Annotation error

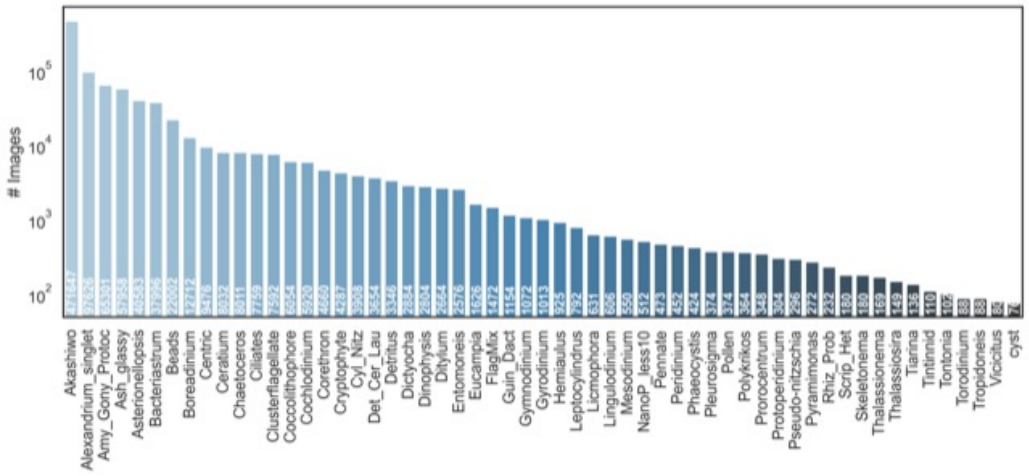


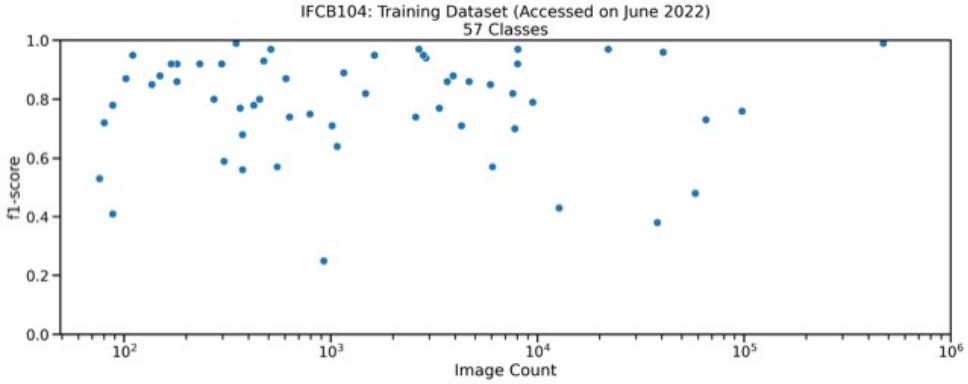


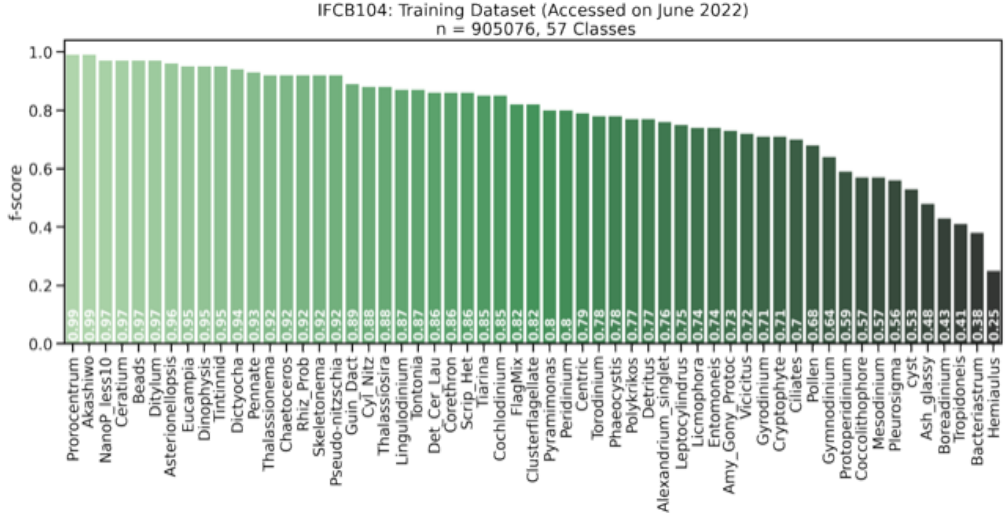
Total of 905,076 annotations:

- ➔ 57 genus/species classes
- ➔ CNN (Xception)
- ➔ F_1 = 0.93

UCSC Image Database







Initial Analysis: F1 score is not particularly dependent on the number of images….

Accuracy:			0.93
Macro avg:	0.82	0.76	0.78
Weighted avg:	0.93	0.93	0.93

ICHA 19th INTERNATIONAL CONFERENCE ON HARMFUL ALGAE MEXICO 2021

Development and comparison of IFCB classifiers in coastal California

Anna R. McGaraghan, Kendra Hayashi, Raphael M. Kudela
Ocean Sciences, University of California Santa Cruz, 1156 High St, Santa Cruz, California, USA

NT-P-4

- Can we use multiple classifiers to operationally identify HABs?
- How sensitive is it changing IFCBs?

- Two instruments, 4 classifiers:
  - SCW Random Forest (24 classes)
  - San Francisco Bay Random Forest (32 classes)
  - TAMU Random Forest (54 classes)
  - SCW CNN (50 classes)

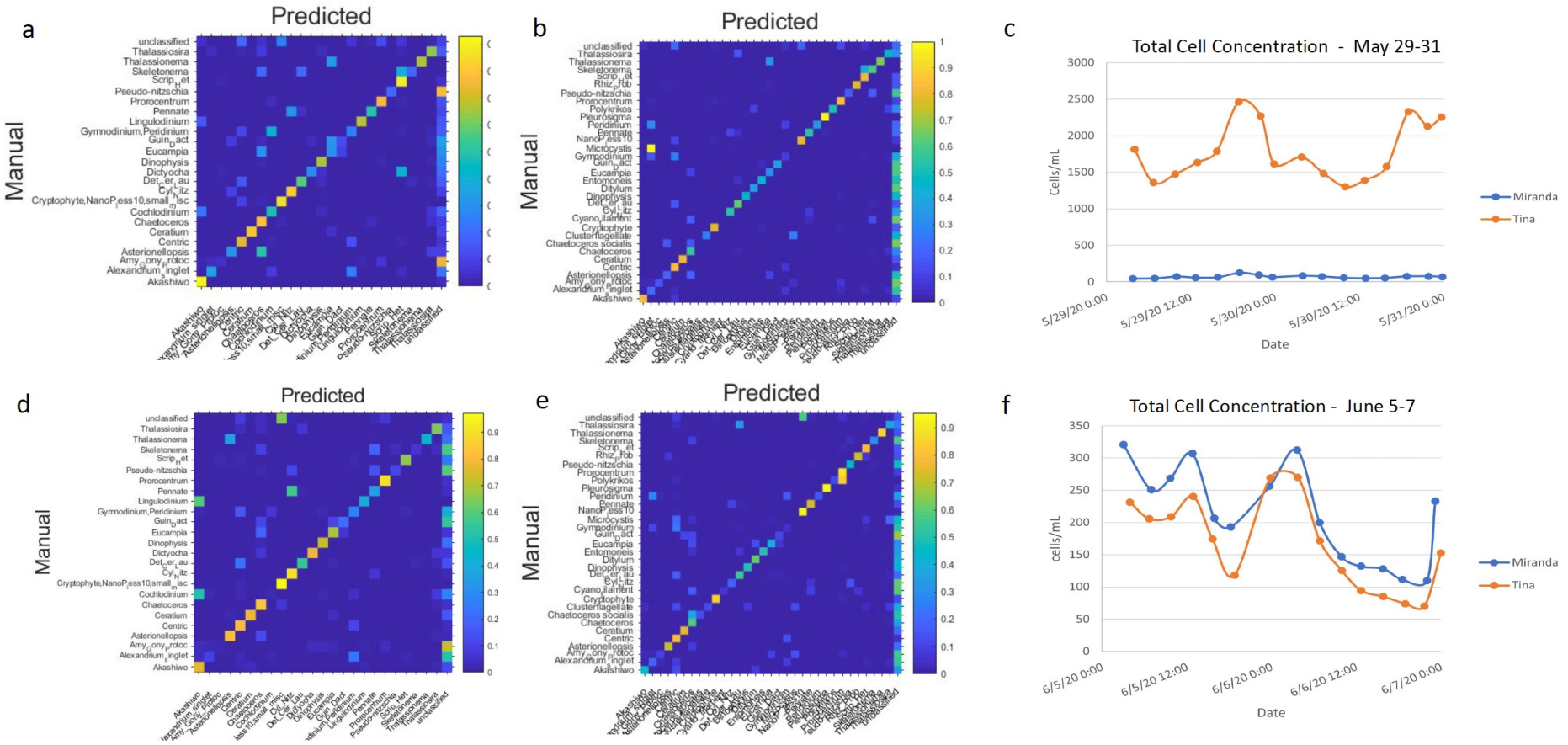
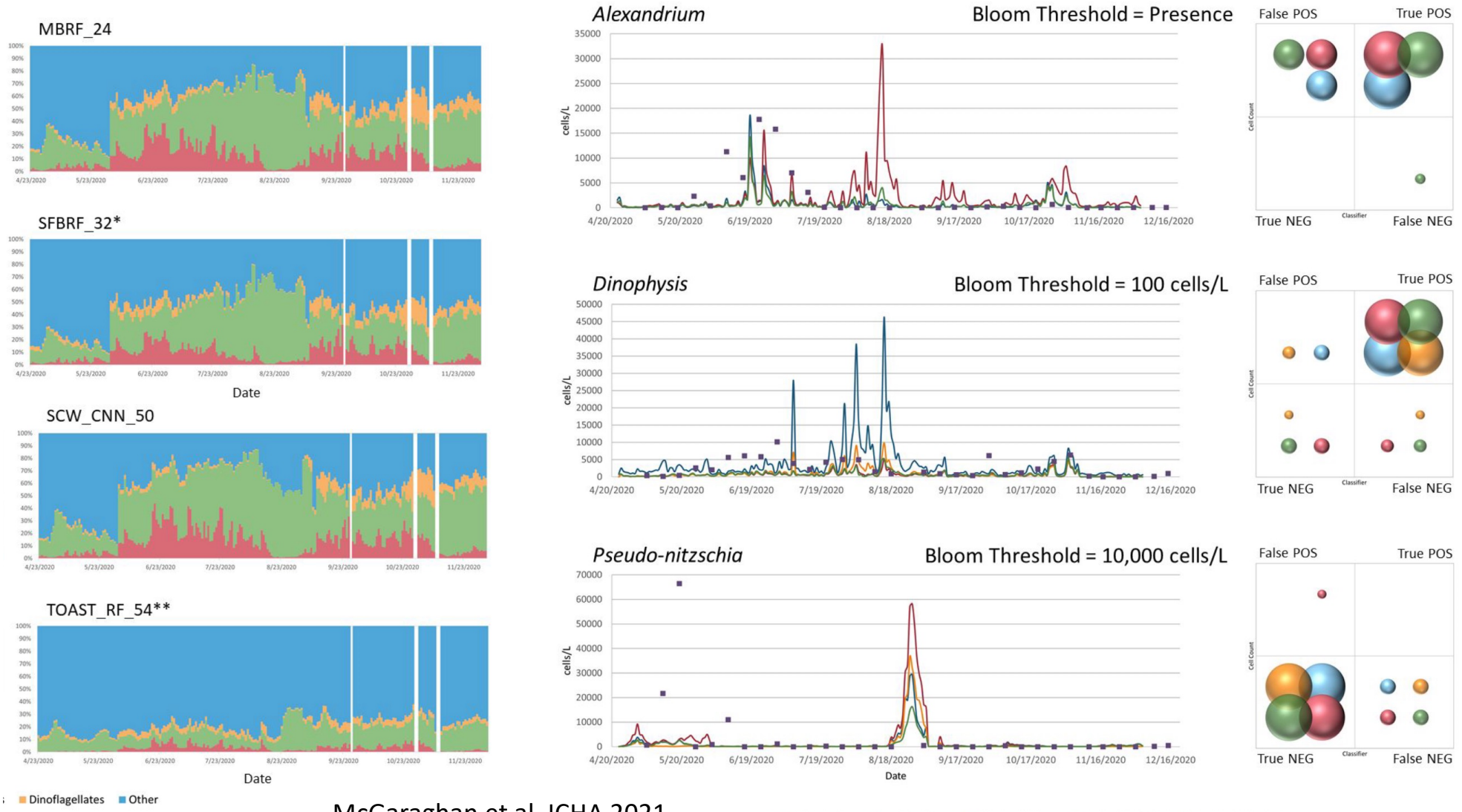


Figure 3: Performance of RF classifiers at the Santa Cruz Wharf in 2020, on IFCB104/Tina and IFCB113/Miranda. Manual annotation vs MBRF_24 classifier for (a) Miranda and (d) Tina, and SFBRF_32 for (b) Miranda and (e) Tina. Line graphs show total cell counts between the instruments at the beginning of deployment before adjusting Miranda's PMTs to reflect cell sizes in Monterey Bay.



McGaraghan et al. ICHA 2021

Development and comparison of IFCB classifiers in coastal California

Anna R. McGaraghan, Kendra Hayashi, Raphael M. Kudela
Ocean Sciences, University of California Santa Cruz, 1156 High St, Santa Cruz, California, USA

NT-P-4

- Initial performance of SFBRF_32 is equivalent to MBRF_24, with 99.1% overall accuracy

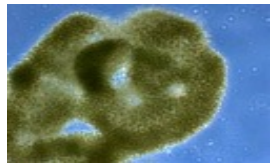
- SFBRF_32 delivered similar performance results as MBRF_24 during the side-by-side IFCB deployment at the Santa Cruz Wharf. All accuracies 94-96%

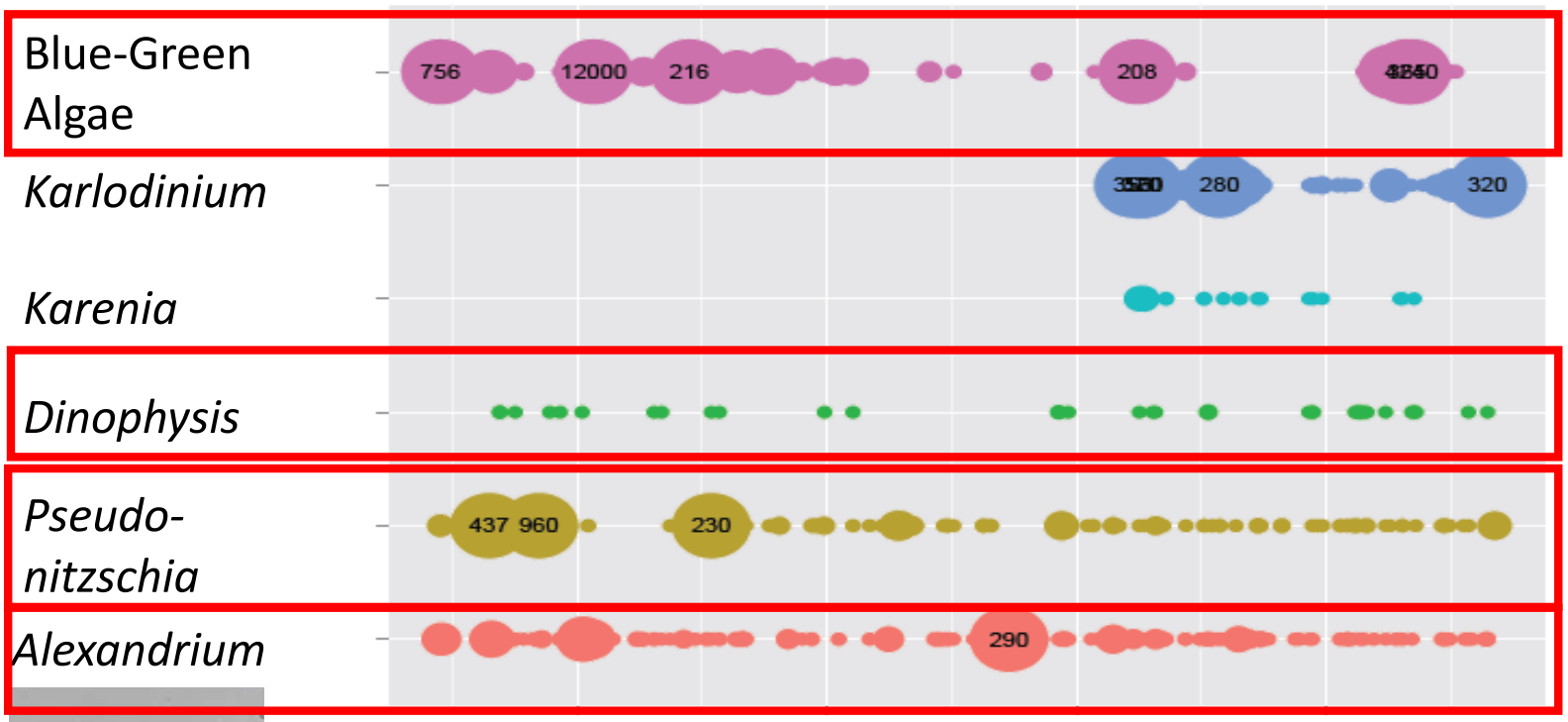
- IFCB PMT settings have a noticeable impact on how data convey overall phytoplankton community structure

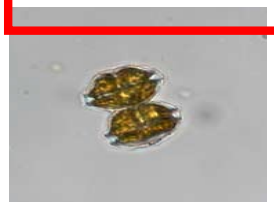
- Choice of classifier influences our perception of community structure

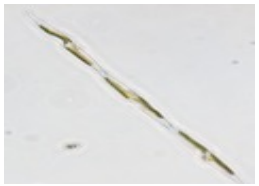
- Out of region classifiers may be applied to effectively assess bloom detection of specific HAB taxa in new IFCB locations. Both classifier accuracies were 99% for the 3 taxa

San Francisco Bay:
Four Toxic Groups of Concern

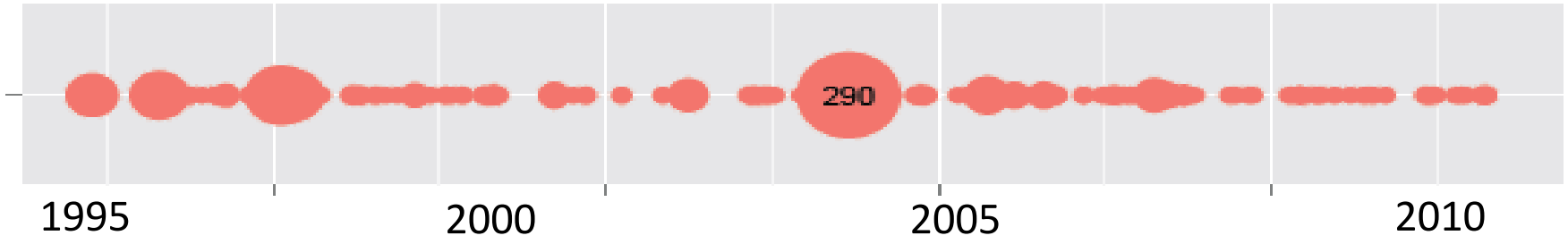




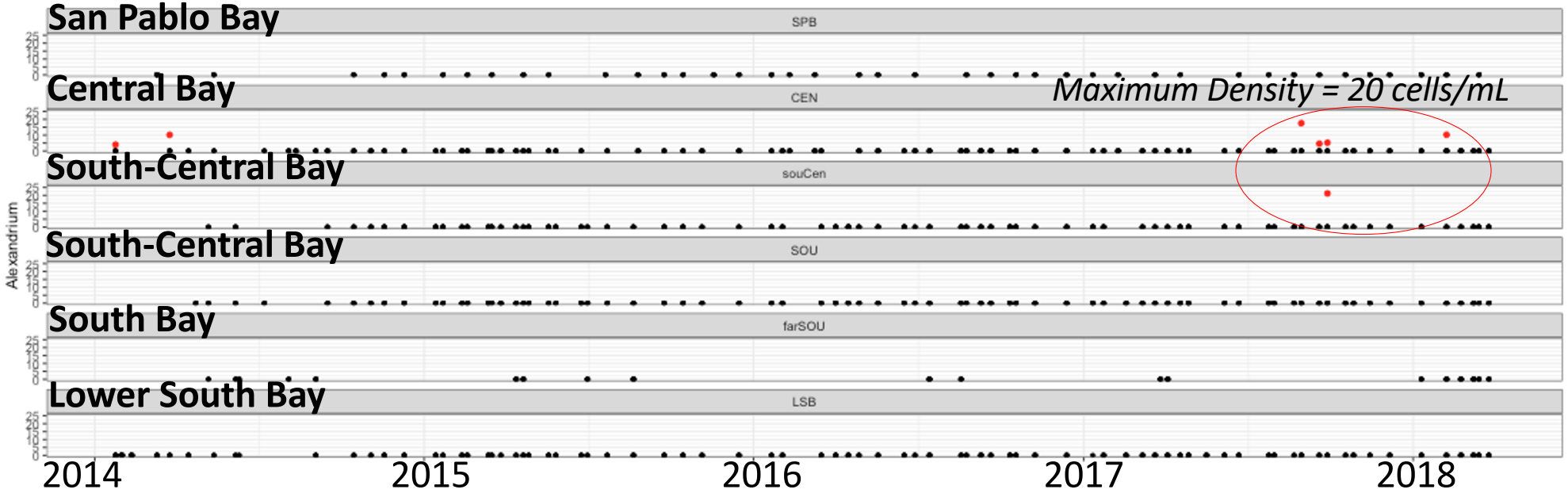


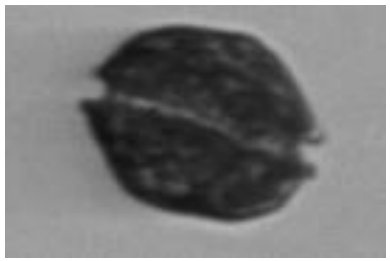


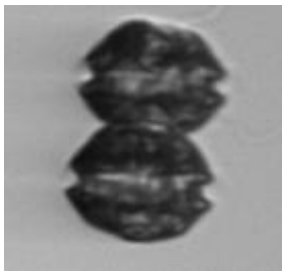
Alexandrium – Original Microscopist

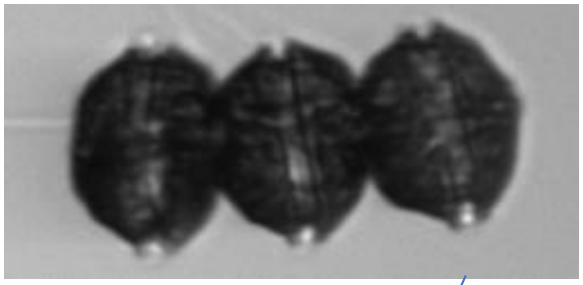


Alexandrium – New Microscopist



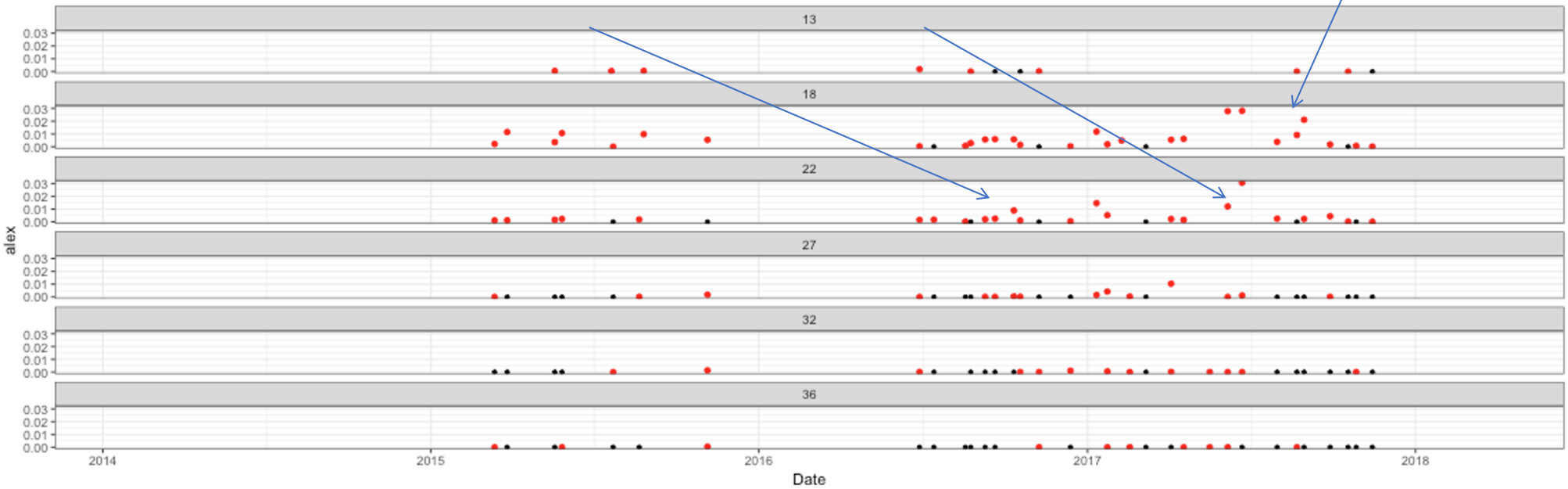




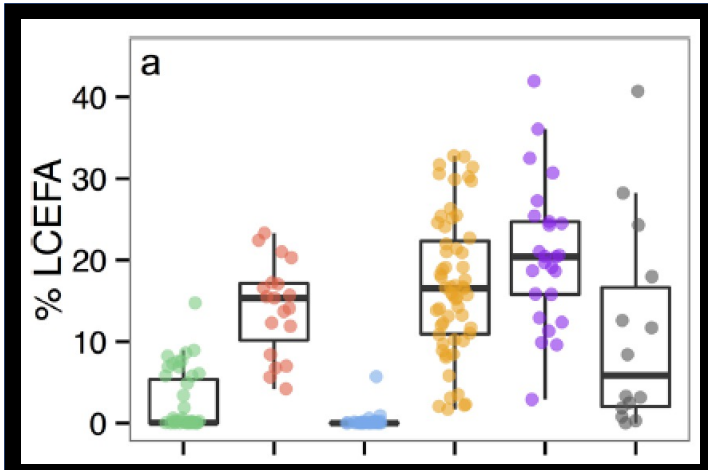


Alexandrium – Proportion of qPCR samples for 18S rDNA primers spanning the variable V5 (Euk1132F) and V7 (Euk1423R) regions

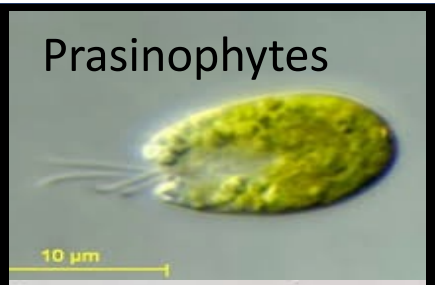
31 cells/mL

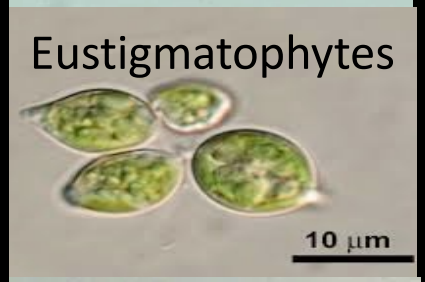


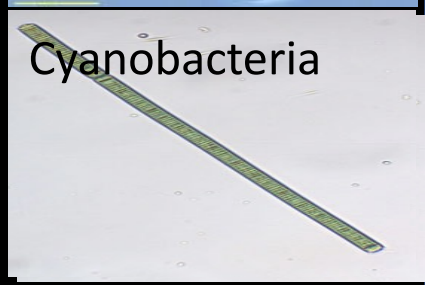
2014 2015 2016 2017 2018



Galloway & Winder, PLoS ONE 2015



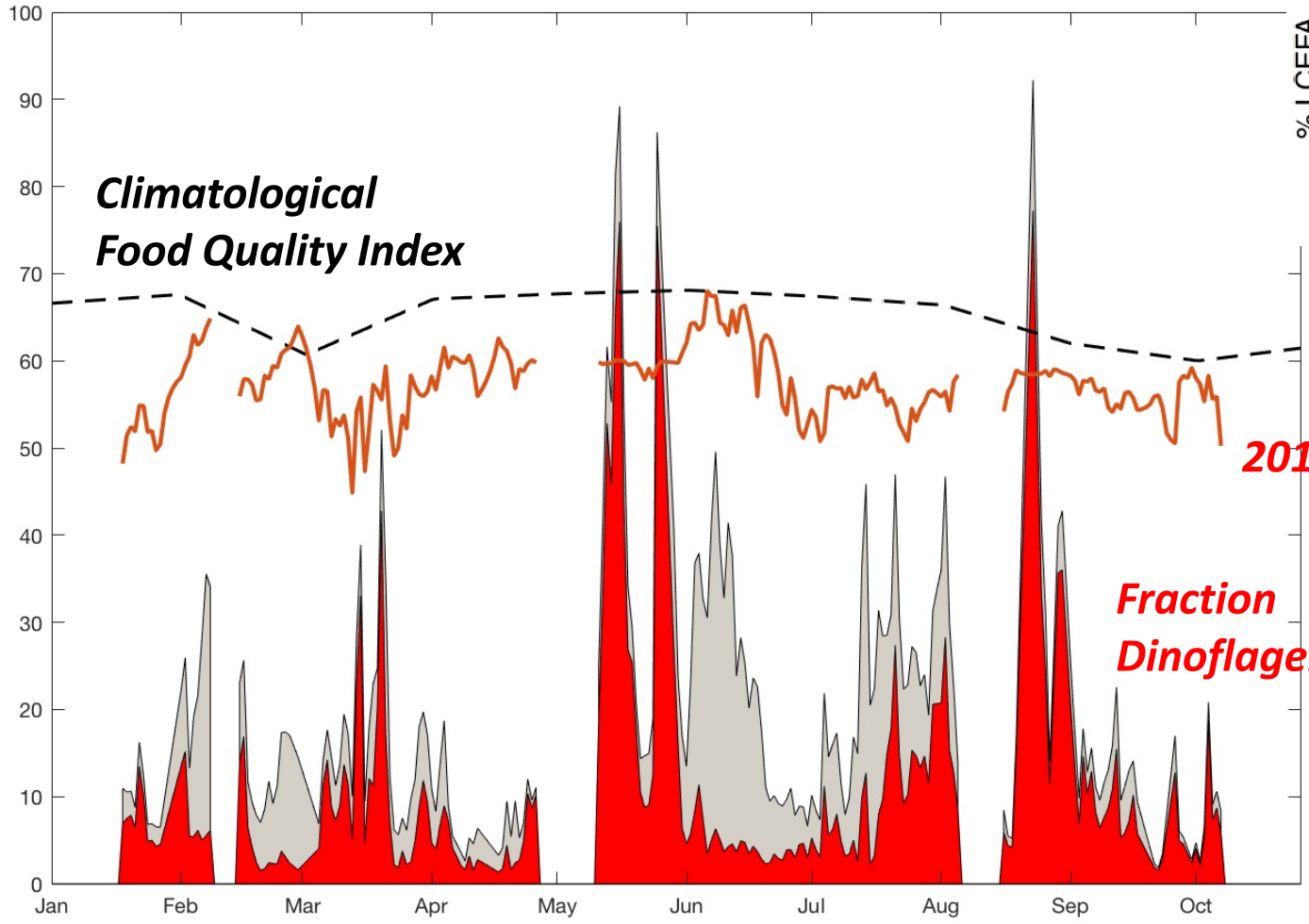


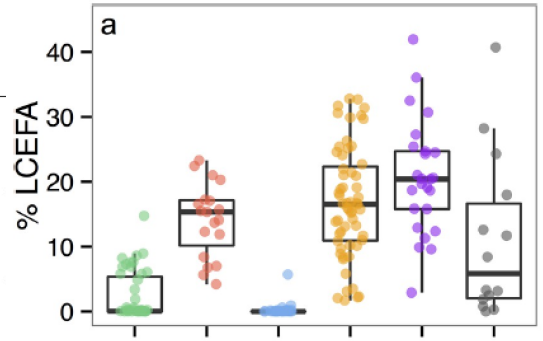


Phytoplankton Food Quality Index

- Each PFT assigned a nutritional value
- Values based on evolutionary traits
- Good correspondence between microscopy, HPLC, IFCB, and Satellite Remote Sensing

Climatology based on Schulien et al., MEPS 2017





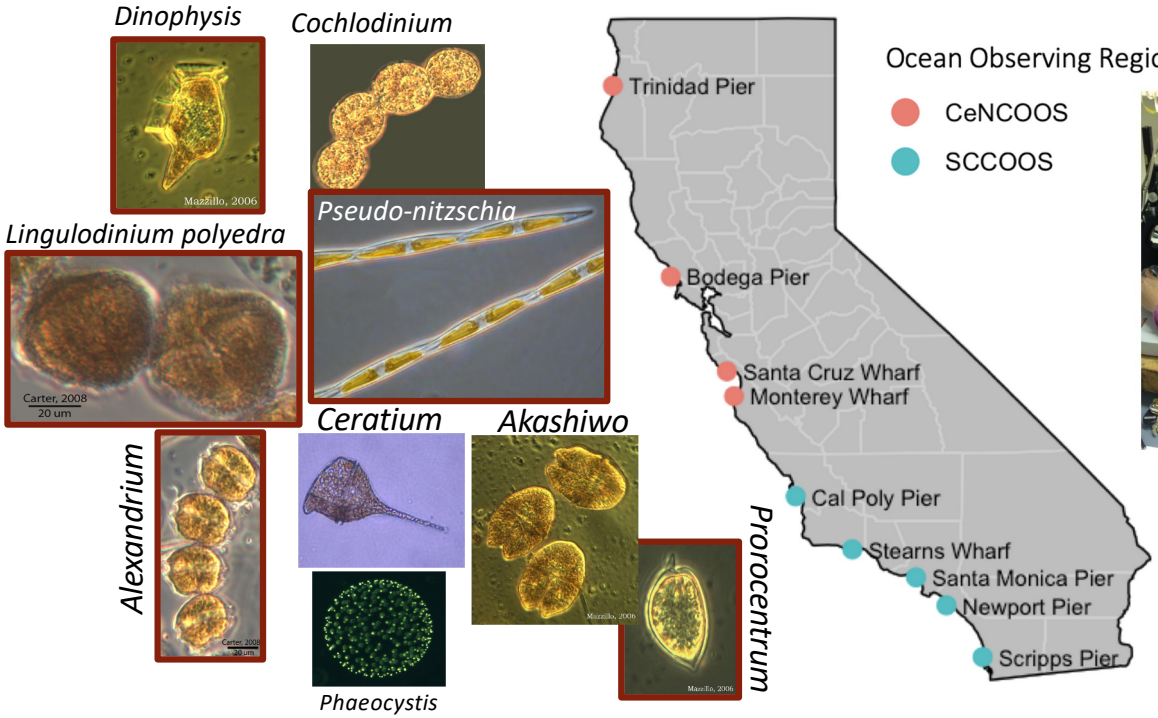
HARMFUL ALGAL BLOOM MONITORING & ALERT PROGRAM

IFCBs are ready to go for operational monitoring!

Weekly measurements:

- **HAB spp. (8-9 taxa)**
- Chl-a, Temp, Salinity, **Nutrients**
- **Domoic Acid + SPATT toxins**
- Weekly alerts to HABMAP listserv
- Monthly QC'd data now served via ERDDAP
- Synthesis with models: **CA HAB Bulletin**

- **10 academic institutions**



SOUTHERN CALIFORNIA COASTAL OCEAN OBSERVING SYSTEM

CENTRAL & NORTHERN CALIFORNIA OCEAN OBSERVING SYSTEM