

Biogeosciences manuscript bg-2019-506

"Potential predictability of marine ecosystem drivers" by T. L. Frölicher, et al.
March 20, 2020.

We thank the editor and both reviewers for assessing our manuscript. The comments have helped to further improve our manuscript. The review comments are given in black and our reply in blue.

Detailed Response to Editor's comments

Dear authors, My compliments for your nice work and revision.

We thank the editor for the positive decision and for efficiently handling the paper.

In the revision, I appreciated your mapping of the limiting nutrients in figure 13. However, I was quite surprised by these distributions. For example, Moore et al. (2013), suggested that nitrogen is limiting primary production at the ocean low latitudes. This is quite in contrast with your findings, which indicate iron and phosphorus as limiting (small) phytoplankton production. Can you please comment this result?

We thank the reviewer for raising this interesting question. We think that GFDL ESM2M is difficult to compare directly with the Moore et al. study for two reasons: 1) ESM2M is a coupled Earth system model with an open nitrogen cycle with N₂ fixation, atmospheric deposition, and river runoff as sources, and water column and sediment denitrification as sinks that are free to be out of balance, whereas the Moore study uses an observation-driven parameterization of phytoplankton growth rather than an Earth system model. 2) In contrast with the Moore et al. study, which assesses iron limitation through nutrient uptake alone, the role of iron limitation in ESM2M operates through the ability to synthesize chlorophyll. As such, iron limitation can become severe, but as long as some chlorophyll can be produced, suboptimal productivity can proceed driving down the iron limitation term. We have clarified in the manuscript: "GFDL ESM2M's overall strong iron limitation is in contrast to Moore et al. (2013), who suggests that nitrogen is the limiting nutrient in the subtropical gyres. GFDL ESM2M is a fully coupled Earth system model and assess iron limitation through the ability to synthesize chlorophyll. In contrast, Moore et al., (2013) uses observation-driven parametrizations of phytoplankton growth and assess iron limitation through nutrient uptake alone."

Another minor suggestions: please refer to figure 1 when defining potential predictability in section 2.3.1 This would help the reader.

We have added a reference to figure 1 in section 2.3.1. Many thanks for this pointer.

Detailed Response to Reviewer's comments: Referee #1

This paper evaluates the potential predictability of marine ecosystem drivers (T, pH, O₂, and NPP) and discusses which physical and biological processes fundamentally determines the upper limit of the predictability at global and regional scales using a series of perfect model simulation in an earth system model (GFDL ESM2M). Whereas previous studies focus on a physical process that describes the potential predictability of marine biogeochemical variables, this study shows that the biogeochemical interactions, as well as physical processes, are key drivers to contribute to the upper limit of the predictability, which leads to the understanding of differences in the potential predictability between temperature (physical) and biogeochemical variations. The global and regional characteristics of the biogeochemical predictability that this paper shows are important information to produce skillful multi-year predictions of marine ecosystems. I recommend a minor revision according to the comments given below.

We thank the reviewer for this positive and encouraging review.

(1) To what extent does AMOC variability affect the potential predictability of marine ecosystem drivers in the North Atlantic? The decadal potential predictability of marine ecosystem drivers is prominent in the North Atlantic, which would be strongly related to the variability in AMOC. To understand the mean states and variations in AMOC in perfect and ensemble simulations in GFDL ESM2M, I request to show the time series of AMOC variability, like Figure 1a.

We have included a new Figure C1 (see below) that shows the simulated AMOC maximum from the 300-yr long preindustrial control simulation, as well as the prognostic potential predictability of the AMOC maximum. The text in section 3.3.1 has been modified to say: "In GFDL's ESM2M, the AMOC experiences strong low-frequency variability, consistent with Msadek et al. (2010), and its predictability time horizon is about 9 yr (Figure C1)."

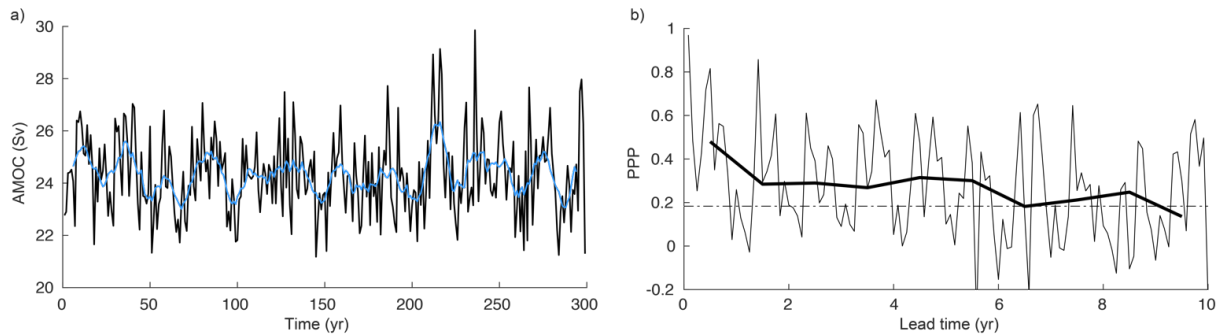


Figure C1: (a) Simulated annual mean AMOC maximum of the 300-yr long preindustrial control simulation. The blue line indicates the 10-yr running mean. (b) Monthly mean (thin line) and annual mean (thick line) prognostic potential predictability for the AMOC maximum. The horizontal black dashed line represents the predictability threshold.

Importantly, we note that the predictability of some marine ecosystem drivers is dependent on a complex interplay between physical and biogeochemical processes. NPP, as an illustrative example, has a relatively short predictability time horizon in the North Atlantic, with this behavior not directly prescribed by AMOC variations.

(2) What limiting nutrient contributes to higher potential predictability in NPP? In this paper, the regions of higher potential predictability in NPP (Fig. 3d) correspond to those of higher predictability in nutrient limitation, which suggests that higher potential predictability in NPP is fundamentally constrained by the availability of limiting nutrients. Since the model uses the formulations of limitation by multiple nutrients, I wonder what nutrient is key to contribute to higher potential predictability in NPP at the regional scale. The figure and description of a spatial pattern of limiting nutrients are helpful to understand the characteristics of long-term biological variability.

Many thanks for this suggestion. We have included in Figure 13f (former Figure 13g) with contours the limiting nutrients (nitrogen, phosphate or iron) for the growth rate of small phytoplankton production at the surface. We added following sentence to section 3.3.4: “In GFDL’s ESM2M, the subtropical gyres are mainly iron limited (hatching in Figure 13f) and therefore iron fundamentally constrains the predictability of the growth rate of small phytoplankton there. Exceptions are the boundary region between the subtropical and subpolar gyre in the North Pacific (nitrate limited) as well as the tropical Atlantic (phosphate and nitrate) and the northern Indian Ocean (phosphate).”

However, we would like to stress here that the relatively long predictability horizon for NPP in the subtropical gyres is also strongly influenced by the relatively long predictability horizon of the small phytoplankton stock, as can be seen in Fig. 12e.

Other comments:

L137. Remove “The prognostic “
Done.

L159. reflect, not reflects
Done.

L. 281. largely, instead of large

We have reformulated the sentence to: “Therefore, the influence of high frequency atmospheric variability is large, which leads to diminished potential predictability around Antarctica.”

L 363. Why do you refer to Fig.10?

Typo. Many thanks for spotting this. We deleted the reference to Fig. 10, but included a reference to Fig. 3 in the previous sentence.

L. 527. Figure B1 should be Figure A1 on page 37.

We changed the labelling of Figure A1 to Figure B1, as the figure is described in Appendix B.

Page 800. The unit of the iron half-saturation coefficient would be wrong.
Changed and acknowledged.

Figure 3d, Contour lines, and contour number information are too dark to be identified.
We changed the color to red and increased the thickness of the contours.

Detailed Response to Reviewer's comments: Referee #2

This manuscript is a well written, interesting analysis of limits to prediction of biogeochemical quantities in a global Earth system model. I particularly liked the oceanographic explanations for the outcomes of the numerical experiments, and deconvolution of drivers. I am not an expert in global model predictability, but I suspect it is important for the field to undertake a number of these experiments on different models (the authors say this themselves). Even if similar studies exist, or follow this one, this study will remain important. Therefore, I recommend publication with the following, relatively easily-addressed points considered.

We thank the reviewer for this positive and encouraging review.

Major comments:

1. As someone not familiar with this form of model sensitivity analysis, I found some jargon that could have been avoided, or better explained. The term 'perfect modelling framework' was introduced as though the reader should know what it means. More simply the study is a test of the sensitivity of biogeochemical quantities to temperature initial conditions.

We clarified in the method section the term 'perfect modelling framework': "By perturbing the initial conditions of the GFDL ESM2M and quantifying the spread of initially nearby model trajectories, the limit of initial condition predictability was assessed. The underlying assumption is that we have a perfect model (e.g. the model accurately represents all physical and biogeochemical processes relevant to assess marine ecosystem drivers at adequate temporal and spatial resolution), near perfect initial conditions and that we exclude a role for external forcing in determining or limiting predictability."

2. By using temperature the emphasis is on limits to prediction of physical driving of biogeochemical quantities. Other initial conditions could have equally been perturbed, such as salinity, nutrients etc. Although this wasn't tested explicitly, we found with the same coupled Earth system model (GFDL ESM2M) that the choice of prognostic (interactive) versus climatological chlorophyll with identical initial conditions leads to at least qualitatively similar divergence of the model over only a few model time steps. We found that weather patterns are distinct after less than one month under such perturbations. From this we infer that any field that interacts with the dynamical state of the model can serve as an analogous perturbation to those considered here, but this is left as a topic for future testing and quantification. We also note that we used six different starting points from the preindustrial control simulation to start the 40-member ensemble simulation. Therefore the six 40-member ensemble simulation have different physical and biogeochemical states.

And of course there are many other factors limiting predictability, such as the model parameterisation. This isolation of one source of limits to prediction is appropriate but should be made clearer in the introduction, and then discussed, in the light of the results, more thoroughly in the Discussion. For example, changing BGC models, or even the remineralisation rate of organic matter, would change the time scale of AOU.

The question of how specific parameterizations within a given model impact predictability is clearly an important question, and to our knowledge this remains unexplored for marine biogeochemistry and ecosystem drivers. This has in fact been explored for the case of weather prediction (Palmer and Williams, 2008), where the inclusion of stochastic parameterizations has been argued to increase predictability. Such parameterizations are not available in the model used here, and from our understanding existing work on this topic has not considered predictions on the timescales considered here. Returning more directly to the topics raised by the reviewer, testing of the sensitivity of the biogeochemistry model will be facilitated in the future for the case of the MOM6 (isopycnal model) that will be used by both GFDL and CESM, so that one can "switch" between GFDL's newer COBALT model and CESM's Marbl/BEC model. Likewise one could explore the sensitivity to the mixing scheme (GFDL's ePBL versus CESM's KPP scheme) through its effect on the dynamical behavior of the model modulating biogeochemistry. So we anticipate that there should be a valuable opportunity to explore these questions moving forward. We have modified the caveat section in the discussion section to address this comment. See also our reply to comment 6 below.

3. [Most important point that needs addressing]. The terms “lead time” and “predictability time horizon” are used interchangeably in the last paragraph of p8, which demonstrates an inconsistency. The term lead time makes sense to me in Fig 2, 4 and 5. It is the time axis, starting at the perturbed time, along which the variability of the ensemble and controls are measured. But figures 3, 6, 7 etc. the surface plotted is labelled “lead time” when it should be “predictability time horizon”. Predictability time horizon loosely describes as $PPP < 0.183$ (also, is prognostic potential predictability the same thing as predictability time horizon)? But this is problematic since PPP varies with time. Should it be $\min t$ for which $PPP(t) < 0.183$? Check every use of lead time, PPP and predictability time horizon and make sure it is consistent in the manuscript. Predictability is also loosely defined, and would in many cases be best replaced with “predictability time horizon”.

We agree with the reviewer that the terms ‘lead time’, ‘predictability time horizon’ and ‘PPP’ are sometimes used interchangeably in the manuscript. We have now carefully streamlined the usage of these terms throughout the manuscript. As a result, we often replaced ‘predictability’ with ‘predictability time horizon’. In addition, we changed ‘lead time’ to ‘predictability time horizon’ in Figure 3 and Figures 6 to 13.

As the reviewer noted, the prognostic potential predictability is not equal the predictability time horizon. The prognostic potential predictability metric (i.e. PPP) varies with lead time. However, as defined in section 2.3.1, the predictability time horizon is defined as the lead time at which PPP falls below 0.183.

4. Fig. 1b. This figure would be more instructive if it was used as an example of the calculation of the PPP. If the control variance could be plotted (maybe a grey between +/- sigma) and then PPP, and the point at which PPP drops below 0.183. This would set up the rest of the manuscript better.

Many thanks for this suggestion. We have added the standard deviation (not the variance as it becomes too small to be shown) of the control simulation as horizontal dashed lines and indicate with a vertical line the predictability time horizon (i.e. when PPP falls below the predictability threshold).

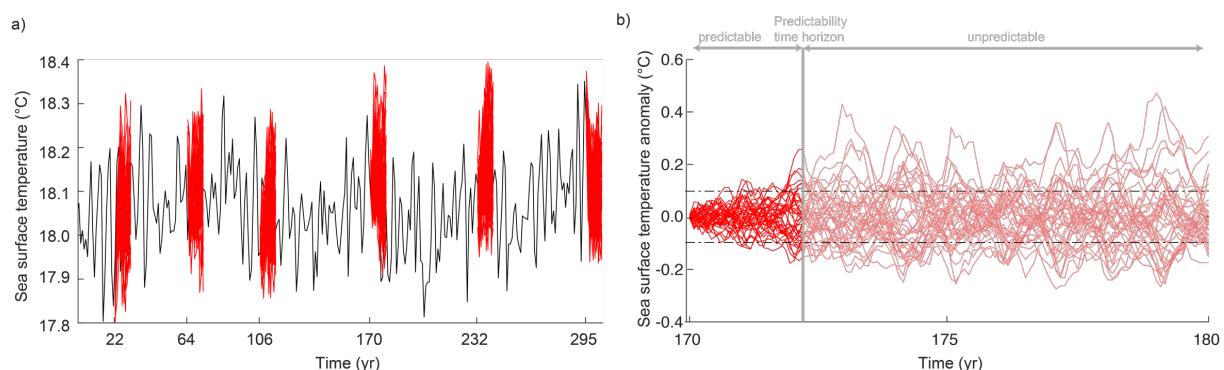


Figure 1: Illustration of the model setup and the calculation of the predictability time horizon. (a) Simulated global mean SST of the 300-yr reference control simulation (black line) and of the six 10-yr long 40 ensemble simulations (red lines). (b) Global mean SST anomaly (i.e., deviation from the control simulation) for the ensemble simulation starting in year 170. Thick red line indicates the period over which SST is predictable (i.e. $PPP \geq 0.183$), and thin red lines indicate period over which SST is unpredictable (i.e. $PPP < 0.183$). The dashed horizontal lines indicate one standard deviation of the control simulation and the vertical line indicates the predictability time horizon.

Also, in the caption, why do you call it the “first ensemble simulation”.

This was the first set of 40 ensemble members we have performed with ESM2M, but this is not important for the reader of this manuscript. We therefore deleted the word ‘first’.

5. Paragraph 396 – 405 needs re-writing for clarity. I think it is trying to say that you can have different time scales for predictability for perturbations in forcing (such as anthropogenic CO₂) to perturbations in initial conditions (as studied here).

We compare in this paragraph the Time of Emergence (ToE) of the marine ecosystem drivers with the predictability of these drivers. Earlier papers have found that the ToE strongly differs among the drivers. This is in stark contrast to the predictability horizon discussed in this paper, which is almost identical for all four ecosystem drivers at the global scale. We hope that the broader suite of revisions we have provided to the manuscript will clarify this message and the context in which it occurs.

6. The discussion has too much focus on obvious limitations (such as ensemble size, years started etc.) and less on more subtle limitations like time-scale of coefficients in the BGC model. The second are particularly worth of discussion here because the effort at deconvolution of the processes allows for an insightful discussion of these.

TOPAZv2 represents a hypothetically "optimal" phytoplankton physiology, i.e. it assumes that the fastest growing phytoplankton group always wins in all environments via the upper limit in growth rates as per Bissinger et al., (2008) and Eppley (1972). Similarly, because TOPAZv2 represents a steady-state ecosystem, there are no time lags between primary production and the grazing response. In the subsurface, the remineralization of particles is set to reproduce the vertical scale of the nutricline on the timescale of sinking particles, and the sinking particle velocity is fast, i.e. 100 m/day. All three factors would tend to decrease the memory associated with the real world surface ecosystem and minimize predictability. We expect coefficients relating to the longer time scales to include the dissolved organic matter remineralization and the depth scale of sinking particle remineralization, both are fairly well constrained from observations at the global scale, but we expect considerable regional and temporal variability in these. As with our comments above, quantifying these effects is outside of the scope of this study, as testing this would require an independent and more expansive site of simulations that are beyond the physical resources available for this study. We have modified the caveat section in the discussion section to address this comment.

Minor comments.

1. Whenever referring to time values, try to keep the adjectives to ones with a sense of time such as low -> short (L254), elevated -> lengthened (L195), high -> long (L391). This aids readability. There are many examples of this.

We thank the reviewer for this and changed the text accordingly.

2. The sentence L91 starting "The six" should come before the "Note" for better readability.
Changed and acknowledged.

3. L92 replace "are" with "were"?
Done.

4. Line 100. For those interpreting the equation, maybe a sentence after it "Thus the range of perturbations is evenly spread from -0.002 to 0.002 C with the control in the centre."
Many thanks for this suggestion, which we included.

5. L110 replace "underrepresented" with "underestimated"
Done.

6. Description of Eq.2 (L119-120) doesn't mention the six ensembles.

The equation is further explained on lines 125-131, where it is stated that N is the total number of different ensemble simulations (N=6) and M the number of ensemble members (M=40)

I didn't fully understand the rationale for 6 ensembles of size 40. Why not 240 members starting at all different times?

It is a common procedure for perfect modelling frameworks that different multi-member ensemble simulations started at different points in time of the control simulations are used to assess predictability. Using different points in the control simulation randomizes the initial conditions (e.g. yielding different ENSO phase states), with this intended to average across biases that may result from predictability being different across different phase of climate modes. We clarified this in section 2.2.

Also is sigma of the control the same for all ensembles?

Yes, the sigma of the control is the same for all ensembles.

Just a little bit more help here to those unfamiliar with the approach.

We hope that our answers above helped to clarify things.

7. L150 pH is approximately $-\log_{10}([H^+])$. I know you didn't mean to define it here, but the use of (or X) sort of implies it.

We deleted the bracket to avoid confusion.

8. L184 meaning of “PPP with lead time” not clear.

We simplified it to “PPP over time”.

9. L206 replace “across” with “for each of the”

We changed the sentence to “.. differences between each of the four ..”

10. L234. How can a coupling enhance predictability? Sentence needs to be more carefully constructed.

We changed to sentence to: “However, biogeochemical processes lead to enhanced predictability below 500 m for O₂ and pH.”

11. L387 “predictability of each variable”.

Changed and acknowledged.

References

Bissinger, J. E., Montagnes, D. J. S., harples, J. and Atkinson, D.: Predicting marine phytoplankton maximum growth rates from temperature: Improving on the Eppley curve using quantile regression, *Limnol. Oceanogr.*, 53(2), 487–493, doi:10.4319/lo.2008.53.2.0487, 2008.

Eppley, R.: Temperature and phytoplankton growth in the sea, *Fish. Bull.*, 1972.

Palmer, T. N. and Williams, P. D.: Introduction. Stochastic physics and climate modelling, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, doi:10.1098/rsta.2008.0059, 2008.