Machine Learning as an Antidote to Chaos: Decoding Ecological Complexity

Maike Sonnewald (maike_s@mit.edu)[1,2], Stephanie Dutkiewicz[1], Chris Hill[1] and Gael Forget[1]
1: Massachusetts Institute of Technology 2: Harvard University

1.Framework

A framework is presented to establish ecological niches. Niches are sorted according to plankton community structure similarity. Using the Longhurst provinces as a benchmark, the presented framework is both:

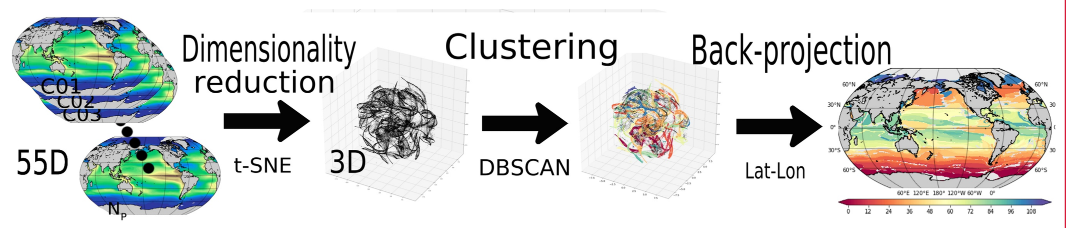
1: Capable of global analysis

2: Nested to be relevant regionally and globally

Determining provinces in the pelagic ocean has relied on parameters such as SST, PAR, Chlorophyll and expert assessment. The presented framework uses instead the local planktonic community structure, and highlights that even when there is similar total biomass (or Chl), very different ecological compositions are found even at the coarsest nesting grade (benchmarked by Longhurst).

The framework uses output from a physical/ecosystem model. The physical setup is the ECCO state estimate at 1° resolution. We use the biomass of the 51 model plankton species and the nutrient fluxes of N,P,Si and Fe to define the eco-provinces. This 55D dataset is is overwhelmingly complex when analyzed naively. First we sort the plankton into 7 functional groups, and sum their biomass so as to reduce to a 11D space.

Fig.1



Further dimensionality reduction using t-SNE highlighted key regions projecting the 11D (Q) only 3D (P) by maximizing the Kullbach-Leibner (KL) distance between the probability that points would be close according to:

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

The 3D features are then clustered using the unsupervised machine learning method DBSCAN setting the min points in a cluster and the max distance (ε) between them:

1. Select point at random
2. Count nr. of immediate neighbors within distance ε
3. The cluster boundary is determined repeating step 2 iteratively. If "min" is reached a cluster is identified.
4. A new point is chosen, repeating 1-4.

The ε and "min" are chosen using connectivity and coverage arguments seen in Fig.2 and 3 respectively (below, red dot).

Fig.2

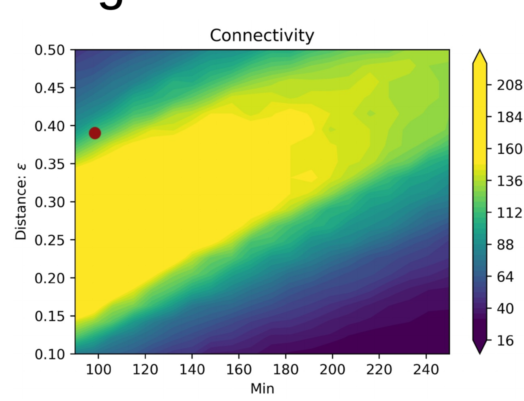
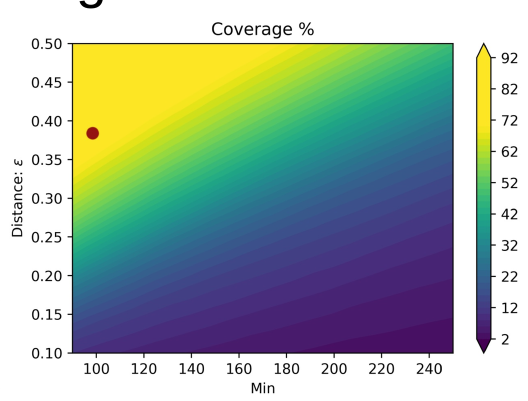


Fig.3



2.Global Niches

Fig.4

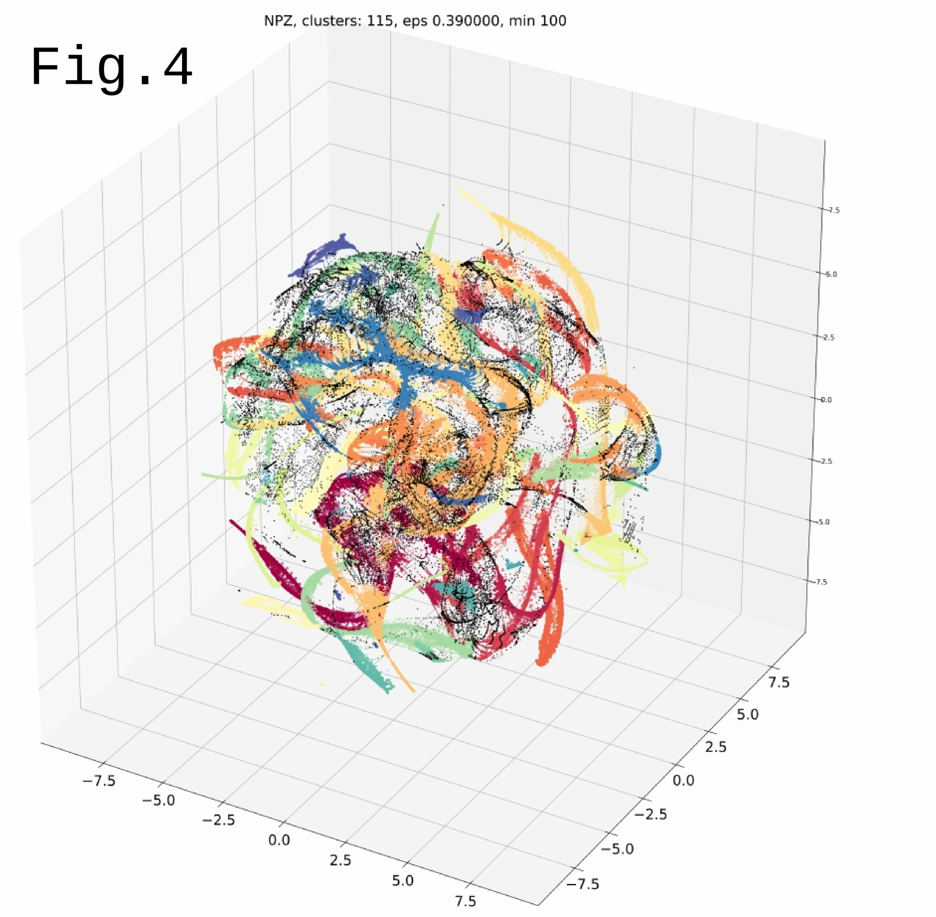
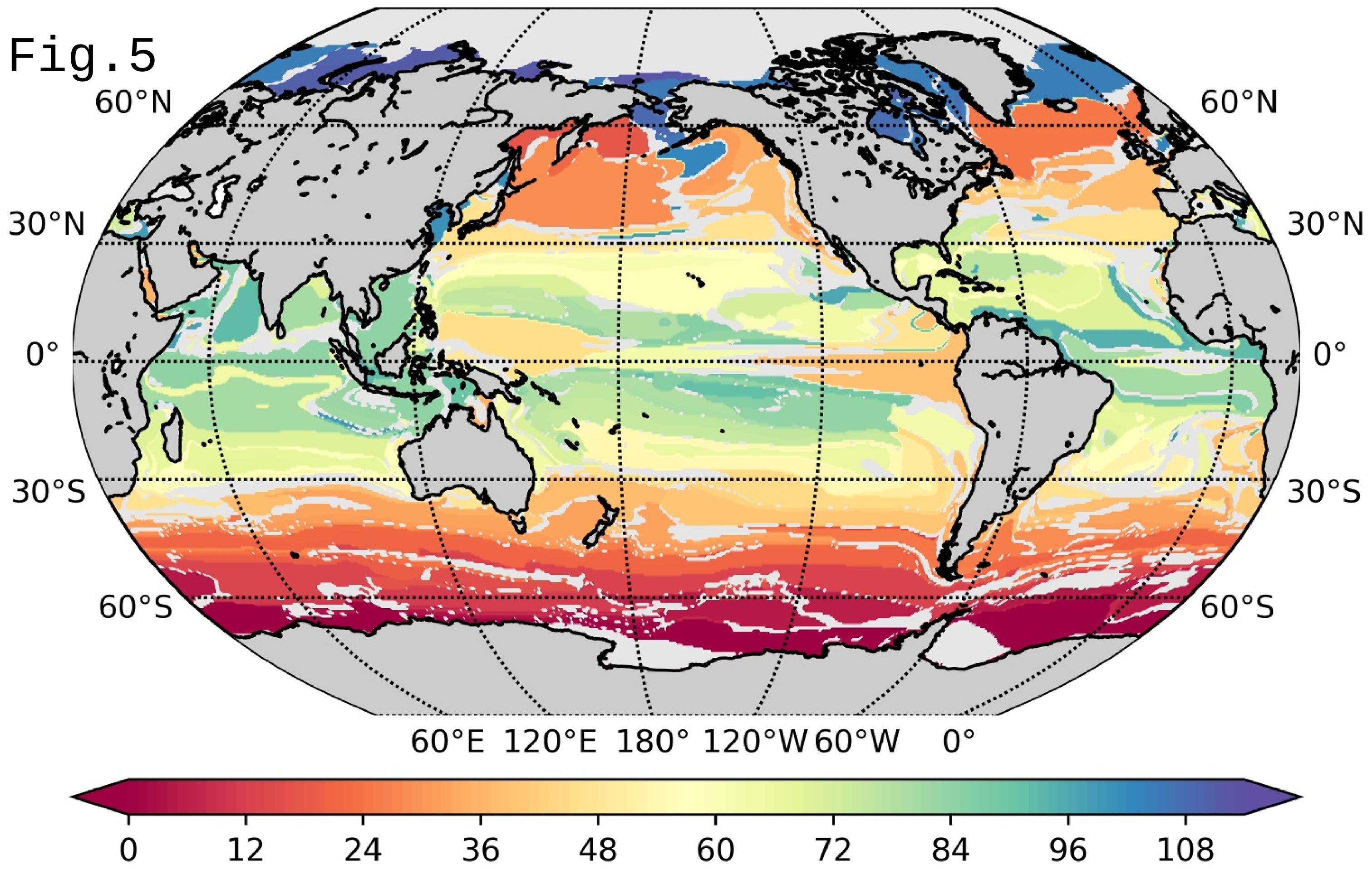


Fig.5



- The data is clustered into distinct regions, with each lat-long occupying one point in Fig.4 (above left). Colors represent found clusters.

- The 3D data in Fig.4 (above left) is projected back onto the global lat-long in Fig.5 (above right). Points not assigned to a cluster are left blank, as well as points of persistent ice cover or low mean biomass.

- t-SNE is probabilistic, so a repeat x10 of Fig.1 method assures robustness.

- The framework determines a full complexity of ca 120 clusters/niches over 10 realizations, one is depicted in Fig.5 (above right).

- Intricate mosaics emerge, with global coverage. Distinctions of zonally coherent, equatorial upwelling and oligotrophic gyre regions are apparent.

- Areas appearing as "noise", are areas of large variability/streaked areas in Fig.4 (e.g. seasonal), or too unique/small to be captured.

3.Ecological sorting

Due to the large number of niches (n), it is difficult to work with these clusters. To reduce the complexity, the provinces can be sorted according to ecological similarity.

The inter-niche Bray-Curtis-dissimilarity (Fig.6 above) for every niche to every other niche is determined according to the biomass of each of the 51 species (C) and total biomass (S):

$$BC_{n_i n_j} = 1 - \frac{C_{n_i n_j}}{S_{n_i} + S_{n_j}}$$

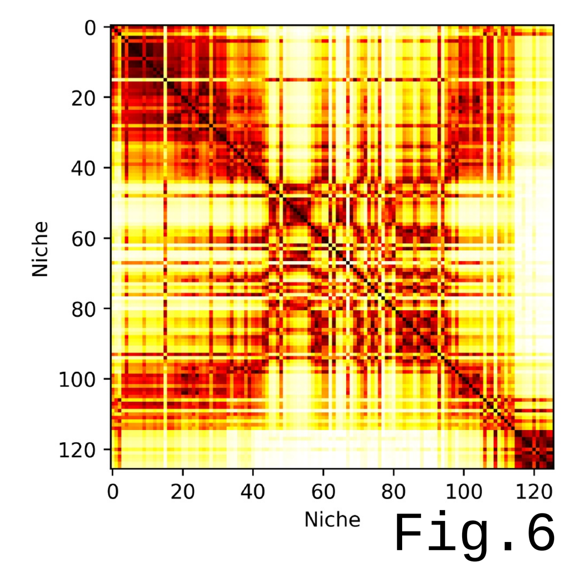


Fig.6

Based on the similarity matrix (Fig.6) the system complexity can be reduce to one province, two province etc (Fig.7), for the 10 realisations. A minimum of 8 clustered provinces the intra-province Brey-Curtis provides the same mean dissimilarity as using the Longurst provinces (Fig.8). Increasing complexity thus offers an improvement using the concept of ecologically driven clustering. Fig.9 and 10 illustrates the intra-province BC for full complexity and for using 12 provinces.

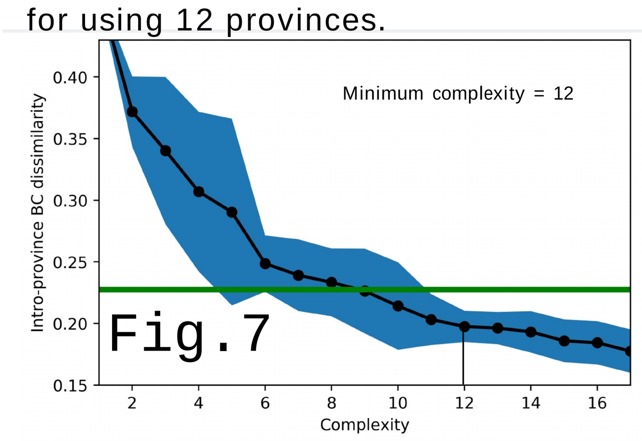


Fig.7

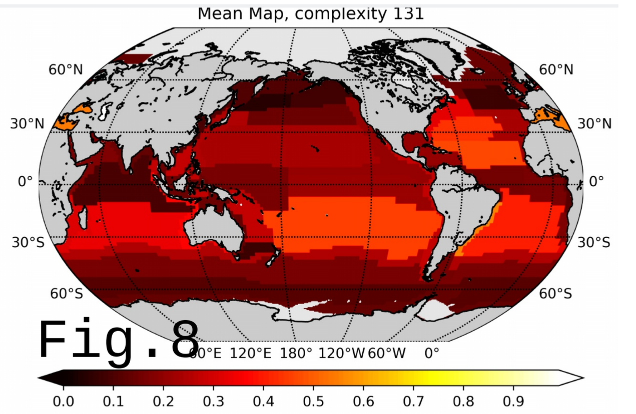


Fig.8

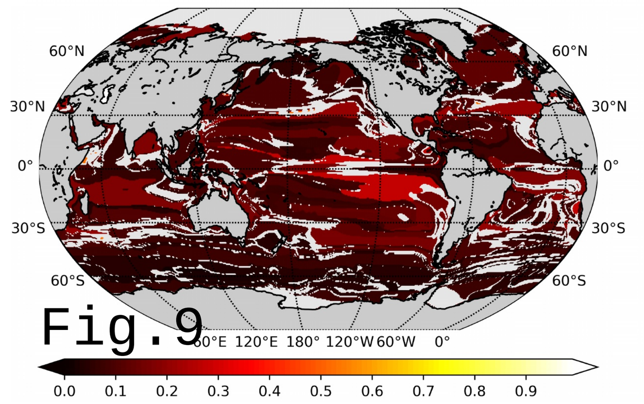


Fig.9

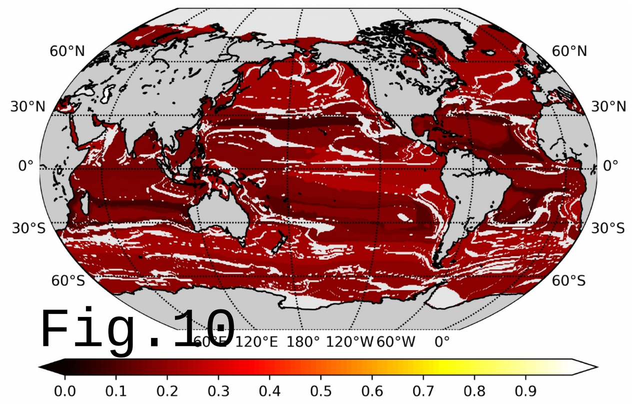
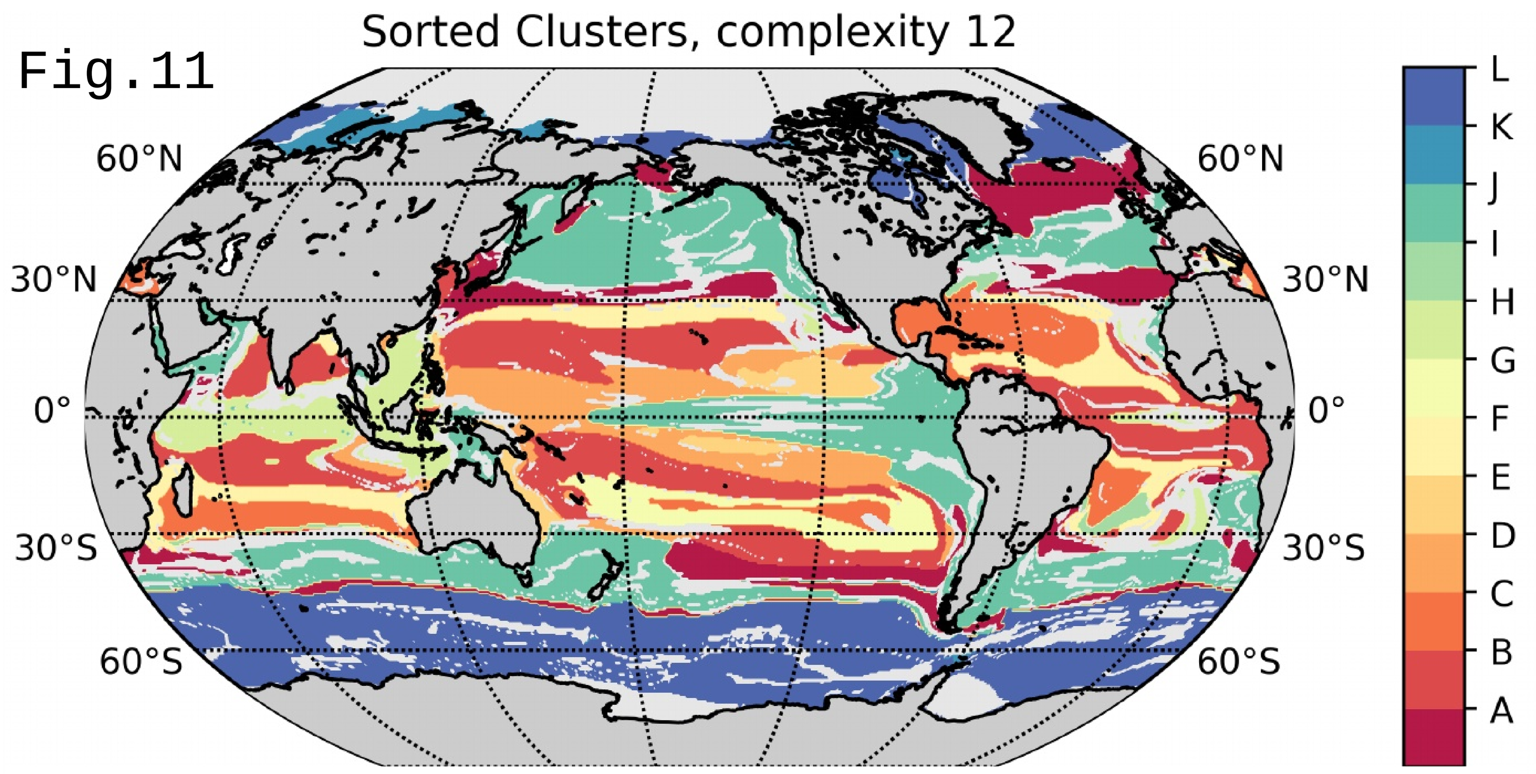


Fig.10

4.Decoded Complexity: Global Provinces

Fig.11



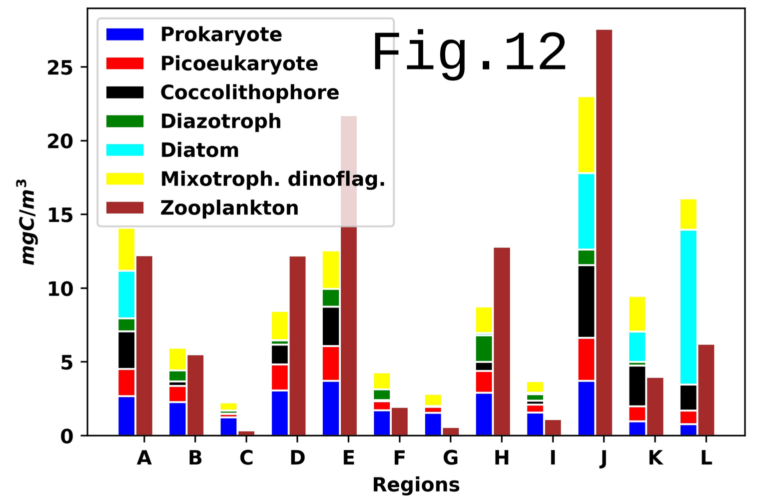


Fig.12

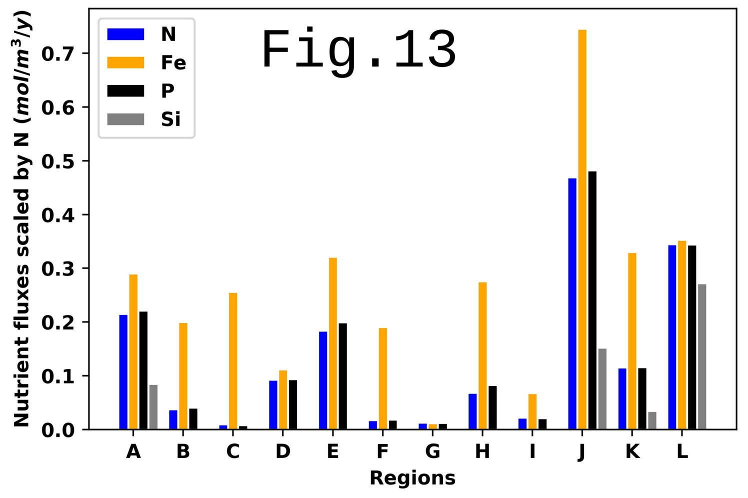


Fig.13

- Fig.11 illustrates the spatial extent of complexity 12, arbitrarily colored, determined from Fig.7.

- The low StD in Fig.7 illustrates that all 10 realizations converge once a complexity >12 is used.

- The area determined by the provinces A-L are assessed and the ecological assemblage and nutrient fluxes examined in Fig.12 and 13 respectively. Fig.13 scales the nutrient fluxes accorfing to the Redfield ratio 1:1:16:16x10^{-3}.

- Identified provinces are all unique.

- These ecological motivated provinces indicate that:
  1) Provinces may have similar community structures, but very different total biomass (or Chl) and very different grazers biomass (e.g. D and E).
  2) Provinces may have very different total biomass (or Chl), but similar community structures (e.g. D and K)

- Province utility includes assessing controls on community structure. For instance:
  1) The relative amount of diatoms is set by the imbalance in the Si to N,P, Fe supplies. With balanced supplies the community is diatom dominated (L) and where they are less balanced diatoms comprise only a smaller fraction (K)
  2) Diazotrophs survive where the Fe and P supplies are in excess of the N supplies (e.g. E and H).

- Biomass is seen to be a poor predictor of of zooplankton (e.g. D+E or K+L).

5.Take Home

- The presented framework successfully decodes the chaotic pelagic modeleled ocean.

- Unique and robust niches were determined, subsequently sorted into provinces, where a complexity of 12 provinces is recomended.

- The presented framework is both global, and can be nested down to the full complexity for more regional applications.

- The framework reaveals a detailed community structure showing:
  1: Similar provinces can have the same biomass, but radically different species assemblages.
  2: Regions with similar species assemblages can differ in productivity by orders of magnitude.

- Biomass is seen to be a poor predictor of of zooplankton.

- Regional and higher resolution efforts will use ecological fingerprinting to extend this work, as well as determining if unique optical identifiers exists.

Paper+poster:
github.com/maikejulie/publications