Here, Dutkiewicz and colleagues using a biological rich model embedded in a global circulation model to examine underlying controls on global pattern of plankton alphadiversity. Overall, it is a really nice study. It is well written, the results a clearly presented, the results are very interesting and the paper generally include a very thoughtful discussion. As such, I only have minor comments.

We thank the reviewer for these positive comments and appreciate the improvements that they make to the article. Below which we respond to the reviewer's comments (black text) in blue text.

I really appreciate that the authors are very explicit about these results being found in a 'model' world. This distinction is often blurred.

Thank you. It is good to maintain this distinction, but also to show how insight from the model can be applied to the real world.

Figure 1 is very convincing.

The study would benefit from a formal comparison between observations and model outputs. Right now, we are left with a visual test. Most global ocean model studies suffer from this issue but I just don't like statements like 'similar pattern' and such. These statements sometimes cover an awful match. I don't this is the case here but nevertheless. . .

We now include a new section in the supplemental with more formal evaluation (biases, spatial correlation and standard deviations) of the model against a variety of satellite and in situ observations (Revised version Supplemental Section S2 and new Supplemental Figs S3-S8). This includes both global level and also more explicit evaluation against the AMT data. However, we note that it is difficult to make a clean comparison between snapshots from the cruises and the model. We describe this more in the Revised Supplemental Section S2. We believe that the figures (e.g. Fig 1 and 5) in the main text are much clearer as they are (rather than the figures designed to show the bias explicitly, Supplemental Figs S6,S7), and plan to leave these as is in the main text. We do point the reader to this more formal evaluation in the revised version of the paper (Near lines 185, revise version lines 196):

"Model development was guided by evaluating against a range of in situ and satellite-derived observations (see Supplemental text S2 and Supplemental Figures S3-S8). We refer the reader to the fuller evaluation in the Supplemental, but provide a brief version here."

At the end of this response, please find the proposed new Supplemental section and figures.

There is obviously a lot to learn from using an R* type framework. However, the framework (in general and as applied here) ignores a key ecosystem feature, whereby organisms switch between different variants of the same resource (e.g., ammonium, nitrite, nitrate, urea, other DON, etc.) – each likely less palatable. This possibility for resource substitution changes the dynamics of diversity in relation to nutrient levels. For instance, it is likely much harder to have competitive exclusion and specialization in one resource might come at the expense of others. This does not invalidate the current study in anyway and it would be challenging to model all these additional tracers. However, I think it would useful to discuss this limitation – especially as it relates to the emergent diversity patterns.

This is an interesting comment. We believe that the R* framework could be altered to address this issue, and it would be interesting to see what that would suggest for diversity, especially if different species had more/less affinity for any of the variants of some resource. This is beyond the scope of this current study.

However, importantly, the numerical model does include ammonium, nitrite, nitrate, with phytoplankton preferentially consuming ammonium. The insight from the R* framework as provided here does still help us understand the results even from this more complex system (ie. the model). We make this clearer in the text (near line 274, revised version lines 294-297)

"We note that the model is significantly more complex than the simple theoretical framework, including multiple limiting nutrients, multiple variants of one of those resources (NH_4, NO_2 and NO_3) with differing affinities, additional loss terms (e.g. sinking) as well as more complicated grazing and foodweb (rather than food chain). However, this framework still helps us understand the patterns of size diversity in the model."

Do you have any issues with the smallest or largest size class? In other words, are there biological boundary problems due to less competition at the edges.

The model does not capture as many larger size classes as observed. This is likely because there are other traits (shape, chain formation, buoyancy control) that we do not include. This is noted in the text (lines 200-202, revised version 215):

"The model captures biomass in almost all size classes (Fig 6, Supplemental Fig S10a), though the largest size classes are likely underestimated. Traits not included in the model (e.g. buoyancy regulation, chain formation, symbiosis) are possibly more important for maintaining these large size classes."

But now also mention in Section 6 (Limitation of this Study), near Lines 407 (revised version lines 436)

"The model considers only three axes of phytoplankton traits. We anticipate that additional axes such as morphology (e.g. shape, spines), motility (e.g. flagella), chains, colony formation, nutrient storage abilities, and symbiosis will each have their own controlling mechanisms. Such traits might allow the model to capture more species, and particularly, more larger types."

L449: I think it is a mistake to think of latitude as an environmental factor. Also, I think it is unfair to characterize past studies as simple statistical correlations. When people are looking for relationships to latitude, they are not arguing that plankton respond to where they are located on a map. Rather latitude is a placeholder for a range of abiotic and biotic interactions. Thus, I think it is reasonable to look at relationship with latitude and I found this section a tad too negative about past efforts.

We agree that it is potentially confusing to call latitude an "environmental" factor and have altered the text in many locations to reflect this. And we agree that latitude is often used to represent many other factors. However, we argue that there are many different nutrient supply rates, and different levels of mixing along longitude at any given latitude. And as such we suggest latitude will not be able to fully explain diversity patterns. We have revised the text to elaborate on this point and to be a little less negative. We do not quote all revised text on these issues, but include an example here (line 448, revise version 490):

"Correlating with factors such as temperature, latitude is a logical first step for trying to understand observed patterns of diversity, as these are often the only additional data that is available from a field study, and for instance "latitude" could potentially stand in for a range of biotic and abiotic processes. Our study however suggests that to some degree these may not be able to help disentangle controllers of diversity."

NEW SUPPLEMENTAL MATERIAL:

S2. Model Evaluation

We evaluate the model against a range of in situ and satellite-derived observations (Main text Figs 1,5,7, and Supplemental Figs S3-S8). The model captures the patterns of low and high surface nutrients seen in the compilation of in situ observation from World Ocean Atlas (Garcia et al., 2014, Supplemental Fig S3). Nitrate is slightly too high in the Pacific gyres and too low along the equator. This reflects that iron limitation may be too strong in this region. But the correlation to observations is good (Supplemental Fig S5). Phosphate has similar, but accentuated, biases in the Pacific Equatorial region, and is also too high in the Southern Ocean. Phosphate is thus more evenly distributed than observed (Supplemental Fig S5). Likely the fixed stoichiometry of the model leads to phosphate concentrations not being sufficiently biologically modulated. Silicic acid also shows similar biases in the Equatorial Pacific and is too high in the Southern Ocean. This latter bias is likely a reflection of constant Si:C we impose. In the Southern Ocean, diatoms are more highly silicified (Tréguer et al 2017). This overestimation in the Southern Ocean leads to a higher spatial standard deviation relative to the observations (Supplemental Fig S5).

Chl-a compares well to satellite estimate (Supplemental Fig S4, S5). Note that the satellite estimates have large uncertainties (Moore et al, 2009 estimates more the 35% errors) and, moreover, the values shown for the satellite Chl-a estimates in Supplemental Fig S4 are not true annual means, but rather compilations of all available data, missing values when there are clouds or the light levels are too low (e.g. polar winters). The coarse resolution of the model does not capture important physical processes near coastlines, and lack of sedimentary and terrestrial supplies of nutrients and organic matter lead to Chl-a being too low in these regions. Chl-a is under-estimated by the model in the subtropical gyres, likely due to lack of mesoscale processes in the model that would supply additional nutrients in these regions (see e.g. Clayton et al 2017). The model Chl-a is higher than the satellite estimates in the high latitudes. Regional biases in the satellite algorithms are likely, particularly an issue in the Southern Ocean (e.g. Szeto et al., 2011, Johnson et al. 2013). The model though has a good correlation with the observations and captures the spatial variability well (Supplemental Fig S5).

We further compare the model to satellite-based estimates of Chl-a in different size classes (Main Text Fig 7, Supplemental Fig S4, S5), using the product from Ward et al (2005). Here we capture the ubiquitous pico-phytoplankton and the limitation of the larger size classes to the more productive regions. The model pico-phytoplankton size class Chl-a is potentially slightly too low and the nano size class too high. Though we note that if we set the pico/nano break at the model 5th size class (just under 3μm) instead at the 4th (2μm) size class, the relative values are much more in line with the satellite product. We suggest that the satellite product division might not be that exact. The micro-size class matches in location to the satellite product but is slightly too low as discussed above, but has the least impressive correlation to the observations (Supplemental Fig S5).

We also compare the model functional group distribution to the latest compilation of observations (Main Text Fig 7b, MAREDAT, Buitenhuis et al 2013, and references therein). The observations are sparse and here we average all observations regardless of season in 5 degree bins (Main Text Fig 7b). With such spatially and temporally sparse observations, we do not believe it makes sense to calculate biases or correlations between the model and observations, and we rely on visual evaluation. Though the observations are sparse, we do capture the ubiquitous nature of the pico-phytoplankton, the limited domain of the diazotrophs (including observed lack of diazotrophs in the South Pacific gyre), the pattern of enhance diatom biomass in high latitude, and low in subtropical gyres. We over-estimate the coccolithophore biomass relative to MAREDAT in many regions, but note that the conversion from cells to biomass in that compilation was estimated to have uncertainties as much as several 100% (O'Brien et al., 2013). The MAREDAT compilation did not include a category for dinoflagellates.

We further evaluate the model against the in situ observations as captured during the Atlantic Meridional Transects (AMT) 1,2,3, and 4 (Main Text Fig 1, 5, Supplemental Figs S6,S7,S8). AMT2 and 4 occurred during April and May of consecutive years, while 1 and 3 took place during September and October. Here we compare the range of values found in the two cruises in each time period to the range of values in the model during the two-month period (Supplemental Figs S6,S7). Similar to the global evaluation above, we find that silicic acid is too high in the Southern Ocean (Supplemental Fig S6) and that Chl-a is underestimated in the subtropical gyres. We note that the model Chl-a compares better to the Southern Ocean in situ observations than they do to the satellite estimates. Though the correlation is reasonable, the spatial variability is too low (Supplemental Fig S8a,b). The phytoplankton functional groups compare less well to observations than the nutrients and Chl-a, but are still plausible. Coccolithophore biomass however drops too low in the Southern Ocean, likely due to the model smallest diatom being parameterized as too competitively advantaged. However, pleasingly, the relative abundances of the three groups (diatoms, coccolithophores and dinoflagellates) are captured: Diatom biomass is much lower in the subtropical gyres than the other two functional groups, and higher in the Southern Ocean and coccolithophores and dinoflagellates as having much more even distributions.

As a final model evaluation, we compare the model estimates of richness against those found along the AMT (Main Text Fig 1, Supplemental Fig S7, S8b,c). As expected, given the only 350 species parameterized in the model, the model has lower diversity than seen in the AMT. But, the model does captures the low and high patterns of total richness along the AMT (Supplemental Fig S7a,d), though underestimates the diversity in the subtropical gyres. In these regions it is likely that traits axes (e.g. symbiosis, colony formation etc) not captured in the model provide additional means for phytoplankton to co-exist. The richness within different functional groups is also captured, though much better for diatoms than the other two groups (Supplemental Figs 7b,e, Supplemental S8c,d). Excitingly the model also captures the differences in the diversity within functional groups and in size classes. Diatoms have much larger diversity in the Southern Ocean than the other functional groups, while coccolithophores and mixotrophic dinoflagellates diversity is much more uniform across the transect. AMT richness was also calculated for different size classes. The model does well in capturing these divisions as well (Supplemental Fig S7c,f, S8c,d). The model captures the much higher diversity within the smallest size category (2-10μm) and the lower and much more regionally varying diversity in the larger size category, including the lack of diversity in the largest size class (>20μm) in the subtropical gyres.

NEW SUPPLEMENTAL REFERENCES

Clayton, S. Dutkiewicz, O. Jahn, C.N. Hill, P. Heimbach, and M.J. Follows. Biogeochemical versus ecological consequences of modeled ocean physics. *Biogeosciences*, **14**, 2877-2889 (2017).

Garcia, H. E., R. A. Locarnini, T. P. Boyer, J. I. Antonov, O.K. Baranova, M.M. Zweng, J.R. Reagan, D.R. Johnson, 2014. *World Ocean Atlas 2013, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate, silicate).* S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 76, 25 pp.

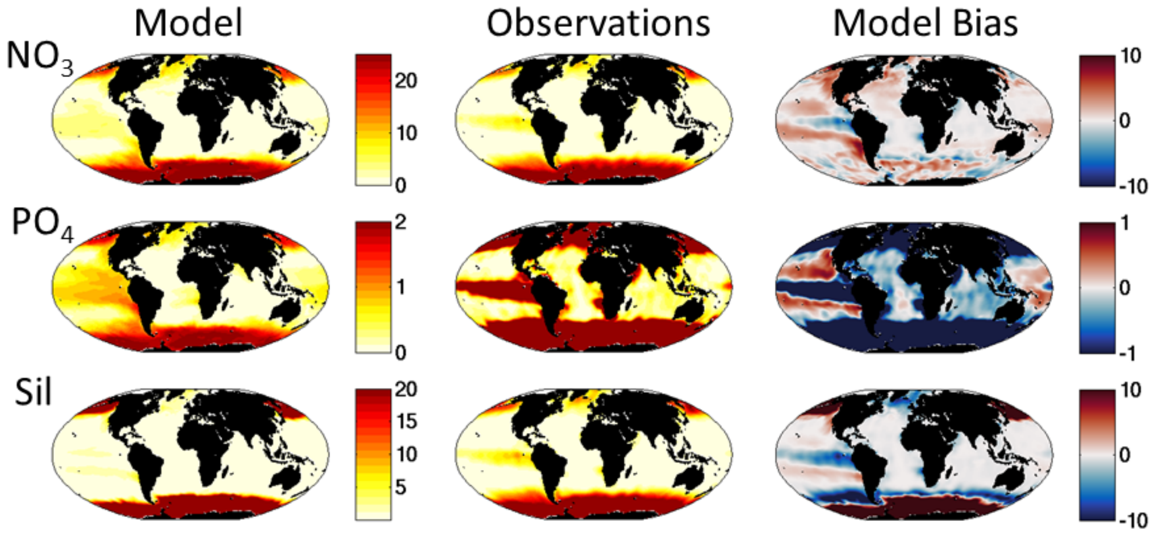
Johnson, R., Strutton, P.G., Wright, S.W., McMinn, A., and Meiners, K.M.: Three improved Satellite Chlorophyll algorithms for the Southern Ocean, J. Geophys. Res. Oceans, 118, doi:10.1002/jgrc.20270, 2013

Moore, T.S., Campell, J.W., and Dowel, M.D.: A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sensing of Environment, 113, 2424–2430, 2009.

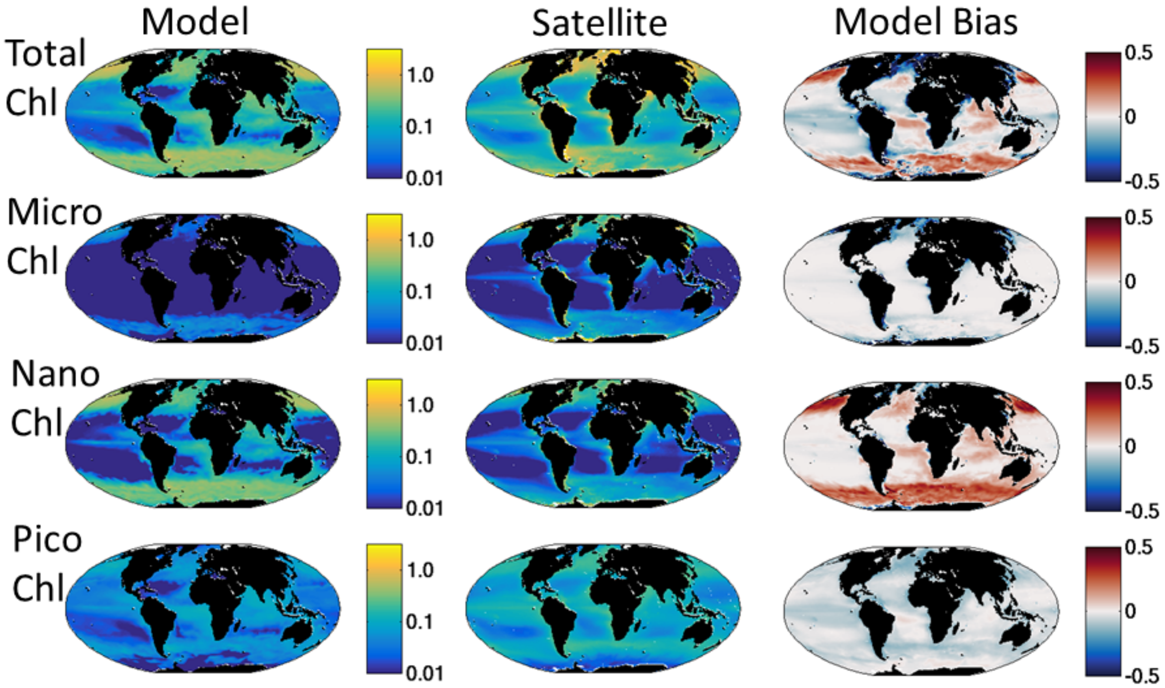
Szeto, M., Werdell, P.J., Moore, T.S., and Campbell, J.W. :Are the world's oceans optically different?, J. Geophys. Res., 116, C00H04, doi:10.1029/2011JC007230, 2011.

Tréguer, P., Bowler, C., B. Moriceau, S. Dutkiewicz, M. Gehlen, K. Leblanc, O. Aumont, L. Bittner, R. Dugdale, Z. Finkel, D. Iudicone, O. Jahn, L. Guidi, M. Lasbleiz, M. Levy, and P. Pondaven. Influence of diatoms on the ocean biological pump. *Nature Geoscience*, doi:10.1038/s41561-017-0028-x (2017).

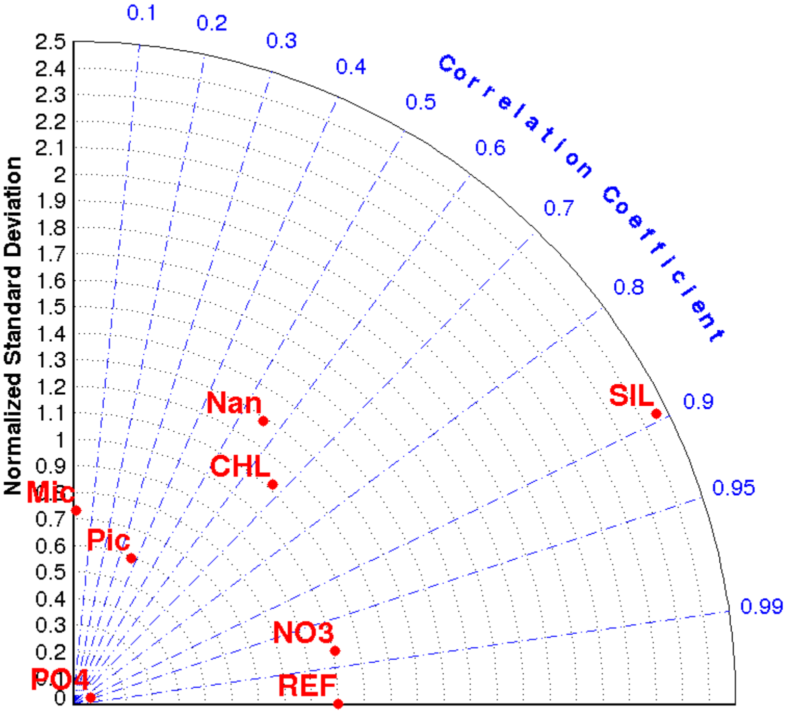
NEW SUPPLEMENTAL FIGURES



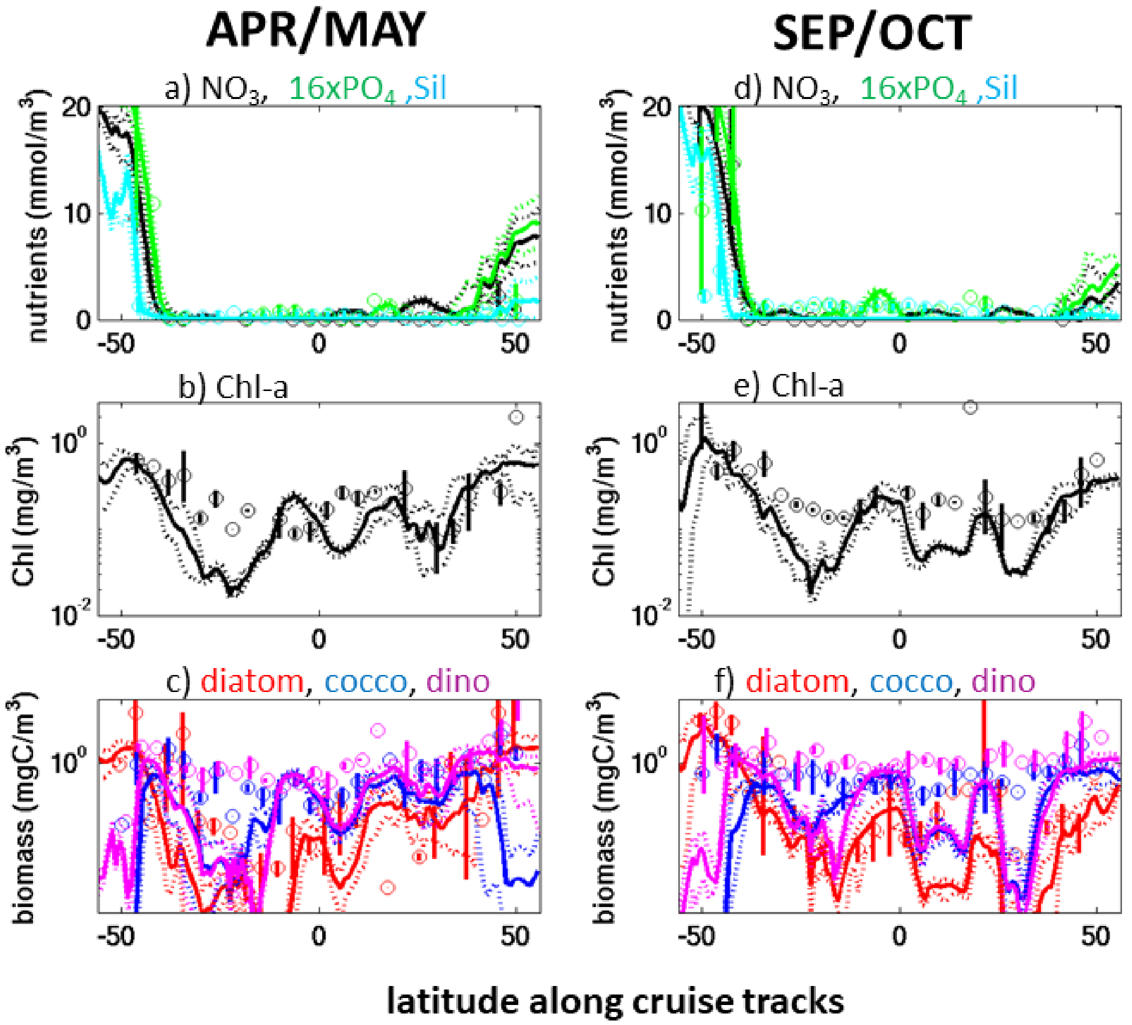
Supplemental Figure S3: Annual Mean Surface (0-10m) Nutrients. (Top row) Nitrate (mmolN/m^3); (Middle row) Phosphate (mmolP/m^3); (Bottom row) Silicic acid (mmolSi/m^3). (Left column) Model, 5th year annual mean; (Middle column) Observations , annual climatology, from World Ocean Atlas (Garcia et al 2013); (Left Column) Model bias determined as model minus observation.



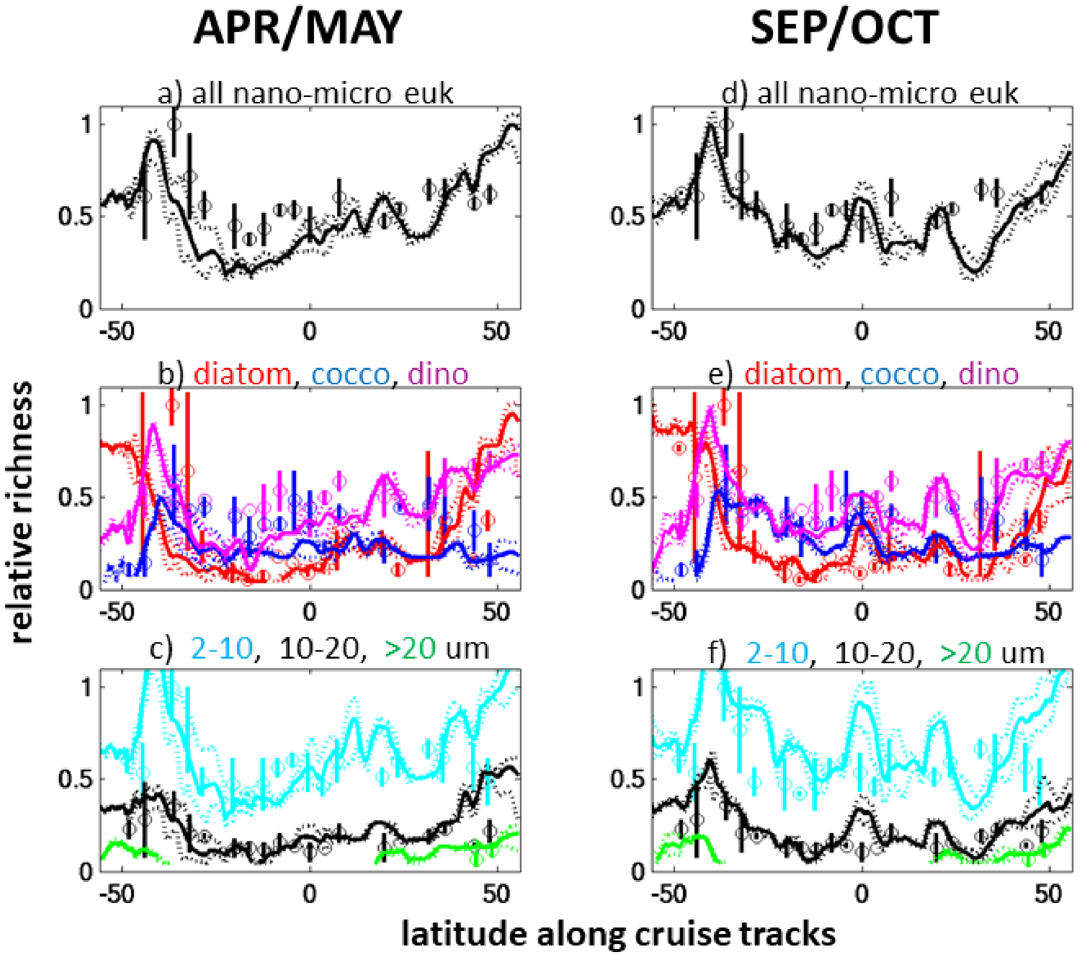
Supplemental Figure S4: Annual Chl (mgChl/m^3). (Top row) total Chl-a; (Second row) Chl in micro (>20µm) size class; (Third row) Chl in nano (2-20µm) size class; (Bottom row) Chl in pico (<2µm) size class. (Left column) Model, 5th year; (Middle Column) Satellite Observations, top from NASA MODIS; other three panels are the satellite based estimates from Ward (2005); (Right Column) Model bias determined as model minus observations. The middle column shows annual "climatology" of all available satellite measurements, with missing observations in the polar winters; while model results are annual mean (0-10m).



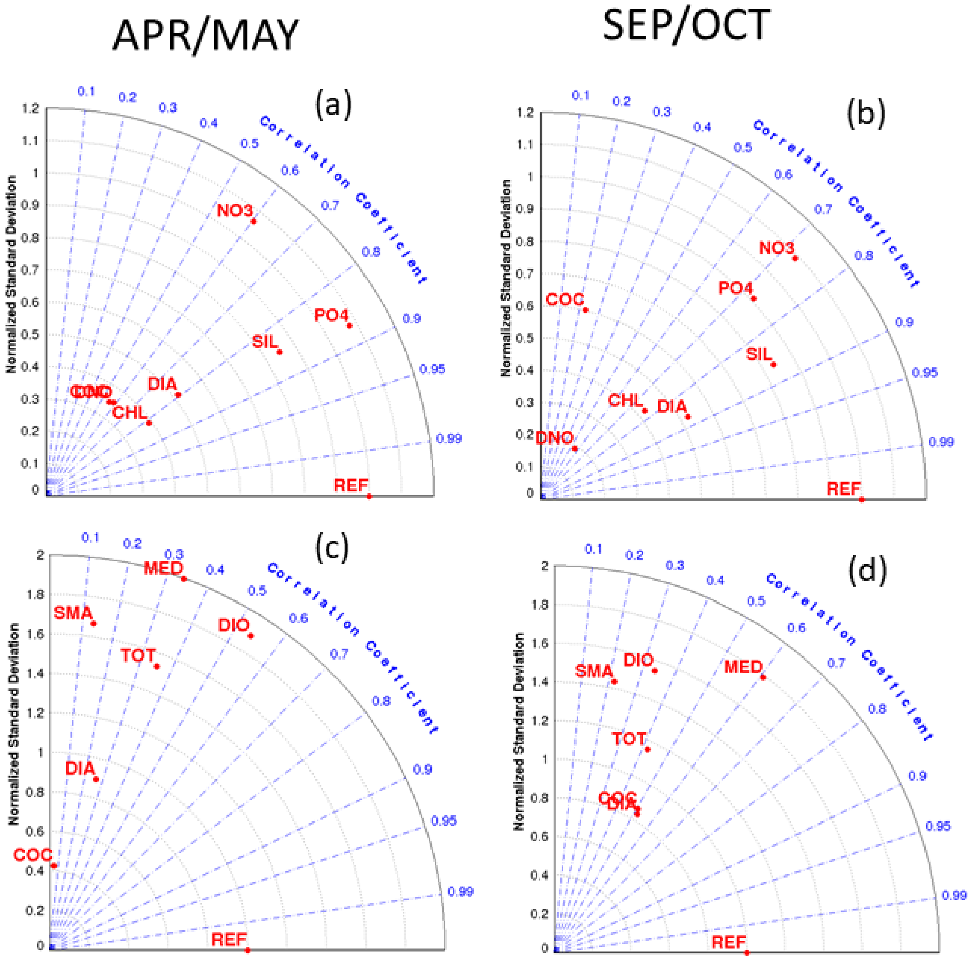
Supplemental Figure S5: Taylor Diagram of Global Annual Surface Fields. This polar coordinate plot shows correlation (angular position) and the normalized (by observed spatial STD) spatial standard deviation (radial position) between model and observation for the fields shown in Supplemental Figures S3 and S4. Statistics are performed on log-normalized fields. REF indicates a perfect match between model and observations. NO3, PO4, SIL refer to nitrate, phosphate and silicic acid respectively; Observations are from World Ocean Atlas (Garcia et al 2014). CHL refers to total Chl-a; Observations are satellite estimates from NASA MODIS. Mic, Nan, Pic refer to Chl-a in the micro (>20µm), nano (2-20µm), pico (<2µm) respectively; Observations are the satellite-based estimates in each size class from Ward et al (2005).



Supplemental Figure S6: Atlantic Meridional Transect Model and In situ Observations. Left Column is for April/May (AMT2,4) results, Right Column for September/October (AMT1,3). Circles indicates average of the two AMT cruises in 4° latitude bins in each time period, and the vertical line across each circle shows the range of the observations. Solid lines indicate the model two-month mean and dashed lines indicate the model minimum and maximum from that two-month period. (a), (b) surface nutrients (black=nitrate, mmolN/m^3; green=phosphate, 16xmmolP/m^3; light blue=silicic acid, mmolSi/m^3); (c), (d) surface Chl-a (mg Chl/m^3); (e), (f) surface phytoplankton biomass (mg C/m^3); red=diatoms; blue=coccolithophores; purple=dinoflagellates).



Supplemental Figure S7: Atlantic Meridional Transect Model and In situ Observations of richness. Left Column is for April/May (AMT2,4) results, Right Column for September/October (AMT1,3). Circle indicates average of the two AMT cruises in each time period in 4° latitude bins, and the vertical line across each circle shows the range of the observations. Solid lines indicate the model two-month mean and dashed lines indicate the model minimum and maximum from that two-month period. Normalized richness of (a),(d) all diatoms, coccolithophores and dinoflagellates together; (b),(e) each functional groups separately (red: diatoms, dark blue: coccolithophores, purple: dinoflagellates); (c),(f) 3 size classes (light blue: 2-10µm, black: 10-20µm, green: >20µm). Model pico-phytoplankton and diazotrophs are not included in the model analysis as they were not analyzed in the observations.



Supplemental Figure S8: Taylor Diagram of Atlantic Meridional Fields. This polar coordinate plot shows correlation (angular position) and the normalized (by observed spatial STD) spatial standard deviation (radial position) between model and observation for the fields shown in Supplemental Figures S6 and S7. Left Column is for April/May (AMT2,4) results, Right Column for September/October (AMT1,3). We compare the in situ two-cruise mean (circles in Fig S6 and S7) against the model two-month average (solid lines) averaged onto the same 4° latitude bins. REF indicates a perfect match between model and observations. (a),(b) NO3, PO4, SIL refer to nitrate, phosphate and silicic acid respectively. CHL refers to Chl-a. DIA, COC, DIO refer to diatom, coccolithophore and dinoflagellate biomass respectively. Statistics are performed on log-normalized fields for the Chl-a and biomass fields. (c),(d) normalized richness where TOT refers to the total richness DIA, COC, DINO refers to the richness in diatoms, coccolithophores, and dinoflagellates respectively, and SMA, MED, LAR to the 3 size classes (2-10µm, 10-20µm, >20µm) respectively.