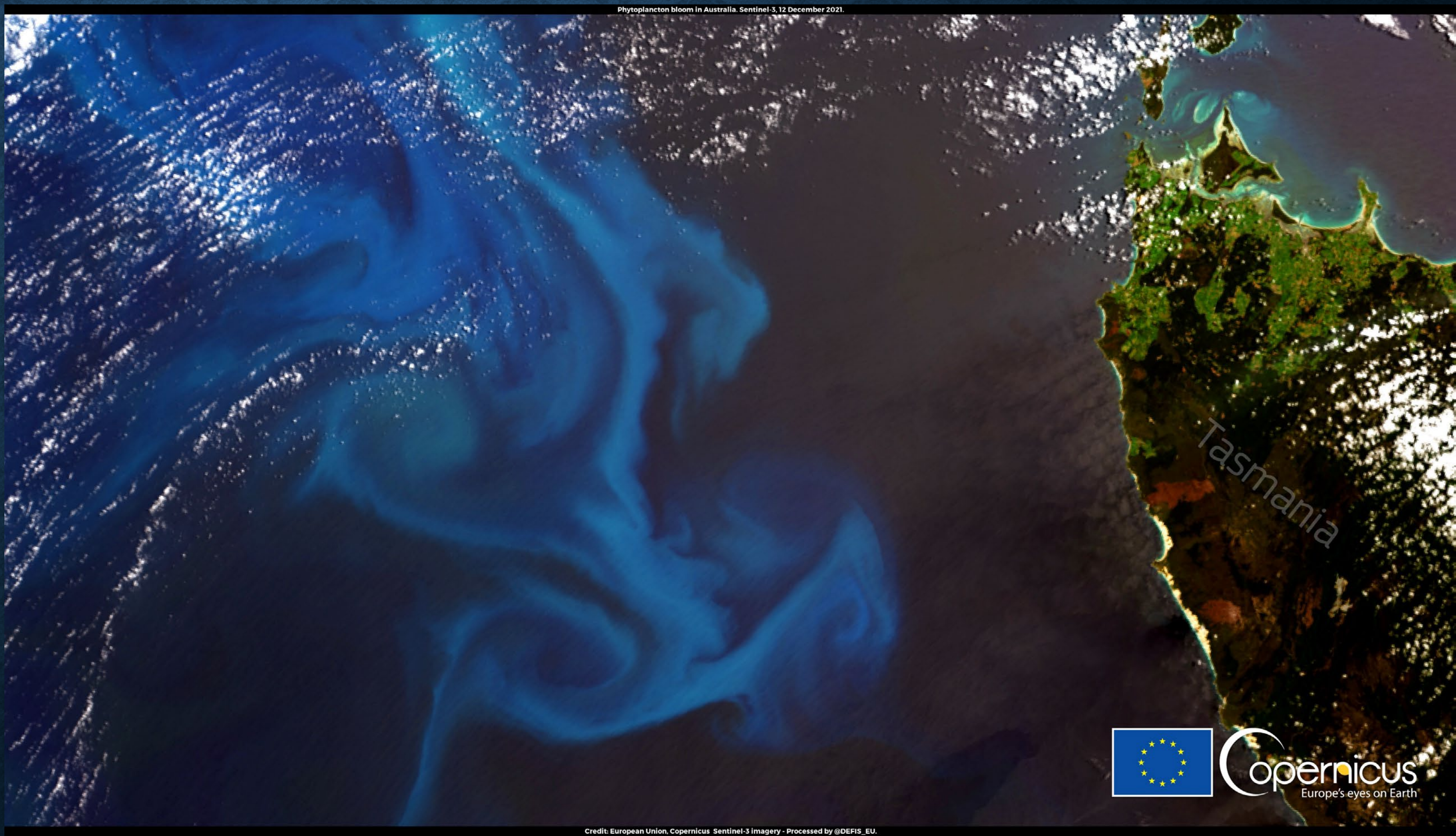
Using Machine Learning to uncover Ecological Mechanisms controlling abundance of Phytoplankton Size Classes from Large-scale observations

- Phytoplankton bloom on 12 December 2021

- Captured by Copernicus Sentinel-3 satellite about 100 Km off the coast of western Tasmania (Australia)



Why use Machine Learning?

- Intrinsic relationships
- Apparent relationships
- Apparent relationships difficult to interpret

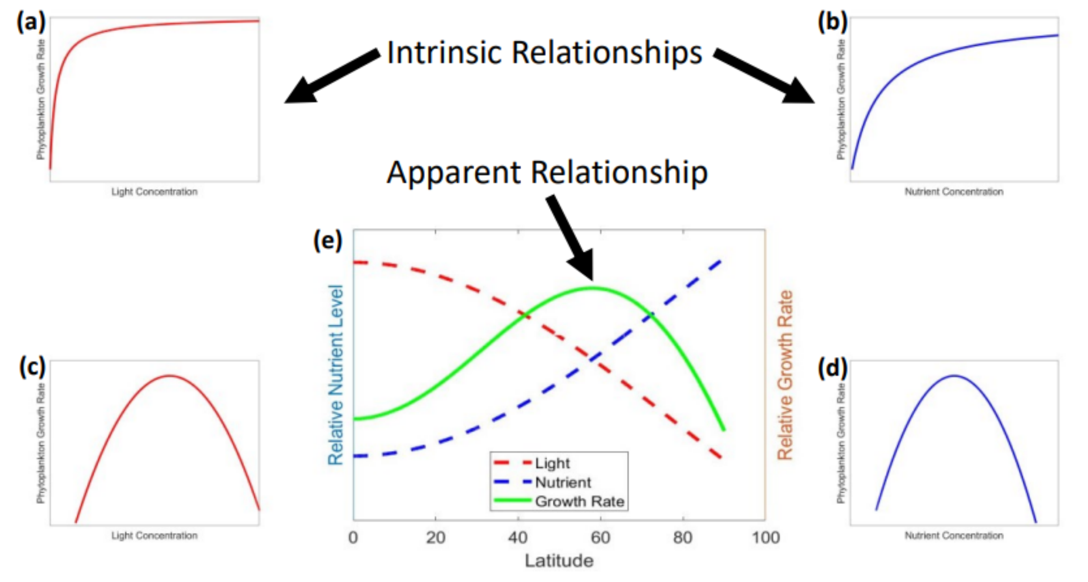


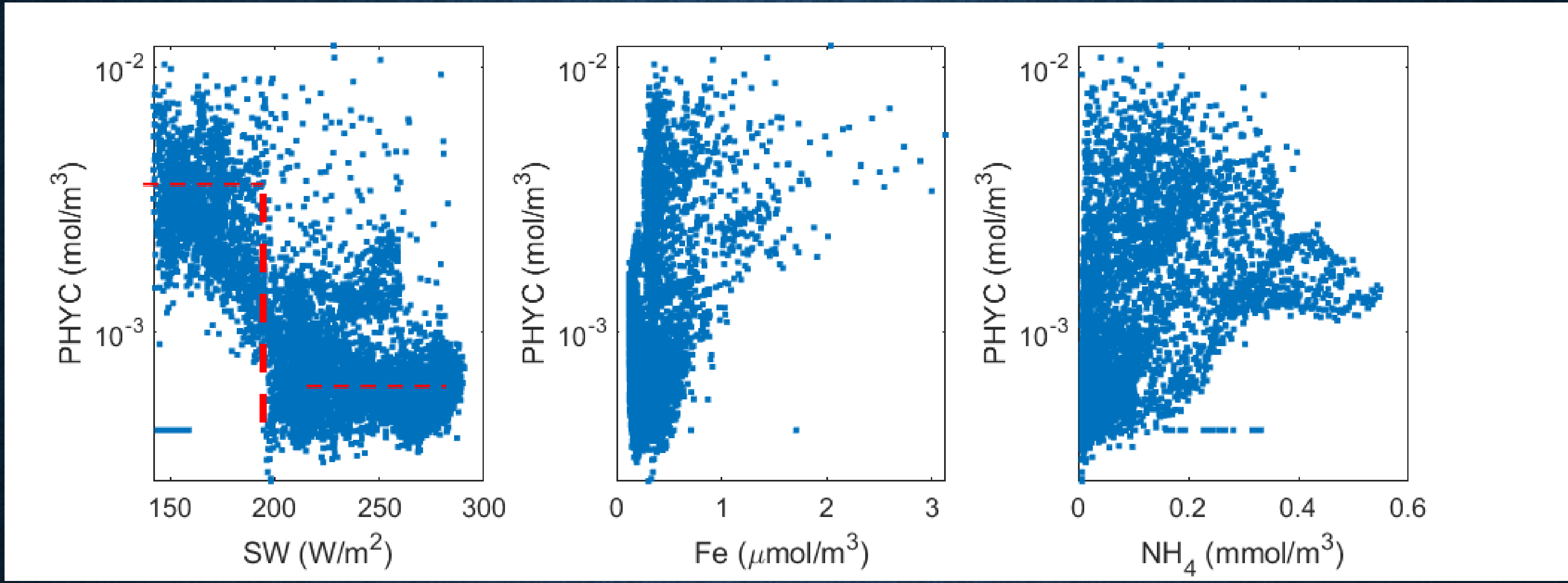
Figure 1.1: Diagram showing the concept of intrinsic and apparent relationships. The top plots are Michaelis-Menten curves with **a** showing the intrinsic relationship between light and phytoplankton growth rate and **b** showing the intrinsic relationship between an arbitrary nutrient and phytoplankton growth rate. The distribution for light across the Northern latitudes is represented in **e** as the red dashed line and the distribution for the nutrient is the blue dashed line. When the distributions from **e** for light and the nutrient are used as input in the intrinsic relationships, the interaction of light and the nutrient produce the apparent relationship represented as the solid green line **e**. The apparent relationship would not necessarily be intuitive if the light and nutrient concentrations were plotted against the phytoplankton growth rate, as seen in **c** and **d**.

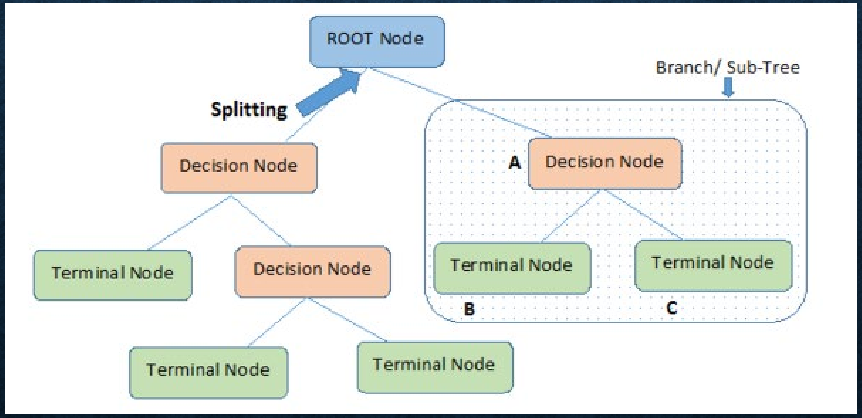
Holder & Gnanadesikan, 2021

Methodology

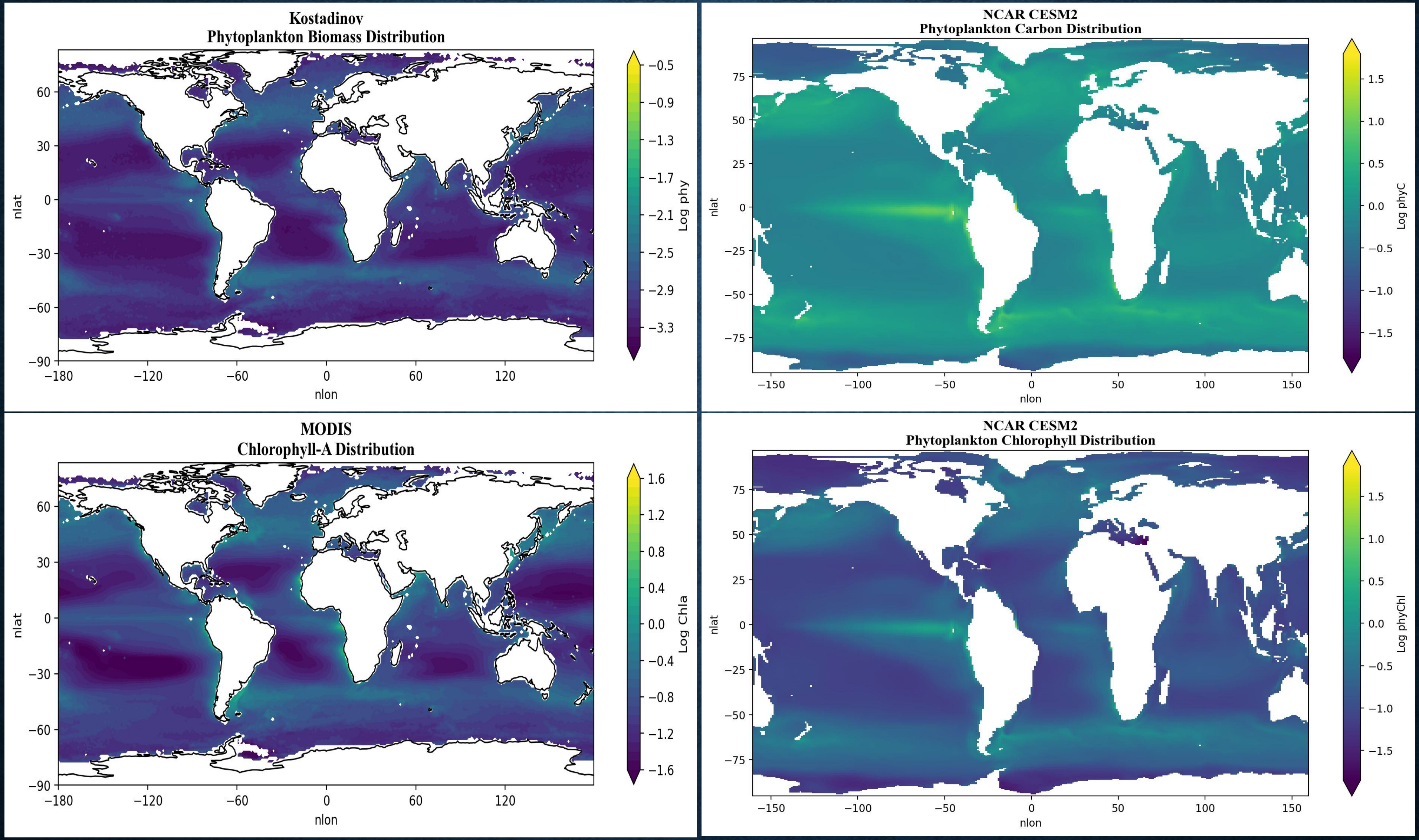
- SeaWiFS, Copernicus (Sentinel-3A), MODIS (Aqua)
- Temperature, salinity, mixed layer depth, silicate, phosphate, nitrate - World Ocean Atlas (WOA) 2018
- Shortwave radiation - International Satellite Cloud Climatology Project (ISCCP)
- Upwelling - Estimating the Circulation and Climate of the Ocean (ECCO) reanalysis data
- Dissolved iron, ammonia(nh4) - ensemble average of ESMs
- RF Regressor trained to predict size classes
- Permutation Importance analysis
- Median Replacement Anomaly analysis

Decision Tree

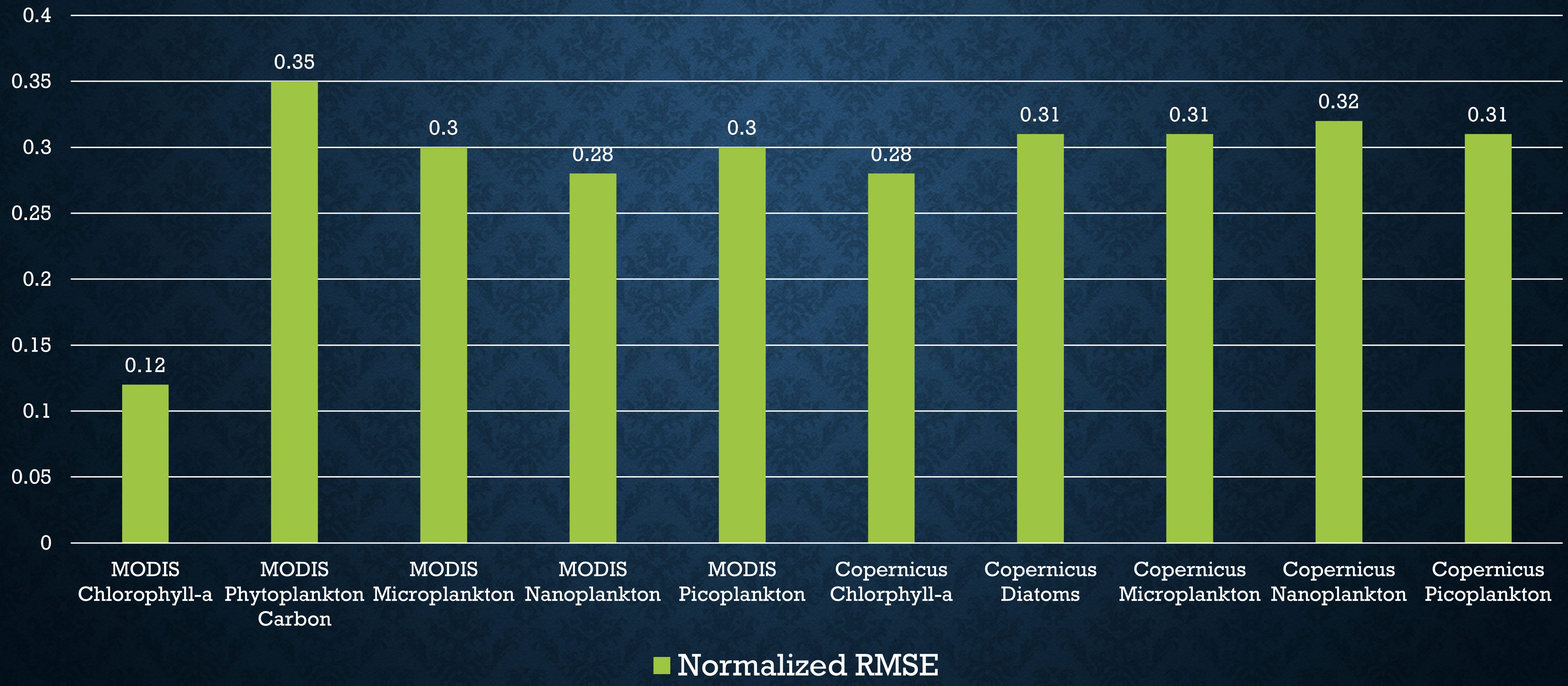




Total Phytoplankton

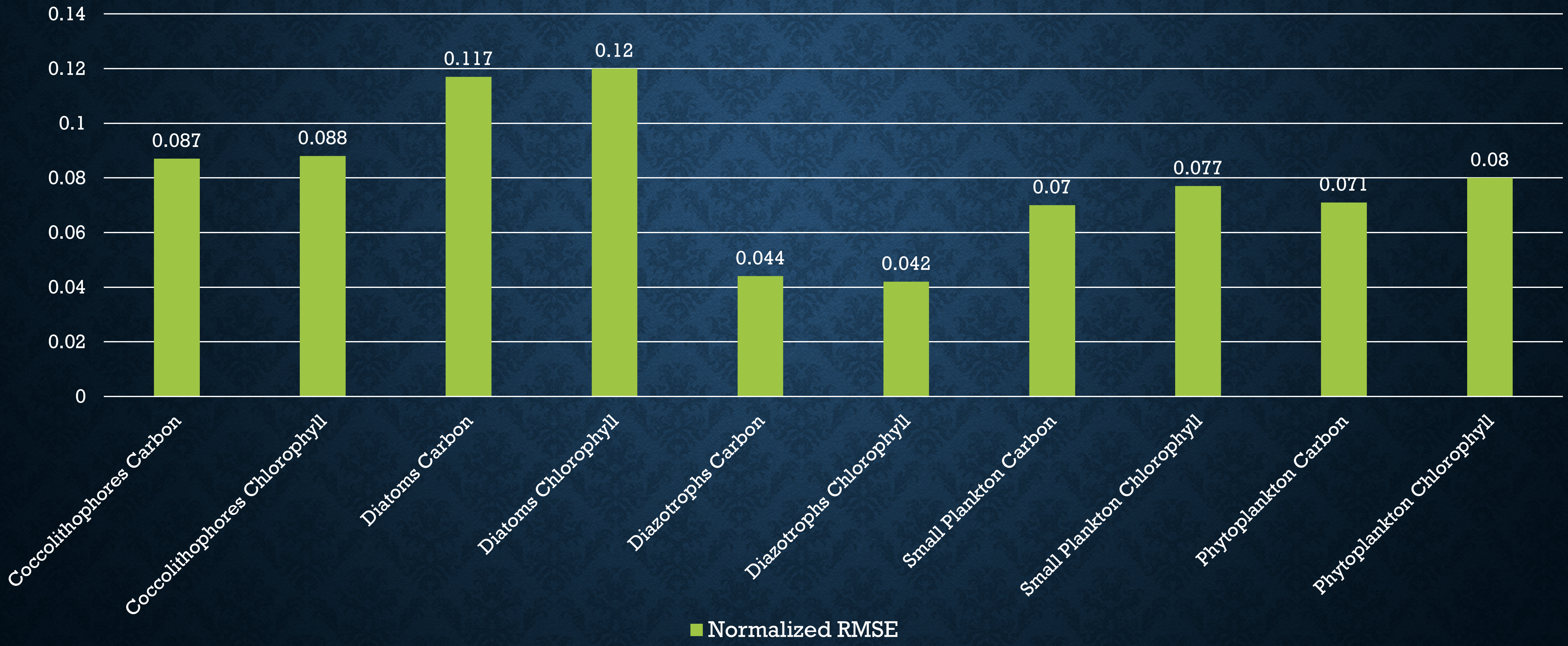


Random Forest : Observations



R2 Score between 0.87 to 0.92. Explains 87% - 92% of the variance

Random Forest : CESM2

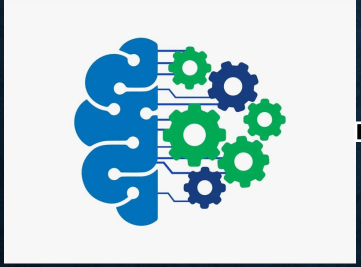


R2 Score between 0.98 to 0.99. Explains 98% - 99% of the variance

Permutation Importance

month	lat	lon	dfe	mld	nh4	no3	po4	rsn	sal	sst	sil	w50	chla	logchla
1	-76.5	-179.5	1.95E-07	62.589	0.00056	0.018248	0.001395	169.4385	34.2867	-0.16111	0.062234	-5.7E-07	2.245918	0.351394
2	-76.5	-179.5	2.3E-07	78.447	0.000796	0.019298	0.00154	109.5923	34.31819	-0.27521	0.063704	-1.2E-07	0.881226	-0.05491
3	-76.5	-179.5	3.14E-07	106.181	0.00059	0.023357	0.001321	22.15769	34.31021	-1.33841	0.060868	2.64E-07	0.031147	-1.50658
4	-76.5	-179.5	4.02E-07	307.635	0.000202	0.027729	0.002066	0	34.33461	-1.81021	0.072237	9.54E-07	0.031147	-1.50658
5	-76.5	-179.5	4.7E-07	394.284	7.28E-05	0.028658	0.002214	0	34.34449	-1.88471	0.074234	2E-06	0.031147	-1.50658
6	-76.5	-179.5	5.17E-07	406.444	4.35E-05	0.027836	0.002124	0	34.4055	-1.88821	0.072959	2.72E-06	0.031147	-1.50658
7	-76.5	-179.5	5.52E-07	427.798	3.48E-05	0.021078	0.001599	0	34.40779	-1.85941	0.064651	2.36E-06	0.031147	-1.50658
8	-76.5	-179.5	5.81E-07	403.002	3.11E-05	0.021078	0.001599	0	34.42471	-1.76001	0.064651	2.12E-06	0.031147	-1.50658
9	-76.5	-179.5	6.06E-07	424.179	2.97E-05	0.021078	0.001599	0	34.42569	-1.87141	0.064651	1.89E-06	0.031147	-1.50658
10	-76.5	-179.5	6.14E-07	369.447	2.78E-05	0.031886	0.002204	0	34.43959	-1.84531	0.077084	1.54E-06	0.031147	-1.50658

Random Shuffling

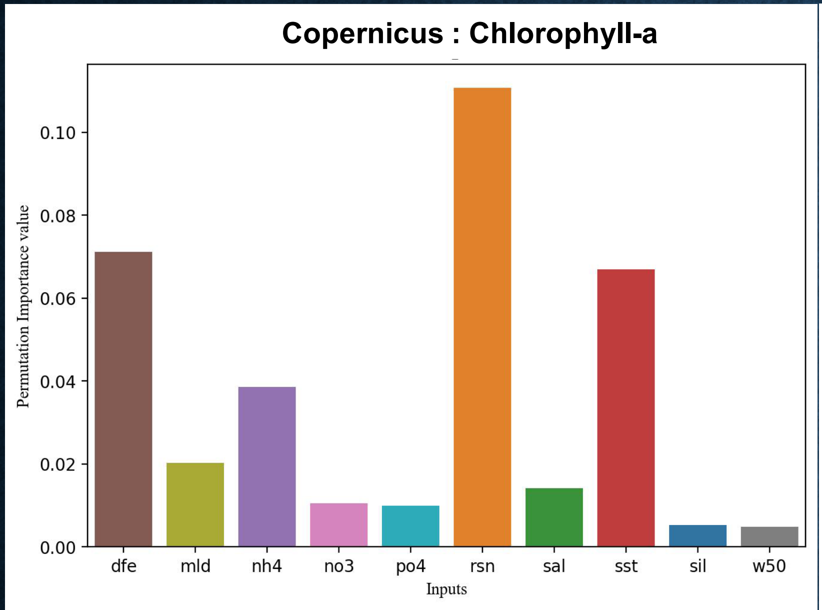


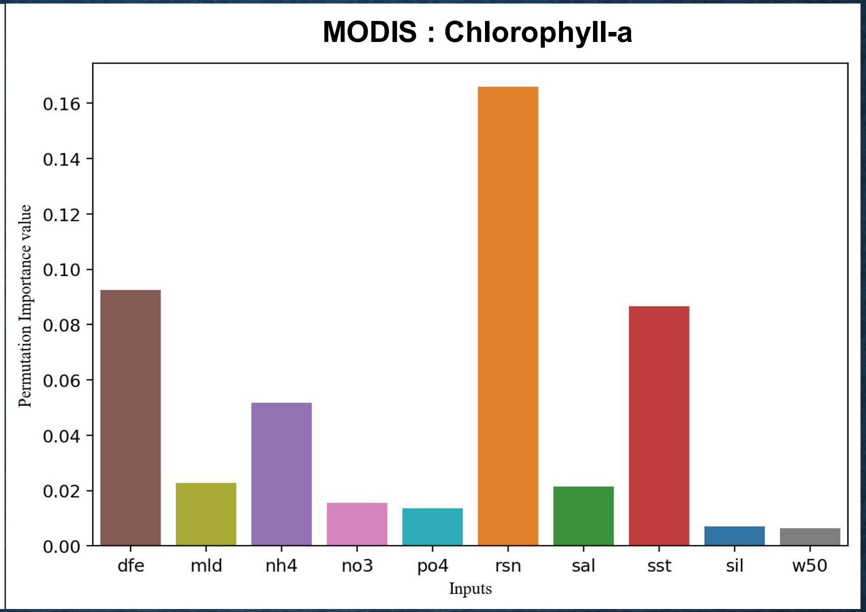
Prediction

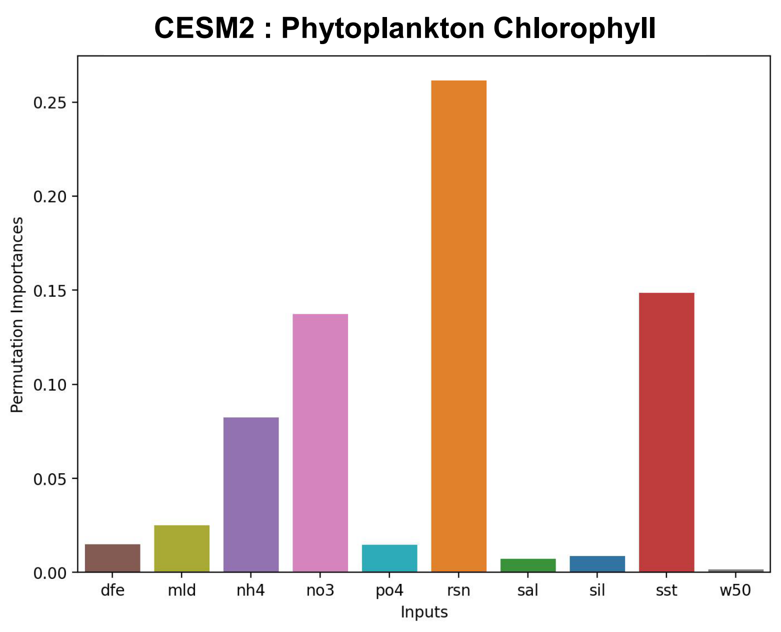
Difference with actual Prediction

Measure of Input Importance

Permutation Importance

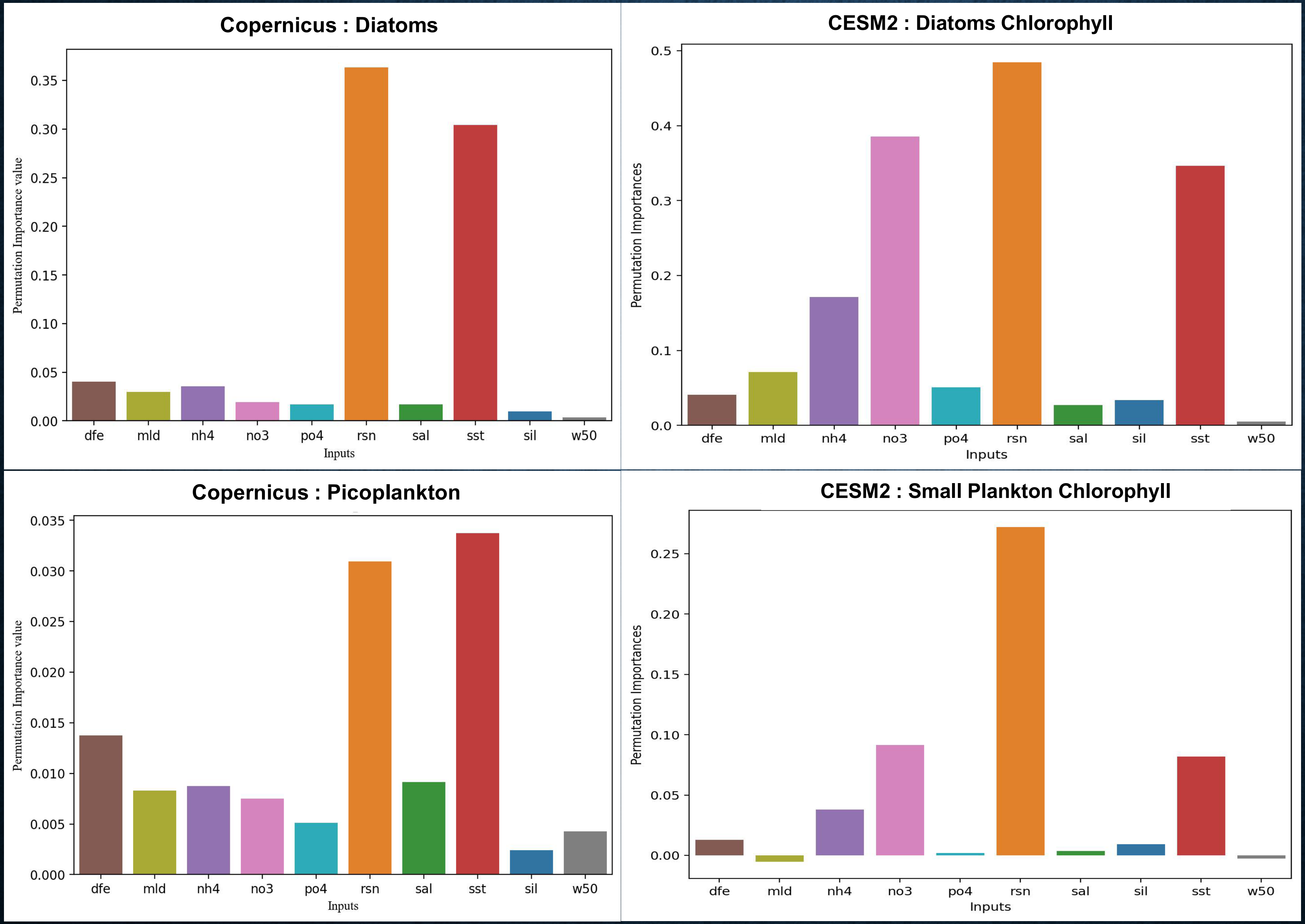






- dfe: iron
- mld: mixed layer depth
- nh4: ammonia
- no3: nirate
- po4: phosphate
- rsn: sw radiation
- sal: salinity
- sst: sea surface temperature
- sil: silicate
- w50: upwelling at 50m

Permutation Importance

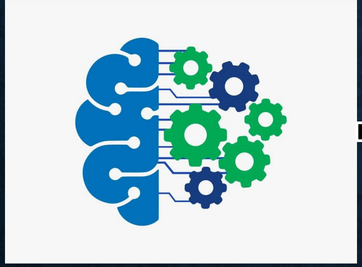


- dfe: iron
- mld: mixed layer depth
- nh4: ammonia
- no3: nirate
- po4: phosphate
- rsn: sw radiation
- sal: salinity
- sst: sea surface temperature
- sil: silicate
- w50: upwelling at 50m

Median Replacement Anomaly

month	lat	lon	dfe	mld	nh4	no3	po4	rsn	sal	sst	sil	w50	chla	logchla
1	**-76.5**	**-179.5**	1.95E-07	62.589	0.00056	0.018248	0.001395	169.4385	34.2867	-0.16111	0.062234	-5.7E-07	2.245918	0.351394
2	**-76.5**	**-179.5**	2.3E-07	78.447	0.000796	0.019298	0.00154	109.5923	34.31819	-0.27521	0.063704	-1.2E-07	0.881226	-0.05491
3	**-76.5**	**-179.5**	3.14E-07	106.181	0.00059	0.023357	0.001321	22.15769	34.31021	-1.33841	0.060868	2.64E-07	0.031147	-1.50658
4	**-76.5**	**-179.5**	4.02E-07	307.635	0.000202	0.027729	0.002066	0	34.33461	-1.81021	0.072237	9.54E-07	0.031147	-1.50658
5	**-76.5**	**-179.5**	4.7E-07	394.284	7.28E-05	0.028658	0.002214	0	34.34449	-1.88471	0.074234	2E-06	0.031147	-1.50658
6	**-76.5**	**-179.5**	5.17E-07	406.444	4.35E-05	0.027836	0.002124	0	34.4055	-1.88821	0.072959	2.72E-06	0.031147	-1.50658
7	**-76.5**	**-179.5**	5.52E-07	427.798	3.48E-05	0.021078	0.001599	0	34.40779	-1.85941	0.064651	2.36E-06	0.031147	-1.50658
8	**-76.5**	**-179.5**	5.81E-07	403.002	3.11E-05	0.021078	0.001599	0	34.42471	-1.76001	0.064651	2.12E-06	0.031147	-1.50658
9	**-76.5**	**-179.5**	6.06E-07	424.179	2.97E-05	0.021078	0.001599	0	34.42569	-1.87141	0.064651	1.89E-06	0.031147	-1.50658
10	**-76.5**	**-179.5**	6.14E-07	369.447	2.78E-05	0.031886	0.002204	0	34.43959	-1.84531	0.077084	1.54E-06	0.031147	-1.50658

Replace with Median

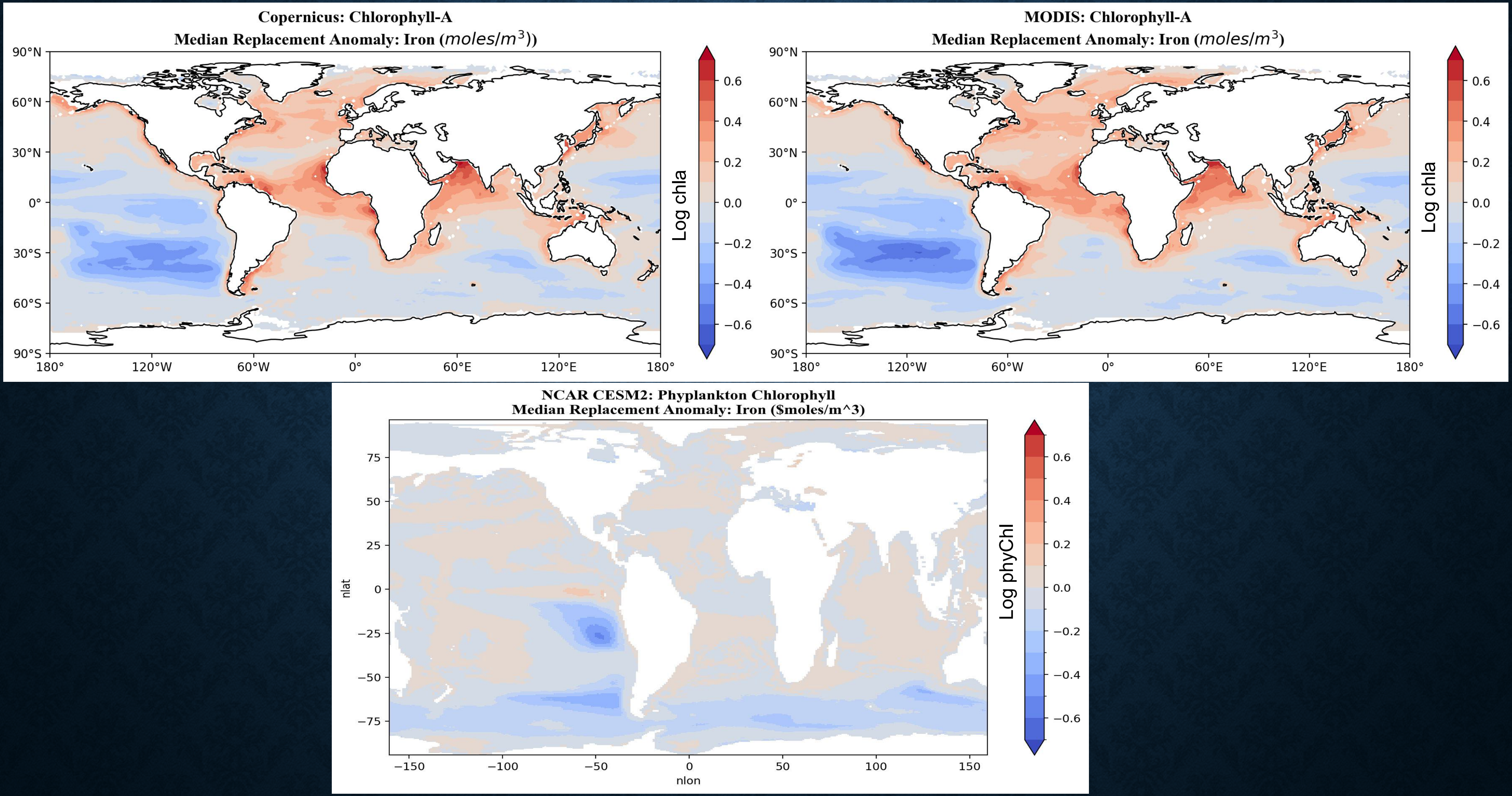


Prediction

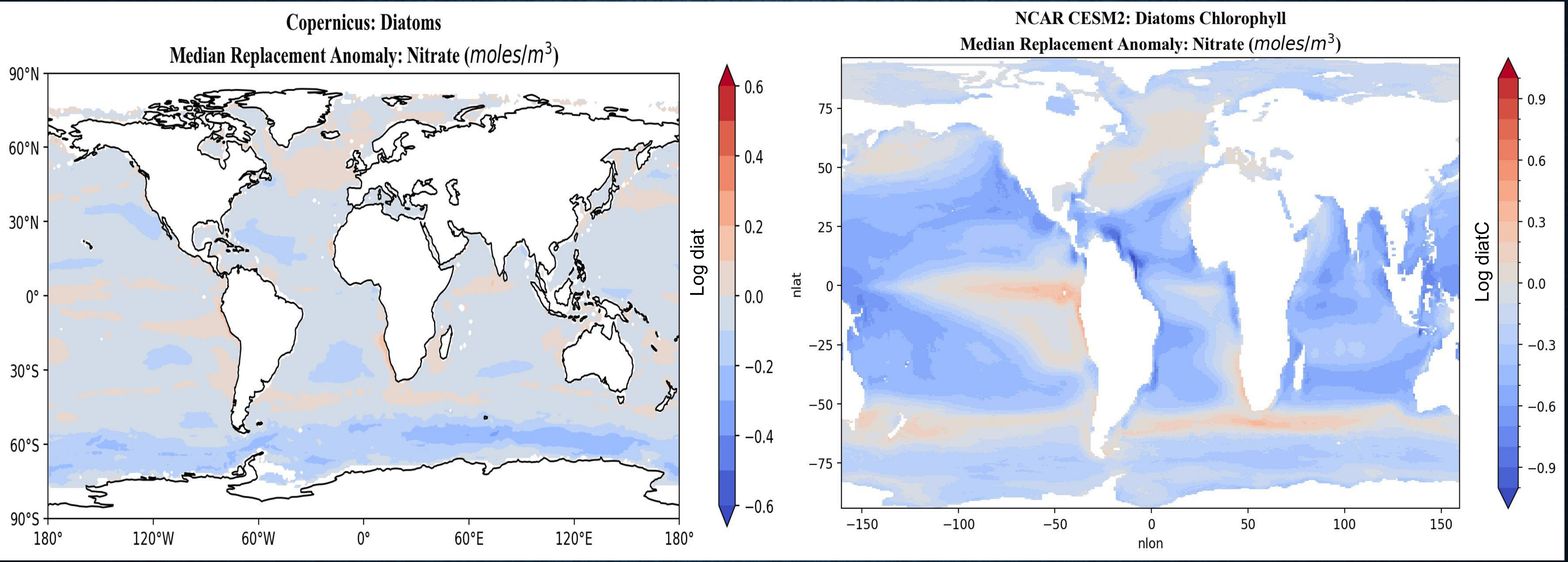
Difference with actual Prediction

Measure of Input Importance

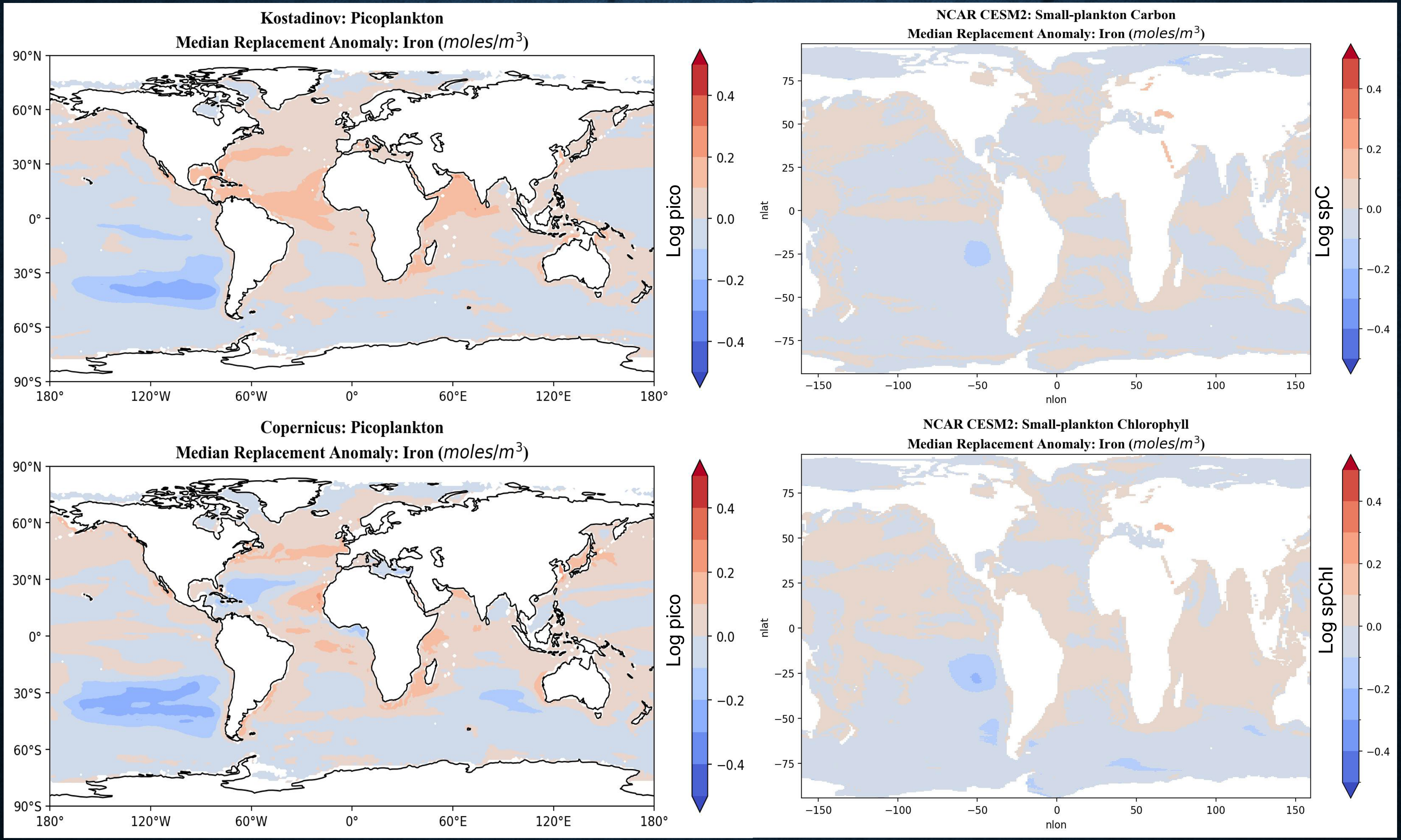
Median Replacement Anomaly



Median Replacement Anomaly



Median Replacement Anomaly



Analysis Results

- Random Forest effective in simulating biomass in observations and CESM2
- Generally shortwave radiation and SST found to be most important
- Iron important for total biomass, picoplankton in observations
- Model does not capture the effect of high iron in the Northern Hemisphere
- Nitrate important in model predictions for diatoms but not important for observations

Future Roadmap

- Compare with ESMs
- Use more XAI methods like SHAP and LIME
- Compare with PACE data
- Improve ESMs

Thank You

- Prof. Anand Gnanadesikan
- NOAA Grant
- Chris Holder