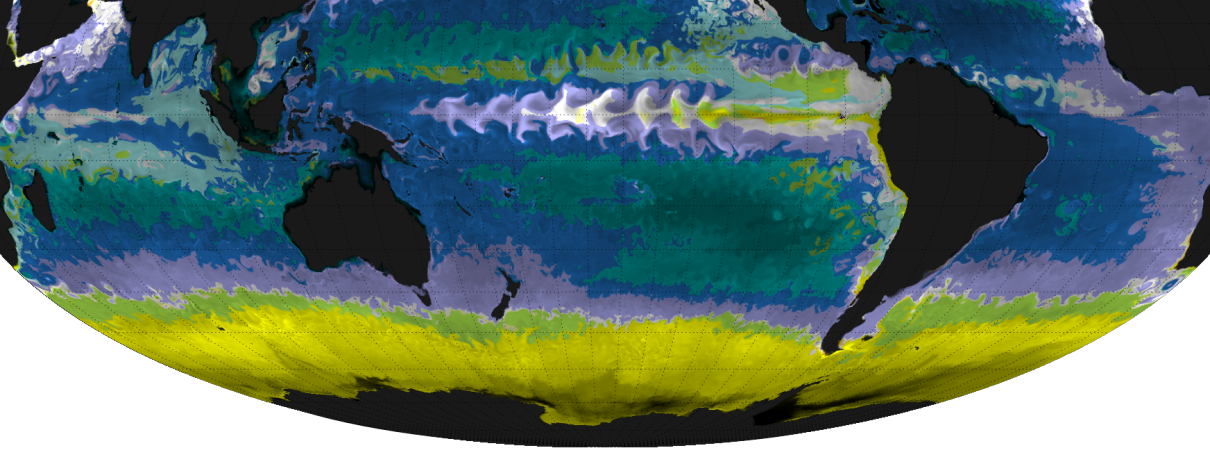
Implementing Darwin in ROMS: Initial results from regional simulations with a complex ecosystem model

J. Paul Mattern[1] Christopher A. Edwards[1] Brian Powell[2] Tobias Friedrich[2]
Stephanie Dutkiewicz[3] Oliver Jahn[3] Mick Follows[3]

cbiomes
Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems

[1] Ocean Sciences Department, UC Santa Cruz
[2] Department of Oceanography, University of Hawai'i at Mānoa
[3] Program in Atmospheres, Oceans, and Climate, MIT

March 2021



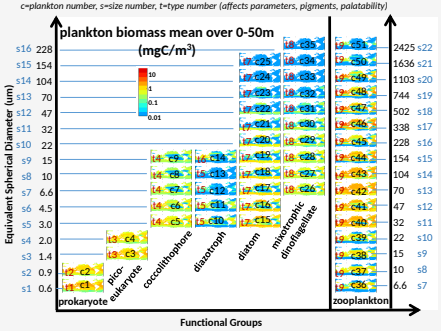
cbiomes

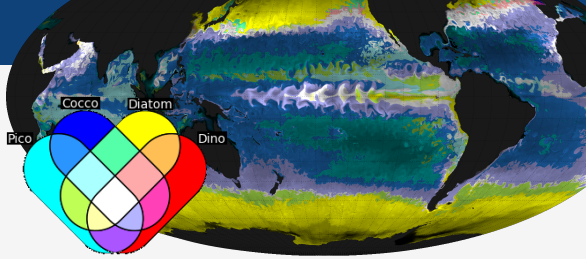
Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems

This work would not be possible without a large network of collaborators from different institutions, especially the efforts of

- Stephanie Dutkiewicz, Oliver Jahn, and the Darwin team at MIT.
- Brian Powell and Tobias Friedrich who have been running ROMS Darwin in their Hawai'i domain.

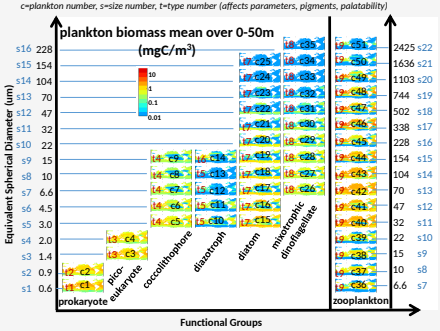
a ROMS implementation of Darwin

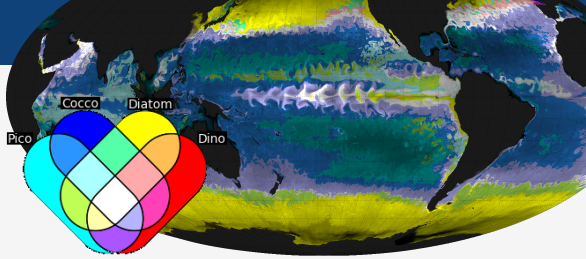




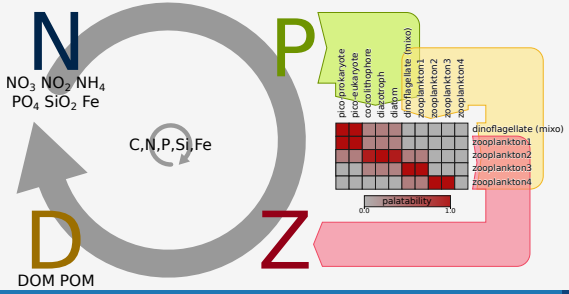
- Darwin is a complex ecosystem model, developed at MIT and originally implemented in the MITgcm and run globally.

a ROMS implementation of Darwin

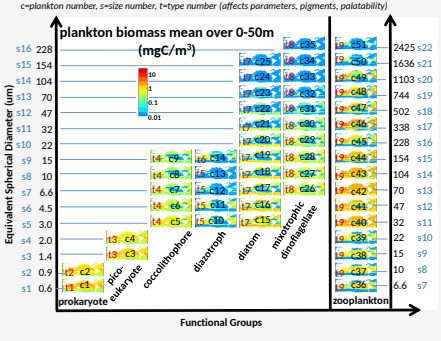


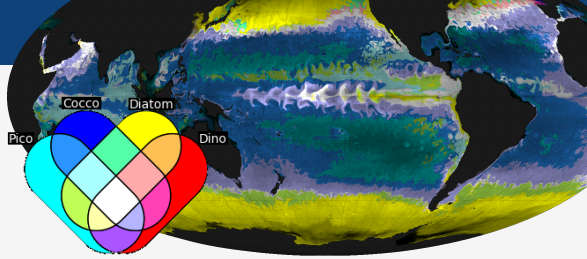


- Darwin is a complex ecosystem model, developed at MIT and originally implemented in the MITgcm and run globally.

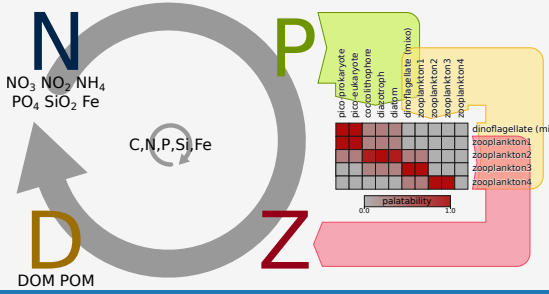


a ROMS implementation of Darwin

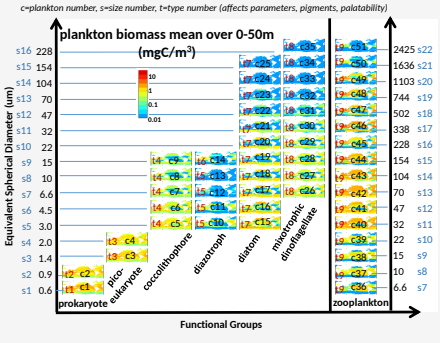


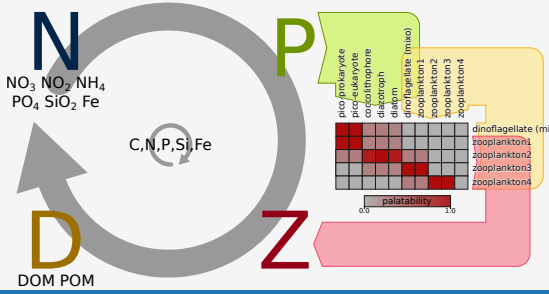


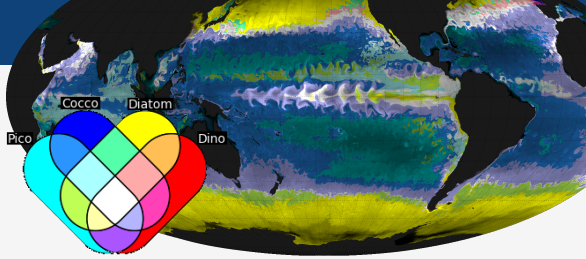
- Darwin is a complex ecosystem model, developed at MIT and originally implemented in the MITgcm and run globally.
- We created a regional version of Darwin implemented in ROMS (the Regional Ocean Modeling System, `myroms.org`).
- The code conversion is performed semi-automatically using a Python program, not all features of the original Darwin code are currently supported (for example the radiative transfer model).



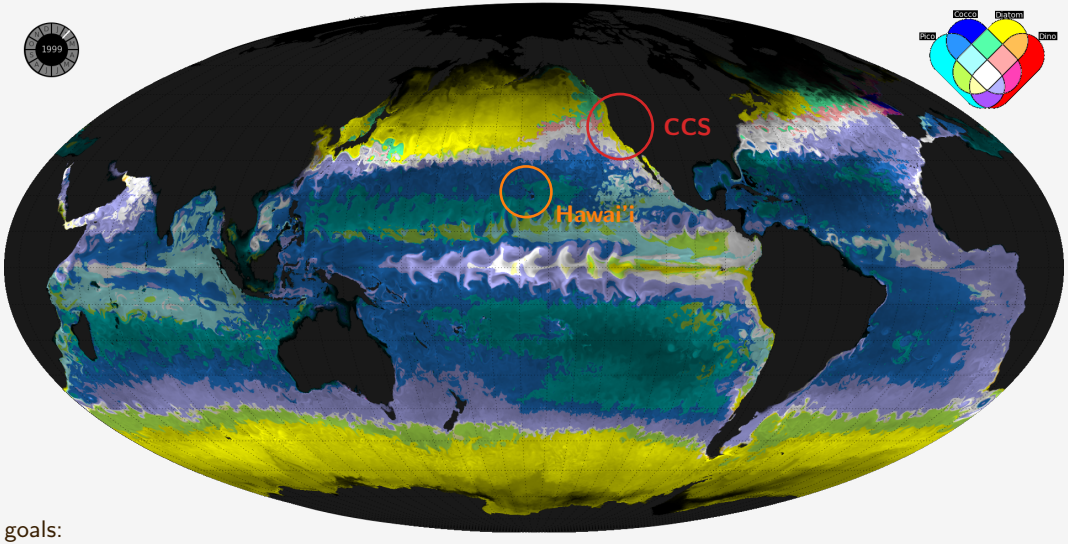
a ROMS implementation of Darwin







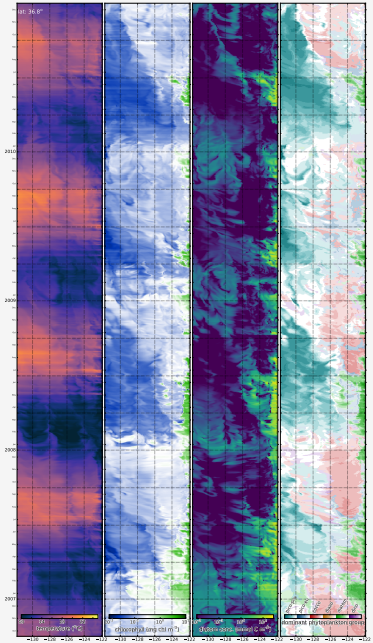
- Darwin is a complex ecosystem model, developed at MIT and originally implemented in the MITgcm and run globally.
- We created a regional version of Darwin implemented in ROMS (the Regional Ocean Modeling System, `myroms.org`).
- The code conversion is performed semi-automatically using a Python program, not all features of the original Darwin code are currently supported (for example the radiative transfer model).
- Darwin can be run in different plankton configurations, here I will mostly show a "6P4Z" setup (still containing 37 biological variables).



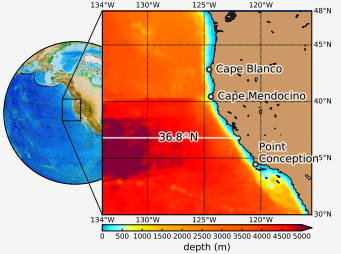
goals:

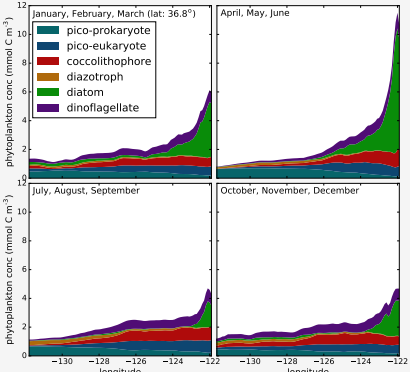
- Explore data assimilation in a complex model and improve Darwin state estimates locally.
- Study Darwin dynamics at a high resolution and in coastal settings.
  (California Current System, Hawai'i)
- Find interesting and informative ways to visualize all the information contained in the Darwin output.
- This is very much a work in progress...
  I am happy to receive input!

a free (no data assimilation) 6P4Z Darwin simulation in the California Current System (CCS)



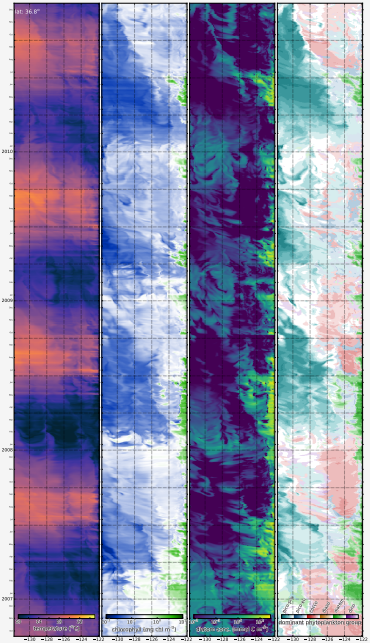
Initial results applying Darwin to our CCS domain after small amounts of manual tuning:



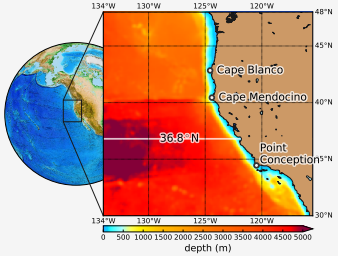


We have not performed a quantitative comparison with data yet but the phytoplankton distribution and dynamics look promising.

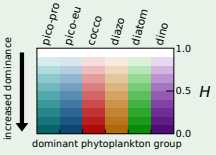
a free (no data assimilation) 6P4Z Darwin simulation in the California Current System (CCS)



Initial results applying Darwin to our CCS domain after small amounts of manual tuning:



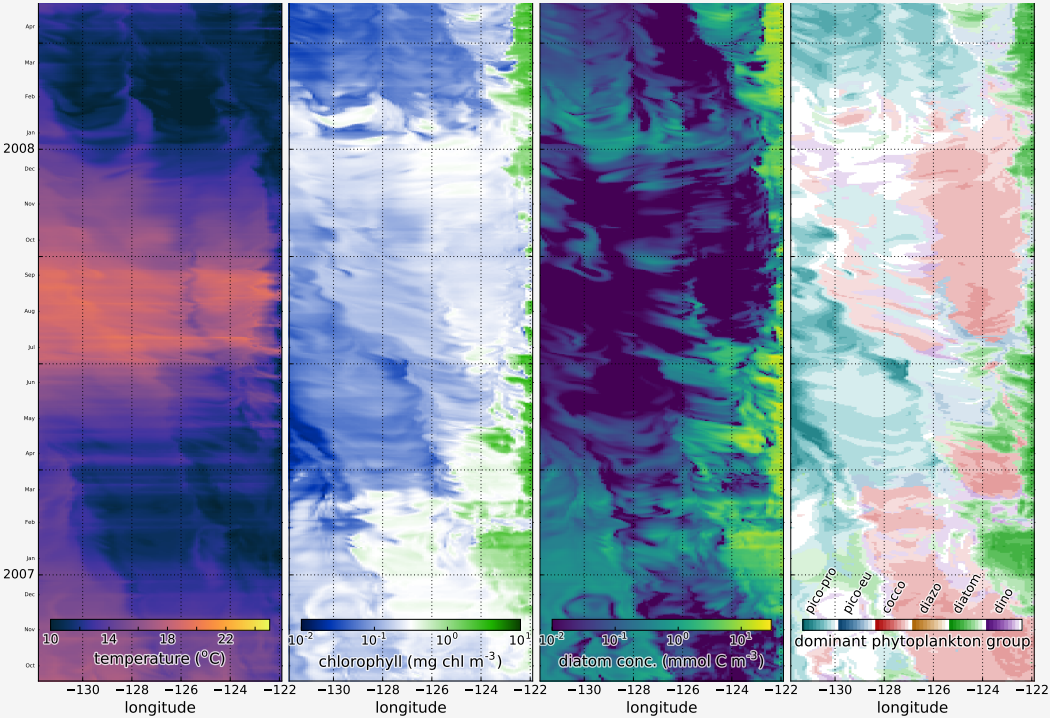
information entropy to visualize the dominant phytoplankton group



Information entropy H is a measure of information content, for 6 phytoplankton groups we define H as

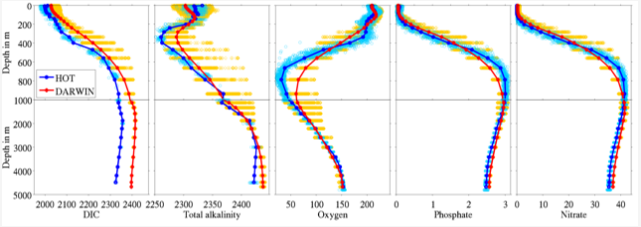
$$H = -\sum_{i=1}^{6} p_i \log_6(p_i) \in [0, 1].$$

Here p_i is the relative abundance of phytoplankton group i. Low values of H close to 0 indicate that one phytoplankton group is dominant and receive a low lightness when plotted (hue indicates which group is the dominant one). High values imply a distribution that is close to uniform and receive a high lightness.


Apr
Mar
Feb
Jan
2008
Dec
Nov
Oct
Sep
Aug
Jul
Jun
May
Apr
Mar
Feb
Jan
2007
Dec
Nov
Oct
10
14
18
22
temperature (°C)
10^{-2}
10^{-1}
10^{0}
10^{1}
chlorophyll (mg chl m^{-3})
10^{-2}
10^{-1}
10^{0}
10^{1}
diatom conc. (mmol C m^{-3})
pico-pro
pico-eu
cocco
diazo
diatom
dino
dominant phytoplankton group
−130
−128
−126
−124
−122
longitude

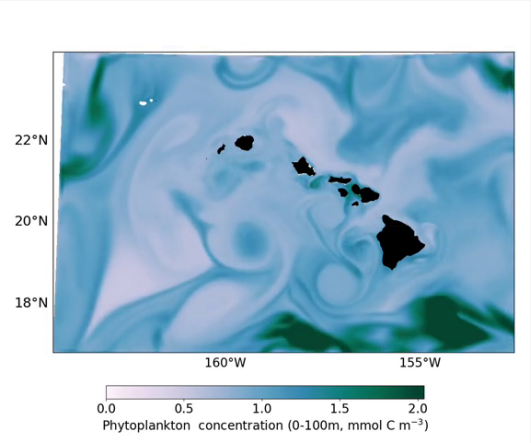
Hawai'i domain

- This is work by Brian Powell and Tobias Friedrich at the Univerity of Hawai'i who have been working with a 2km resolution domain around the Hawaiian islands.
- Two year Darwin 6P4Z simulation with model physics based on assimilative 10 year reanalysis. Boundary conditions for Darwin variables are based on monthly climatology.
- Station Aloha is in the model domain and allows comparison with Hawaiian Ocean Time-series (HOT) data.

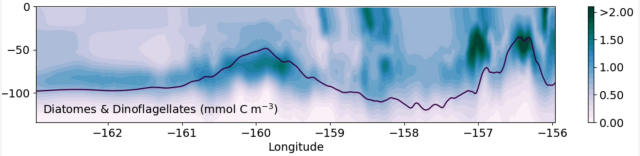


Hawai'i domain

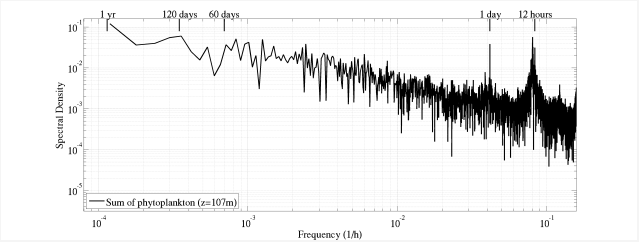
- This is work by Brian Powell and Tobias Friedrich at the Univerity of Hawai'i who have been working with a 2km resolution domain around the Hawaiian islands.
- Two year Darwin 6P4Z simulation with model physics based on assimilative 10 year reanalysis. Boundary conditions for Darwin variables are based on monthly climatology.
- Station Aloha is in the model domain and allows comparison with Hawaiian Ocean Time-series (HOT) data.



- Cyclonic eddies are creating blooms in the wake of the Big Island.



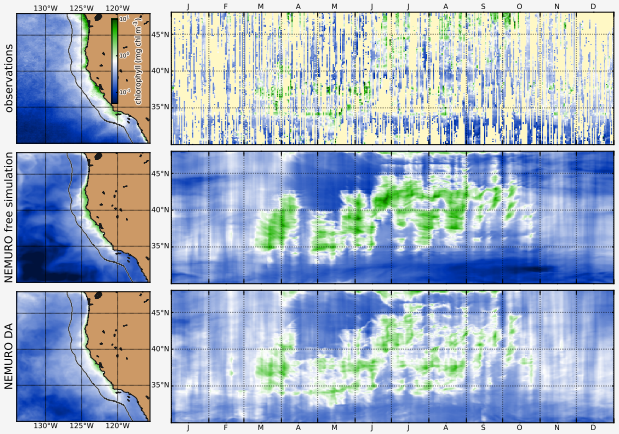
- Physical-biological interactions including the solar cycle and tides play a major role in the variability of total phytoplankton.



- Next step will be to investigate the impact of the physical-biological interactions on the biogeochemistry.

- We have successfully applied 4DVar-based data assimilation to the medium complexity (2P3Z) NEMURO model.

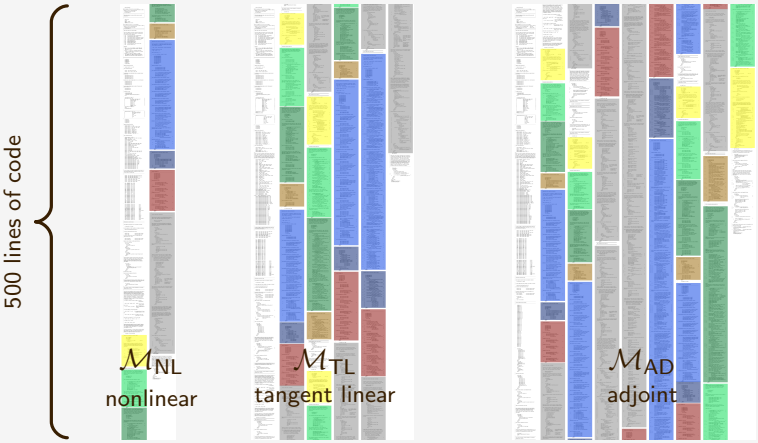
surface chlorophyll *a*



Mattern et al., OM (2017)

motivation: 4DVar physical+biogeochemical data assimilation for the NEMURO model

- We have successfully applied 4DVar-based data assimilation to the medium complexity (2P3Z) NEMURO model.
- But ...



- 4DVar data assimilation requires tangent linear and adjoint code which is difficult and cumbersome to create and maintain.
- We don't want to create adjoint code for Darwin by hand.

alternatives to 4DVar for Darwin state estimation

ensemble-based

EnKF

4DEnsVar ✗

A4DVar

bioEOF-based 4DVar

dual number-based 4DVar

variational

We have been testing adjointless and hybrid approaches that combine ensemble and variational DA:

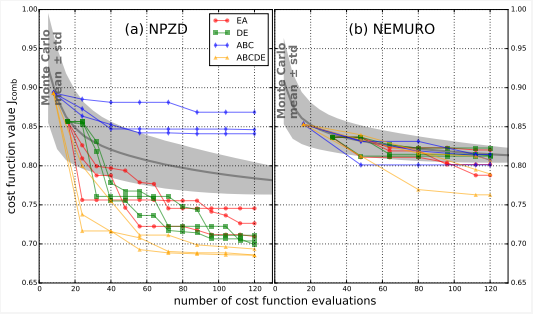
- EnKF: Naive version has been implemented, reliant on a good ensemble, good free run results. Jasen Jacobsen is helping to evaluate the ecosystem dynamics and physical-biological interactions in the free run.
- 4DEnsVar: Disappointing initial results, abandoned in favor of A4DVar.
- A4DVar: *Adjointless* 4DVar, requires ensemble. Easy to implement, being tested right now.
- bioEOF-based 4DVar: Very fast and promising initial results by Jonathan Izett, still in early testing.
- dual number-based 4DVar: Simple way to create exact adjoint code but slow for complex models, would require lower resolution grid for data assimilation (planned feature in future version of ROMS).

Darwin parameter estimation

Variational and especially ensemble-based data assimilation benefits from improved non-assimilative Darwin performance in the ROMS domain.

- Use statistical relationships between physical variables and nutrients to improve nutrient initial and boundary conditions.
- Estimate improved values for key Darwin plankton parameters (small amounts of manual tuning have already yielded good improvements).
- Start out with a simple random sampling approach, then employ evolutionary techniques like ABCDE for parameter estimation.

NPZD: 9 parameters NEMURO: 43 parameters



Mattern & Edwards, JMS (2017)

EA: evolutionary algorithm

DE: differential evolution

ABC: artificial bee colony

ABCDE: artificial bee colony *using* differential evolution

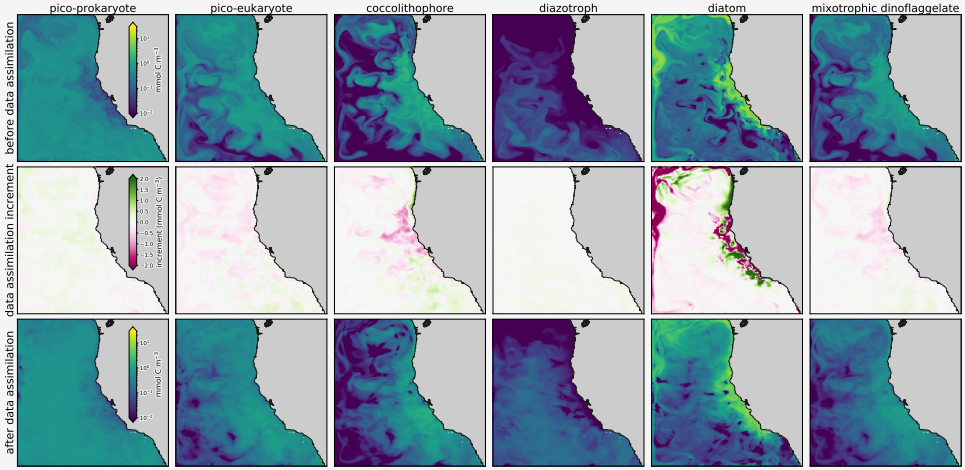
lack of data: challenge and potential solutions

Lack of data is probably the largest challenge to assessing and constraining the wealth of biological variables contained in the Darwin model, especially the different types of plankton variables and their distributions.

- CalCOFI data for the CCS has limited extent, not enough plankton information for a quantitative assessment.
- Size-structured ocean color data for the north pacific (Shubha Sathyendranath and Bob Brewin are part of the CBIOMES team).
- LIR statistical model for relationship between physical and some nutrient variables (Carter et al, L&O (2018)).
- BioARGO?

Ensemble Kalman Filter state estimation – April 1, 2007

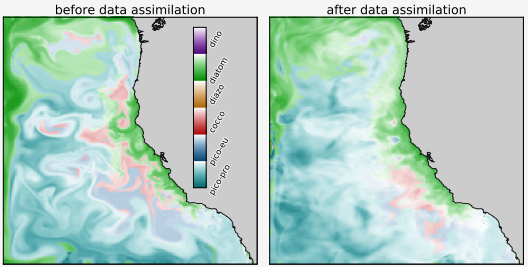
Results of an initial 2-day data EnKF data assimilation cycle, assimilating surface chlorophyll *a* satellite data, jointly with physical observations:



In the first cycle, diatoms, which are blooming along the California coastline, receive the largest adjustment from the EnKF (column 5).

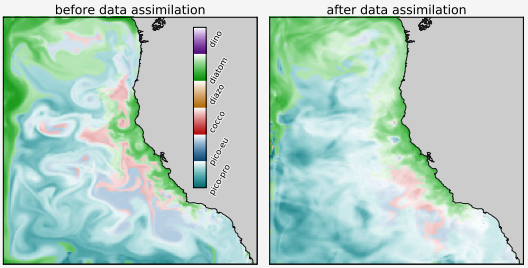
Data assimilation impact on the Darwin ecosystem structure

the dominant phytoplankton group in the model



We observe no large scale changes in the dominant phytoplankton group through the EnKF increment.

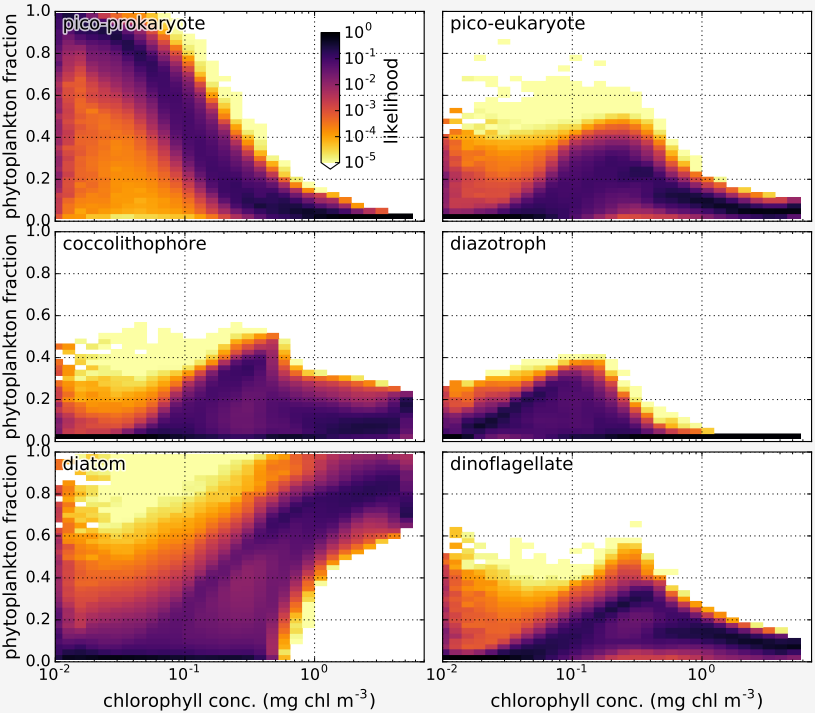
the dominant phytoplankton group in the model



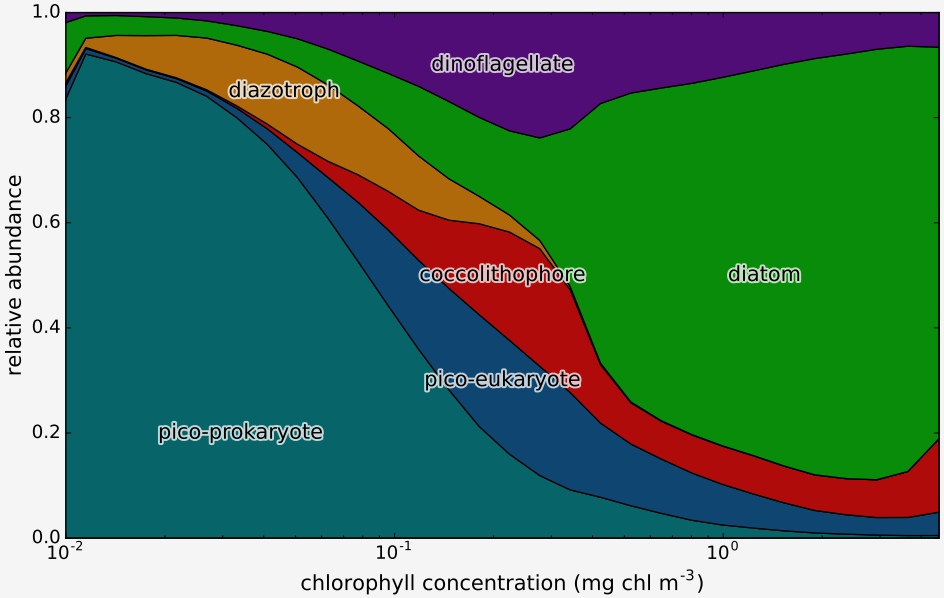
We observe no large scale changes in the dominant phytoplankton group through the EnKF increment.

Thanks!

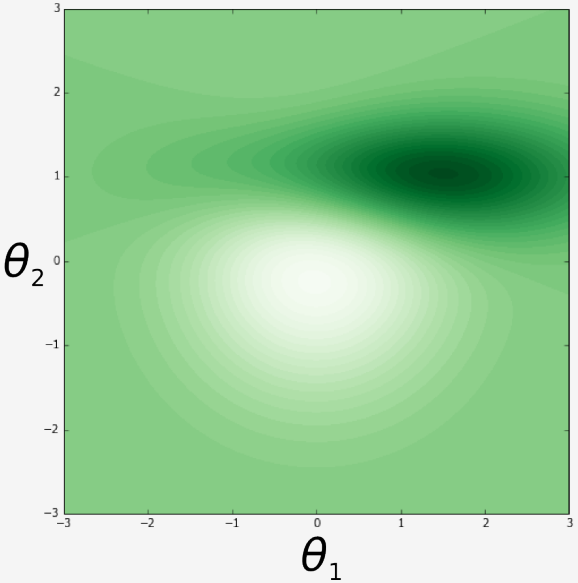
additional slides


pico-prokaryote
pico-eukaryote
coccolithophore
diazotroph
diatom
dinoflagellate
likelihood
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
phytoplankton fraction
0.0
0.2
0.4
0.6
0.8
1.0
10^{-2}
10^{-1}
10^{0}
chlorophyll conc. (mg chl m^{-3})

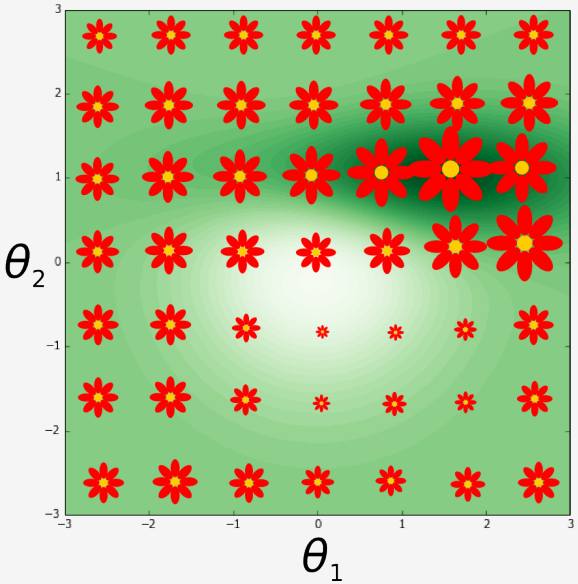
chlorophyll phytoplankton



cost function in parameter space:

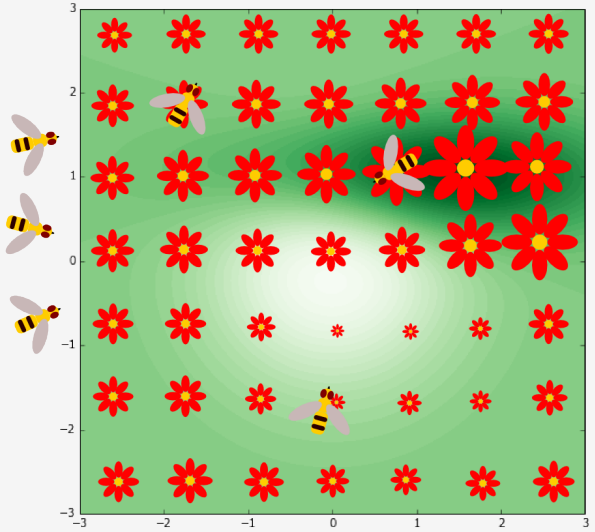


cost function in parameter space:



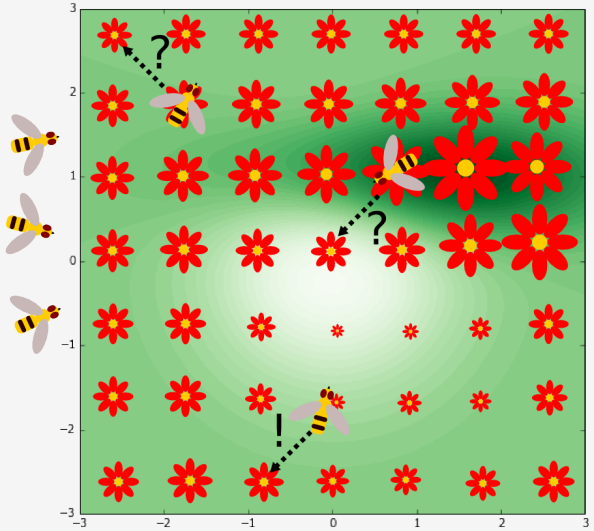
ABC – worker bees and onlookers

cost function in parameter space:



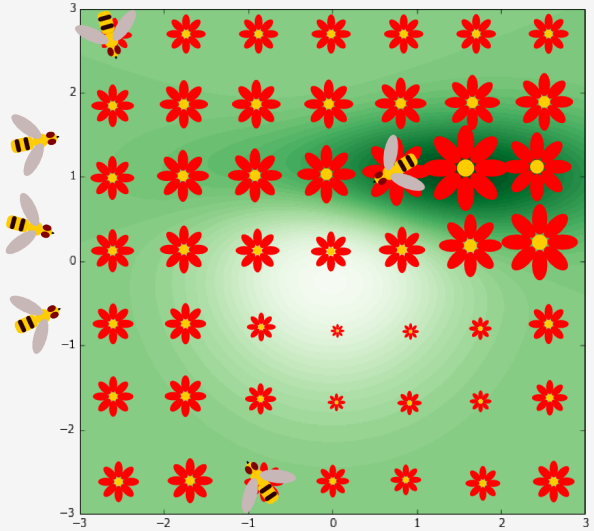
ABC – workers sampling close to their position

cost function in parameter space:



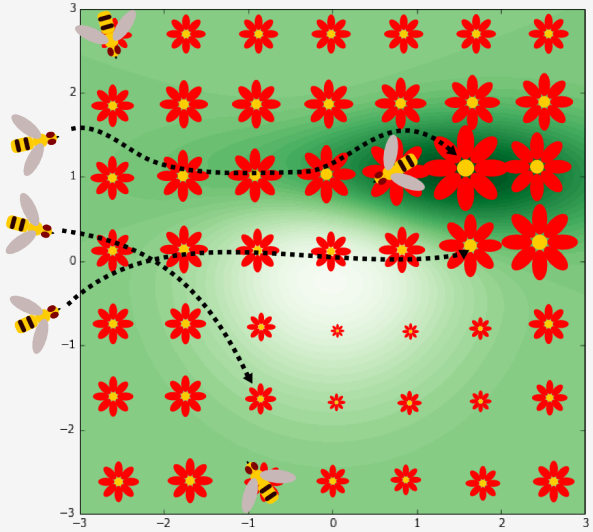
ABC – workers sampling close to their position

cost function in parameter space:



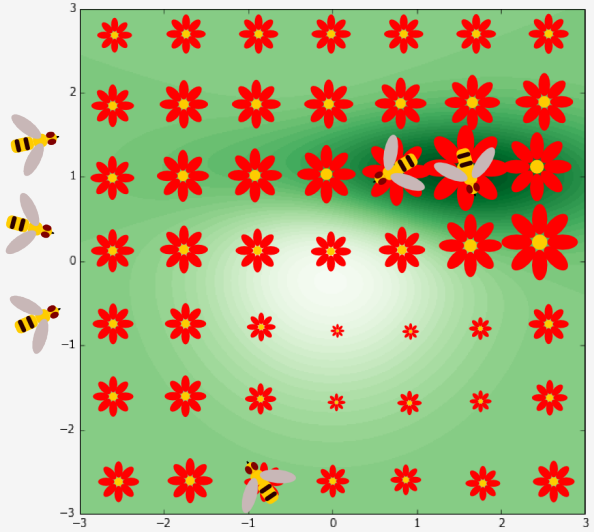
ABC – onlookers sample close to promising positions

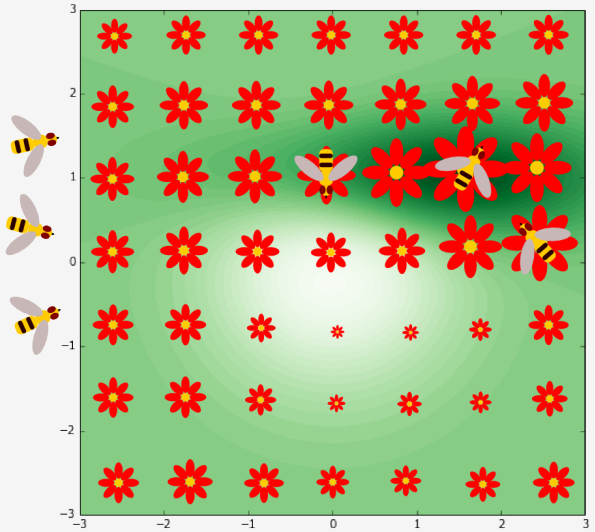
cost function in parameter space:



ABC – onlookers sample close to promising positions

cost function in parameter space:



ABC – end result

cost function in parameter space: