We thank both Reviewers for their time in revisiting this manuscript. Please find our point-by-point response (normal font) to all comments (*italics*) below.

Reviewer #1: John Dunne

General comments:

The manuscript "Trivial improvements of predictive skill due to direct reconstruction of global carbon cycle" by Spring et al. describe results from reconstructed simulations nudging a model simulation to itself for atmospheric physics, ocean physics, and carbon cycle as well as perfect predictability experiments with the original model and these reconstructions. I like this paper very much as a detailed investigation into the limits of injesting "data" into a model.

Thank you for your supportive assessment of our work.

The text is extremely dense, with many concepts and model limitations all being discussed at once with a focus on biases without any description of the mean state, which made me need to read over sections over and over before I was able to put the pieces together. For example, the degredation of ITCZ and Southern Ocean winds is only clear after one puts together a mental map of the base state. It might help to have the base simulation of each parameter in Figure 1 as a new Figure 1 or provided as supplementary material.

We will provide a climatology figure in the supplementary to enhance readability of the manuscript.

The use of language was combersome, however, with such vague words as "indirect", "direct" and "reconstruction" are used when descriptive terms like "physically nudged" "physically nudged atmosphere" and "phsyically and biogeochemically nudged" would have worked. I am guessing that there is a literature precedent for this redirection of terms, but it made the early parts of the manuscript difficult to maintain in scope.

We searched the literature and are not aware of a precedent, which we would happily comply with. Added Fig. 1 & 2 for further clarification. We decided against biogeochemically nudged, because we only nudge DIC and alkalinity, but no further biogeochemical variables. We plan to add the following figures for better understanding of the simulations to the supplementary.

The discussion of Figure 2 is incomplete and extends out through Figure 5.

We agree with the Reviewer's criticism of the current manuscript structure. In the manuscript we deal with multiple variables from the ocean, land, and atmospheric model components, analyze different spatial domains (gridded or global), and apply multiple metrics (bias, correlation, accuracy). An option would be to split the figure content into one figure per variable with all the different metrics to compare, which we find somewhat unpractical. Alternatively, what we choose, is to split the figure content into one figure per metric/method, where all variables are comparable in one figure.

In either structure, it is challenging to avoid the jumps in the discussion between the figures, because data presented in the figures is discussed and analyzed under various angles allowing to formulate our main conclusions. We

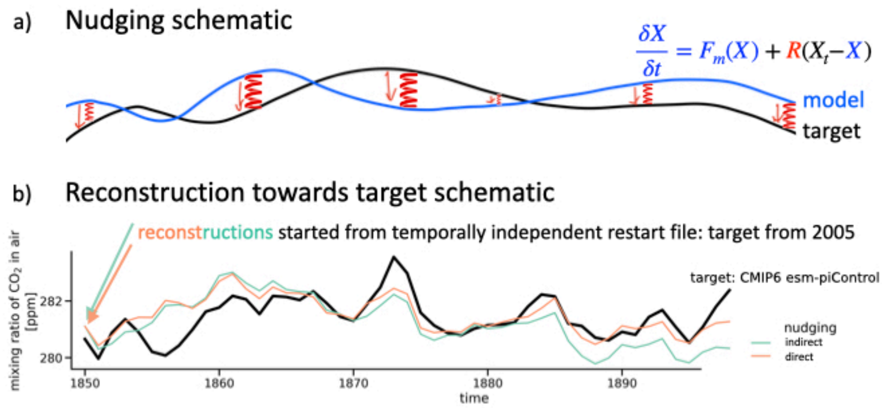

Figure 1: a) Schematic of nudging with relaxation constant. b) Schematic of reconstruction towards a target, where reconstructions are started from temporally independent restart files from the same simulation but 155 years later in time, i.e., 2005.

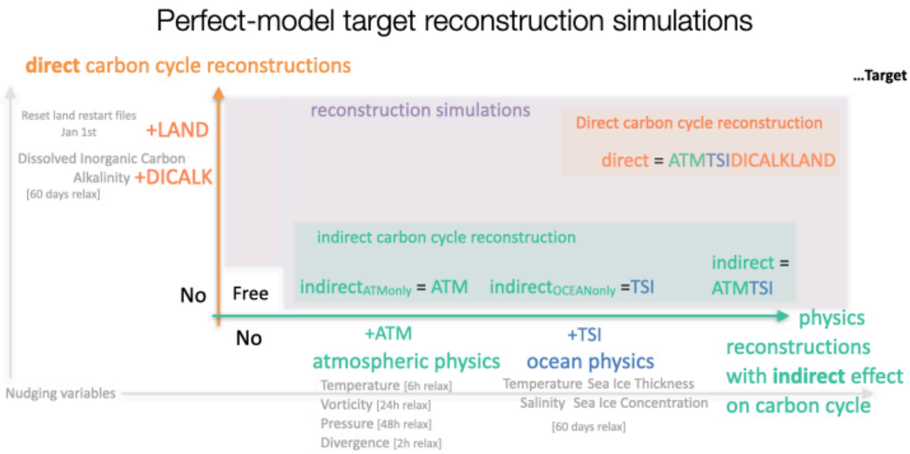

Figure 2: Schematic overview of perfect-model target reconstruction simulations showing which variables are reconstructed in which simulations.

believe that in a full-size manuscript, such approach is adequate. We will revise the manuscript to reduce such jumping, to consolidate the discussion of result with the purpose of enhancing the readability of the manuscript. Please note that in Fig. 2 we discuss land and ocean CO_2, one driver variable and the result in atmospheric CO_2. As we first discuss ocean in section 3.2, then land in section 3.3 and finally the atmosphere 3.4, the discussion of Fig. 2 is indeed spread out.

The conclusions seem a bit wanting of the opportunities for future investigation. Rather than being satisfied with "We conclude that the indirect carbon cycle reconstruction serves its purpose." It would be much more productive to point out what alternatives to nudging might provide superior options for future work. It should also be noted in the conclusion that the present work does not address the potential role for structural uncertainty, and potential for ecosystems to be more complex than represented in the current model and thus needing external constraint and providing a potential advantage to "direct initialization".

In response to this comment, we add to the text: "However, our study is designed and conducted in an idealized framework. When transferring our results into assimilation of real-world observations and its implications on predictability, structural uncertainties (model resolution in space and time) and missing ecosystem processes need to additionally be dealt with." And: "Future studies, especially those aiming to address regional marine ecosystem services, could consider a wider range of assimilation techniques and data breadth. Furthermore, more the implications of more advanced data assimilation techniques for the carbon cycle [Evensen, 1994; Han et al., 2004; Balmaseda et al., 2007; Zhang et al., 2007] should be explored."

Technical comments:

p1,ln8 – "We nudge variables from this target onto arbitrary initial conditions 150 years later mimicking an assimilation simulation generating initial conditions for hindcast experiments of prediction systems" I don't understand how this process works from this description. There is also a comma missing after "later" Instead, it sounds like the authors "nudged variables towards simulations from the same run 150 years earlier" to create a reconstruction of the target dataset.

Indeed, we nudged variables from the target simulation onto a temporally independent restart file, where we took a restart file from the same simulation but from 155 years after the nudging period. We added Fig. 1 for clarity and clarified the sentence to "We nudge variables from this target onto arbitrary initial conditions from a restart file 155 years later, mimicking an assimilation simulation".

P1 ln12 - I don't quite understand the distinction between "direct reconstruction" and indirect reconstruction". It is not defined in the abstract.

We will rewrite the abstract and make these definitions more clear.

Abstract overall – I think this is the longest abstract I have ever seen, yet it only describes concepts vaguely. I recommend the authors strip out the details of internally defined distinctions and spend more time on the implications of "We find improvements in global carbon cycle predictive skill from direct reconstruction compared to indirect reconstruction. After correcting for mean bias, indirect and direct reconstruction both predict the target similarly well and only *moderately worse than perfect initialization after the first lead year."*

We will shorten the abstract accordingly.

Ln41 – "where the forecast is started from" is redundant.

Deleted.

Ln 44 – does "indirectly" translate to "uninitialized"?

No. "Uninitialized" would mean that the model simulation is not initialized from reconstructions/observational products. "indirectly" refers to initialised simulations in which the climate state is initialized, but the carbon cycle is not initialized directly from data. Yet it is not uninitialized like in the historical CMIP6 simulations, where climate evolution of internal variability does not necessarily follow the observed/target climate evolution. We add explanation for the terminology used in the manuscript.

Ln 55 – This sentence is an identity "In this perfect-model target reconstruction framework, we have perfect knowledge about the ground truth and a perfect model"

It is a definition and explicit description.

Ln 58 – "Originally"? A reference should be provided as to the early work that is being invoked.

Added [Schneider and Griffies, 1999] and [Meehl et al., 2009] for an overview of initialized predictions.

Ln 60 and 61 – This appears to be describing results and conclusions of the present work. References should be provided to establish the literature context (as is done on ln 62).

Ln 60 is now changes as introductory sentence for paragraph, more citations now added: We describe what nudging is used for [Jeuken et al., 1996] and list previous studies on the impacts of nudging on circulation of ocean-atmosphere interaction [Zhu and Kumar, 2018] and biogeochemistry [Toggweiler et al., 1989].

Ln 65 – How do you know about these "severe consequences"? what is the citation? I know that this problem is discussed in the following, but there must be others:.

We refer to the studies cited above and missed to include Park et al., 2018 before.

Ln 91 – This is a strange justification. One could make the same argument for N2 or O2... presumably the reason for focusing on carbon has more to do with relevance to society. Is the question being answered why land and ocean are being treated together? If so, perhaps "We focus on the combined ocean and *land aspects of the carbon cycle because this allows us to explore the implications of flux predictability for atmospheric CO2 as well-mixed greenhouse gas."*

The sentence meant to strengthen why we need to address this issue in global simulations and why only looking at CO_2 fluxes would not suffice. Changed to "We focus on the global carbon cycle, because the land and ocean carbon cycle control the internal variability of atmospheric CO_2."

Ln 244-245 – "dominated by the bias of pCO2" instead of the bias in temperature?

CO_2 flux is influenced by oceanic pCO_2, which is influenced by oceanic temperature, circulation and biology.

Ln 248 – The description of this figure suddenly stops without addressing the XCO2 panels.

We discuss land and atmosphere in a later section. We had to decide: either discuss all variables first and the reader has to jump in between indirect and direct reconstructions, or discuss all reconstructions first and then variables are separated between section (our choice). We add a note where the land and atmosphere subfigures will be discussed.

Figure 4 – it would be helpful for the reader to see the comparative lines for Indirect Atm only to compare with Figure 2 and Figure 3

Because the manuscript is already quite long and we want to confront the direct and the indirect method, we choose to only show a comparison of indirect (atmosphere and ocean nudging) with direct (atmosphere, ocean and carbon cycle). We included indirect ATM in figures 1 and 2 to show the improvement of ocean assimilation compared to just assimilating the atmosphere, as it is done in the Global Carbon Budget. However, our primary goal is to compare the current standard of the indirect carbon cycle reconstruction by atmosphere and ocean assimilation in recent studies [Li et al., 2019; Yeager et al., 2018; Park et al., 2019].

Ln 302 – not sure why this sentence has its own paragraph

We tried to structure of the atmospheric CO_2 reconstruction subsection in a way that we first describe the indirect method and then the direct carbon cycle reconstruction. This single sentence section 3.4 introduces the next two sections 3.4.1 and 3.4.2. We will merge this into few sections.

Ln 347 - It is only here, after Figure 5 is presented, that I get to find why Figure 2o looks so much like Figure 2m. If I understand correctly, it is a coincidence – Figure 2m is high because the surface temperature is high, while Figure 2o is high because the land releases CO2 over the course of the year do to the climate mismatch. A statement to this effect near Ln 248 before moving on to Figure 3 would help orient the reader.

We agree that this explanation for the similar plots in 2m,o is missing. We add to section 3.4.2, where we describe direct carbon cycle reconstruction (after previous Ln 358): "There are very similar spatial distribution of the bias for the indirect atmosphere-only and the direct carbon cycle reconstruction [Fig. 2m,o], with different underlying causes. In the indirect atmosphere-only reconstruction, the XCO_2 bias is high [Fig. 2m], because the carbon cycle outgasses due to higher global temperatures due to nudging [Fig. 1d]. However, in the direct reconstruction the bias is high [Fig. 2o], because the carbon stocks from land and ocean are reconstructed without mass conservation. Here, the temperature bias is smaller than in the indirect atmospheric nudging, because the ocean temperature nudging stabilizes the global heat content. "

Ln 407 – "but below the initialized" is unclear, is "but drifts slightly below the initialized value over the course of the simulation" intended?

Indeed unclear. We change to ", but the error stays below the direct initialization error, which suffers from the bias in the direct reconstruction simulation."

Ln 422 – "For a real-world application, our direct land carbon reconstruction method cannot be used." I would disagree with this statement and should change "cannot" to "should not". The easiest form of data assimilation for land would be to simply over-write the vegetation biomass periodically from a satellite product, something very similar in principle to what is being done here. I think the more interesting question that is answered here is why that is a bad idea. I think this is a point very much worth making as satellite products become more diverse and land initialization approaches are considered.

We agree that our statement was a bit too strong. We change to "should not". We append discussing: "In practice satellite products of carbon cycle variables could be assimilated into the model periodically or at each time step. However, just strong interference with the model will likely result in strong drifts, especially in dependent variables. For useful real-world applications of land carbon cycle assimilation, sequential [Evensen, 1994; Balmaseda et al., 2007; Zhang et al., 2007] or variational data assimilation techniques [Han et al., 2004] could be used for initialization. But still the problem of data availability for the reforecast period remains."

Ln 424 – This conclusion appears to be the crux of the paper – that the nudging technique introduces such large biases in climate mean state as to make the "direct" approach incompatible with the original model. I am not an expert on physical data assimilation, but isn't that the reason that ensemble Kalman filter is used rather than nudging? Would one expect these other techniques that do not shift the ITCZ or dampen Southern Ocean winds to also find a "trivial" role for BGC initialization?.

We agree that the biases after nudging are substantial, which is the case for all reconstruction simulations, which all lead to their own peculiar differences as Reviewer notes above. We want to note that we are not aware of other studies trying reconstruct a model climatology with that same model target. Usually, observations are reconstructed and therefore there is no clear target to compare to.

The nudging method is one of the simplest reconstruction methods. Because it can be quite strong and abrupt, ensemble Kalman filter is replacing the method over time. We add in section 5: "Haney reconstruction is the simplest approach to data assimilation allowing little flexibility to the model. Many centers are now transitioning towards the ensemble Kalman filter data assimilation, which allows more variability [Park et al., 2019; Brune and Baehr, 2020]. Applying such techniques to the carbon cycle may lead to better reconstructions."

457 - Rather than being satisfied with "We conclude that the indirect carbon cycle reconstruction serves its purpose." It would be much more productive to point out what alternatives to nudging might provide superior options for future work. It should also be noted in the conclusion that the present work does not address the potential role for structural uncertainty to provide an advantage to "direct initialization"

Inspired by the above comments, we already added more recent methods applicable to carbon cycle reconstruction. We clarify: "We conclude that the indirect carbon cycle reconstruction serves its purpose of reconstructing variation in the global carbon cycle."

We agree that we only use one model to reconstruct to itself. Therefore we do not have any structural uncertainty other than the reconstruction method itself or processes missing in our framework. We agree when reconstructing the real world, our model lacks processes and resolution contributing to, our framework lacks structural uncertainty. These points are added to section 5.

References

Balmaseda, M. A., D. Dee, A. Vidard, and D. L. T. Anderson (2007). "A Multivariate Treatment of Bias for Sequential Data Assimilation: Application to the Tropical Oceans". en. In: *Quarterly Journal of the Royal Meteorological Society* 133.622, pp. 167–179. DOI: 10/czgj3m.

Brune, Sebastian and Johanna Baehr (2020). "Preserving the Coupled Atmosphere–Ocean Feedback in Initializations of Decadal Climate Predictions". en. In: *WIREs Climate Change* 11.3, e637. DOI: 10/ghtnt8.

Estella-Perez, Victor, Juliette Mignot, Eric Guilyardi, Didier Swingedouw, and Gilles Reverdin (2020). "Advances in Reconstructing the AMOC Using Sea Surface Observations of Salinity". en. In: *Climate Dynamics* 55.3, pp. 975–992. DOI: 10/gkbhsh.

Evensen, Geir (1994). "Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics". en. In: *Journal of Geophysical Research: Oceans* 99.C5, pp. 10143–10162. DOI: 10/fpjxh8.

Fransner, Filippa, François Counillon, Ingo Bethke, Jerry Tjiputra, Annette Samuelsen, Aleksi Nummelin, and Are Olsen (2020). "Ocean Biogeochemical Predictions—Initialization and Limits of Predictability". English. In: *Frontiers in Marine Science* 7. DOI: 10/gg22rr.

Friedlingstein, Pierre et al. (2020). "Global Carbon Budget 2020". English. In: *Earth System Science Data* 12.4, pp. 3269–3340. DOI: 10/ghn75s.

Han, Guijun, Jiang Zhu, and Guangqing Zhou (2004). "Salinity Estimation Using the T - S Relation in the Context of Variational Data Assimilation: SALINITY ESTIMATION". en. In: *Journal of Geophysical Research: Oceans* 109.C3. DOI: 10/b34qr8.

Ilyina, T. et al. (2021). "Predictable Variations of the Carbon Sinks and Atmospheric CO2 Growth in a Multi-Model Framework". en. In: *Geophysical Research Letters* 48.6, e2020GL090695. DOI: 10/ghsn7h.

Jeuken, A. B. M., P. C. Siegmund, L. C. Heijboer, J. Feichter, and L. Bengtsson (1996). "On the Potential of Assimilating Meteorological Analyses in a Global Climate Model for the Purpose of Model Validation". en. In: *Journal of Geophysical Research: Atmospheres* 101.D12, pp. 16939–16950. DOI: 10/d64x9q.

Li, H., T. Ilyina, W. A. Müller, and P. Landschützer (2019). "Predicting the Variable Ocean Carbon Sink". en. In: *Science Advances* 5.4, eaav6471. DOI: 10/gf4fxm.

Lovenduski, N. S., S. G. Yeager, K. Lindsay, and M. C. Long (2019). "Predicting Near-Term Variability in Ocean Carbon Uptake". In: *Earth System Dynamics* 10.1, pp. 45–57. DOI: 10/gfvxkc.

Meehl, Gerald A. et al. (2009). "Decadal Prediction: Can It Be Skillful?" In: *Bulletin of the American Meteorological Society* 90.10, pp. 1467–1486. DOI: 10/dpsjbp.

Park, Jong-Yeon, Charles A. Stock, Xiaosong Yang, John P. Dunne, Anthony Rosati, Jasmin John, and Shaoqing Zhang (2018). "Modeling Global Ocean Biogeochemistry With Physical Data Assimilation: A Pragmatic Solution to the Equatorial Instability". In: *Journal of Advances in Modeling Earth Systems* 10.3, pp. 891–906. DOI: 10/gddxmt.

Park, Jong-Yeon, Charles A. Stock, John P. Dunne, Xiaosong Yang, and Anthony Rosati (2019). "Seasonal to Multiannual Marine Ecosystem Prediction with a Global Earth System Model". en. In: *Science* 365.6450, pp. 284–288. DOI: 10/gf7fbj.

Pohlmann, Holger, Johann H. Jungclaus, Armin Köhl, Detlef Stammer, and Jochem Marotzke (2009). "Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic". In: *Journal of Climate* 22.14, pp. 3926–3938. DOI: 10/cdvhcr.

Rast, Sebastian, Renate Brokopf, Monika Esch, Veronika Gayler, Ingo Kirchner, Luis Kornblueh, Andreas Rhodin, and Uwe Schulzweida (2012). *User Manual for ECHAM6*. Tech. rep. Hamburg. URL: https://icdc.cen.uni-hamburg.de/fileadmin/user_upload/icdc_Dokumente/ECHAM/echam6_userguide.pdf.

Ruprich-Robert, Yohan et al. (2021). "Impacts of Atlantic Multidecadal Variability on the Tropical Pacific: A Multi-Model Study". en. In: *npj Climate and Atmospheric Science* 4.1, pp. 1–11. DOI: 10/gkb6tb.

Sarmiento, Jorge Louis and Nicolas Gruber (2006). *Ocean Biogeochemical Dynamics*. eng. Princeton, NJ: Princeton Univ. Press. URL: https://press.princeton.edu/books/hardcover/9780691017075/ocean-biogeochemical-dynamics.

Schimel, David S. (1995). "Terrestrial Ecosystems and the Carbon Cycle". en. In: *Global Change Biology* 1.1, pp. 77–91. DOI: 10/dw5kbg.

Schneider, Tapio and Stephen M. Griffies (1999). "A Conceptual Framework for Predictability Studies". en. In: *Journal of Climate* 12.10, pp. 3133–3155. DOI: 10/cf6zsg.

Séférian, Roland, Laurent Bopp, Marion Gehlen, Didier Swingedouw, Juliette Mignot, Eric Guilyardi, and Jérôme Servonnat (2014). "Multiyear Predictability of Tropical Marine Productivity". en. In: *Proceedings of the National Academy of Sciences* 111.32, pp. 11646–11651. DOI: 10/f6cgs3.

Servonnat, Jérôme, Juliette Mignot, Eric Guilyardi, Didier Swingedouw, Roland Séférian, and Sonia Labetoulle (2015). "Reconstructing the Subsurface Ocean Decadal Variability Using Surface Nudging in a Perfect Model Framework". en. In: *Climate Dynamics* 44.1-2, pp. 315–338. DOI: 10/f6v7kq.

Spring, Aaron and Tatiana Ilyina (2020). "Predictability Horizons in the Global Carbon Cycle Inferred From a Perfect-Model Framework". In: *Geophysical Research Letters* 47.9, e2019GL085311. DOI: 10/ggtbv2.

Toggweiler, J. R., K. Dixon, and K. Bryan (1989). "Simulations of Radiocarbon in a Coarse-Resolution World Ocean Model: 1. Steady State Prebomb Distributions". en. In: *Journal of Geophysical Research: Oceans* 94.C6, pp. 8217–8242. DOI: 10/ffvkfj.

Yeager, S. G. et al. (2018). "Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model". In: *Bulletin of the American Meteorological Society*. DOI: 10/gddfcs.

Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg (2007). "System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies". EN. In: *Monthly Weather Review* 135.10, pp. 3541–3564. DOI: 10/dkvk78.

Zhu, Jieshun and Arun Kumar (2018). “Influence of Surface Nudging on Climatological Mean and ENSO Feedbacks in a Coupled Model”. en. In: *Climate Dynamics* 50.1, pp. 571–586. DOI: 10/gcwrz4.