Pilcher et al. used an ocean-biogeochemical model to investigate carbonate system variability in the Bering Sea, with a focus on the long-term trends of ocean acidification (OA) variables (Ω, pH, H+, pCO2), motivated by the need of relevant OA indices for marine resources management. They extended the temporal coverage of previous modeling studies, quantifying spatiotemporal trends during 1970-2022. Their model results showed a significant acceleration of the OA trend over the last 25-years. The simulated bottom trends are greater than the surface trends, presumably associated with an increased respiration/remineralization in response to enhanced phytoplankton production. This is an interesting and valuable study that contributes to better understand changes in the carbonate system in the Bering Sea, but I think additional work is required to clarify/improve the model settings and further explain the model results.

RESPONSE: We thank the reviewer for the helpful feedback on our manuscript. We have conducted the additional requested analysis and have responded to the comments below. We think incorporating this feedback along with the comments from Reviewer 2 have greatly improved the manuscript.

Main comments

1) Model forcing

The main results in this study relate to the long-term patterns in OA variables. A relevant question is therefore how robust the derived 1970-2022 trends are, and I have some concern about this. The author recognized potential issues associated with the use of multiple products to derive the surface model forcing (CORE for 1970-1994, NCEP-CFSR for 2011-2021, and NCEP-CFSv2 for 2011-2022) and boundary conditions of physical variables (Northeast Pacific model (NEP) for 1970-1994 and CFS for 1995-2022). Model results show that the CORE-to-CFS transition most likely altered the long-term patterns in the OA progression. Maybe I am missing something, but there are available atmospheric reanalysis products that cover the entire study period (e.g., ERA5: 1950-2022), so I am not clear why the authors decided to use those three different products. Could you clarify this?

RESPONSE: We thank the reviewer for this insightful comment. We agree with their concerns regarding the shift in forcing from CORE to CFSR, hence why we made the decision to be explicitly clear on the timeframe over which the relevant trends develop. Our intention is to highlight that, while the shift from CORE to CFSR generates some artificial trends in certain variables (e.g. salinity), our primary result of enhanced bottom water OA rates compared to the surface is based on the CFSR forced simulation and is not artificially generated by the switch in forcing. Whether this result is constrained to just the CFSR forcing (as opposed to a different forcing product) is a separate issue (see next comment response). We chose to use the CORE and CFSR forcing because we have an extensive history of utilizing this forcing for Bering Sea ROMS model projects (e.g. Hermann et al., 2016; Pilcher et al., 2019; Kearney et al., 2020) and have found it to work quite well for the Bering Sea region. Furthermore, we are also in the process of testing 9-month seasonal forecasts, which utilize CFS for the

retrospective seasonal reforecasts. Bottom temperature seasonal forecasts have already been developed (e.g. Kearny et al., 2021) and we are now testing the biogeochemistry (Pilcher et al., in prep). Because these reforecasts need to be skill assessed compared to a hindcast, using the CFSR hindcast is the best option. Lastly, while there are some atmospheric reanalysis products that are available back to 1970, it is often more difficult to obtain an ocean reanalysis product back to that timeframe. For example, while ERA5 starts in 1950, the corresponding ocean reanalysis product (ORA5) starts in 1979. Thus, we would still have to combine multiple ocean reanalysis products to get the boundary conditions back to 1970, and our analysis of the salinity changes highlight that it's the transition between these boundary conditions that can induce the artificial trends. We have added some additional text throughout the manuscript to help clarify.

An additional CORE-forced hindcast ending in 2003 was intended to clarify the CORE-CFS transition impact on the trends, but I do not think this extra analysis significantly helped to that goal. Instead of comparing if the CORE-CFS trends for 1970-2022 are like the CORE trends for 1970-2003, I would compare if trends derived from CORE and CFS forced experiments are consistent during the overlapping period of these two products. A comparison over the overlapping period could also help to clarify potential impacts on seasonal and interannual variability.

Maybe you could re-run the full hindcast using ERA5 as the only atmospheric forcing, so that the "forcing issue" would be limited to the NEP-to-CFS shift in the boundary conditions. In that case, you can run additional experiments to compare trends over the overlapping periods of NEP and CFS.

RESPONSE: Similar to the previous comment, our goal with this comparison is not to suggest that trends between the CORE and CFSR forced products are the same. Forcing products are imperfect and contain uncertainty which will impact our model results (e.g. Jung et al., 2014 doi:10.1002/2013GL059040; Lima et al., 2018 doi:10.1029/2018JC013919). Rather, our goal is to illustrate that our longterm trends are not artificially generated by the switch in forcing. A separate hindcast utilizing the ERA5 forcing for the entire timeframe is an interesting project idea, but this would be a substantial additional effort, that would then shift the focus of the project towards understanding model sensitivity to the forcing (both atmospheric and horizontal boundary conditions), which is not our goal here.

2) Salinity trend shift

Figure S1 show a strong trend shift in salinity associated with the change in the model forcing products (CORE to CFS). I suggest reporting the mean alkalinity series at surface and bottom as supplementary figure, since salinity and alkalinity are usually strongly correlated. Did you get a similar trend change in alkalinity? Since salinity and alkalinity are drivers of Ω, pH, and pCO2, then that shift could significantly impact all the reported OA trends. Could you discuss about it?

RESPONSE: This is a great suggestion, and we have added the alkalinity timeseries to the

supplement, along with the change in the DIC and TA boundary conditions that result from the salinity shift, following the recommendation of Reviewer 2. Similar to salinity, there is an overall decreasing trend in alkalinity, which appears to be largely related to the forcing shift, although there is still a slight decreasing trend over just the CFS timeframe. While this will impact the carbonate chemistry, the decrease in salinity will also decrease DIC, thus mitigating some of the effect. Additionally, the change in salinity from the forcing can modify the incoming boundary conditions, which have a relatively greater effect on DIC than TA due to the empirical relationship (i.e. comment raised by Reviewer 2). We also further expand on the drivers based on the Taylor series decomposition mentioned further below. We have expanded the following text in the manuscript to further describe these details and added a new supplemental figure. We also show below a figure for the TA trend in the 3 ESMs for comparison:

"The lateral boundary conditions for DIC and TA are calculated via linear regressions with salinity through the following equations below, derived from observational data collected primarily from 2008-2010 (Pilcher et al., 2019).

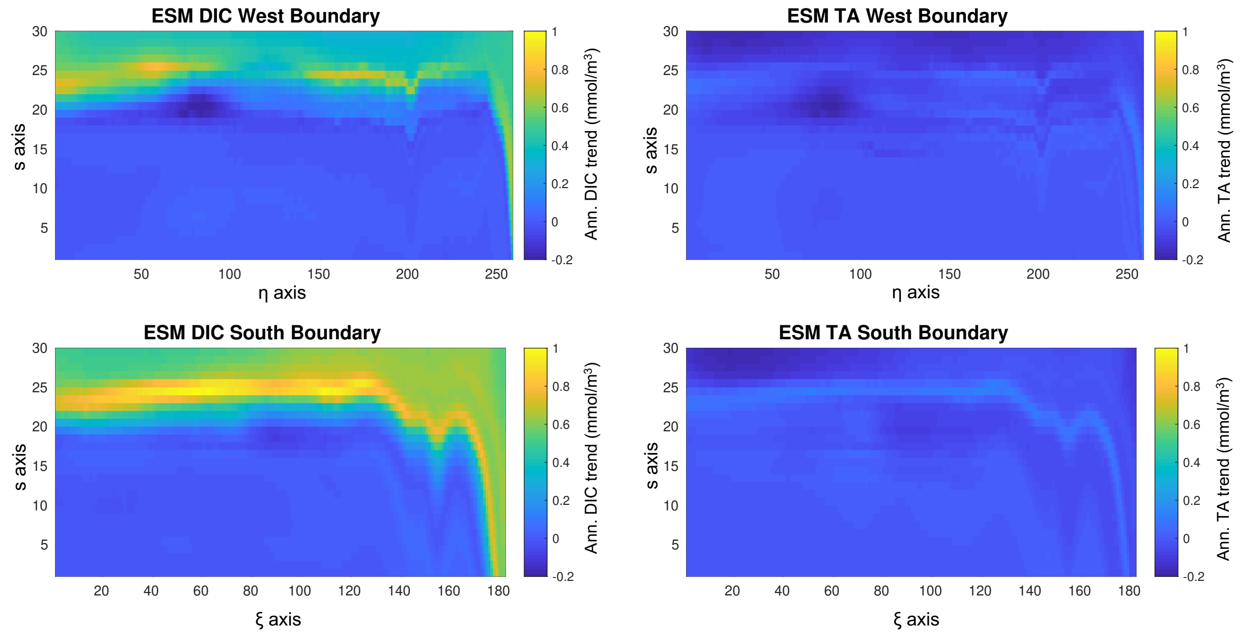
$$\boldsymbol{S < 32.6\ DIC = 58.5 * S + 191.2 + \Delta DIC(t)^{atmo}} \tag{1}$$

$$\boldsymbol{S\ \geq 32.6\ DIC = 140.4 * S - 2478.7 + \Delta DIC(t)^{atmo}} \tag{2}$$

$$\boldsymbol{S < 33.6\ TA = 49.6 * S + 600.6} \tag{3}$$

$$\boldsymbol{S\ \geq 33.6\ TA = 141.8 * S - 2494.4} \tag{4}$$

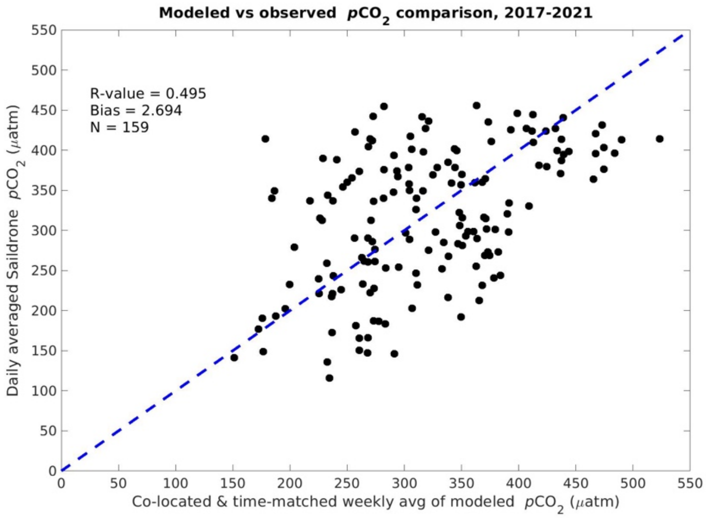
The salinity-DIC regression has changed over time as the oceanic uptake of CO_2 has increased the DIC concentration of waters, with no effect on salinity. Thus, using this same relationship for the boundary conditions at the start of the hindcast in 1970 would artificially increase DIC. To account for changes in DIC over time, we center the DIC-salinity relationship on the year 2009 (i.e. midpoint of 2008-2010 sampling timeframe) and subtract (add) DIC for years before (after) 2009. The DIC value added or subtracted (ΔDIC^{atmo} in equations 1-2) for year(t) is obtained from the linear trend in DIC (Fig. S1) calculated from the historical runs of the Coupled Model Intercomparison Phase 6 (CMIP6) over the 1970-2009 timeframe from the mean of three different Earth System Models (GFDL-ESM4, CESM2, and MIROC-ES2L). These three ESMs were selected as they have been used previously in the Bering10K regional dynamical downscaling (Cheng et al., 2021; Pilcher et al., 2022). We chose to use this method to gain the higher spatial resolution, particularly in the vertical, provided by the ESM output. We only use the DIC trend from the CMIP6 ESMs and omit any TA trend because the TA trends over this timeframe are much smaller and are tied to changes in salinity (Hinrichs et al., 2023), which is accounted for in our salinity-TA relationship at the boundary."



3) pCO2 patterns

There are not many observations during fall-winter, but the M2 records in 2021 suggests that the model is overestimating pCO2 during those seasons. If that is correct, then you could conclude that the model has an overall positive bias in pCO2 (and negative bias in pH), which could have a strong impact on the air-sea CO2 fluxes. You could discuss about it and tone done your results related to CO2 fluxes.

RESPONSE: We agree that the M2 comparison does suggest an overall positive bias in model pCO_2. As noted in the text, this bias appears to evolve from an underestimation in the spring pCO_2 drawdown, which leads to subsequent overestimations of pCO_2 throughout the summer. However, it should be noted that M2 is just a single point in the model domain, and we suggest caution when extrapolating a M2 bias to an overall model bias. For example, the Saildrone comparisons (Figure 5), which provide a much larger spatial footprint, also suggest some positive pCO_2 biases in early spring, but there are also negative biases apparent in the southward transect in 2019. A property-property comparison of all of these Saildrone observations, suggests a very small overall bias (Figure below). We lastly note that our total annual carbon flux values of 1.1-7.9 TgC/year compare reasonably well with the broad observational estimates of 2-67 TgC/year, and specifically the 6.8 TgC/year from Cross et al., 2013, which incorporates winter flux estimates missing from the earlier studies.



4) Underlying drivers

A Taylor series decomposition could provide valuable insights about the underlying drivers of OA variability, helping to identify the causes for the carbonate system trend changes, support the hypothesis of a biological driven increase in the bottom OA trends, and identify if salinity and alkalinity play any role on the OA progression changes. Although the authors mentioned that diagnostic mechanisms are beyond the scope of this study, I strongly recommend adding a Taylor decomposition analysis, especially considering that a paper describing historical OA pattern was already published (Pilcher et al., 2019).

RESPONSE: This is an excellent suggestion by the reviewer, and we have now included a new figure along with the accompanying text:

"To further understand the drivers behind changes in the carbonate chemistry, we also use a first-order Taylor series to decompose changes in pCO_2, Ω_{arag}, and $[H^+]$ into the four primary drivers:

$$\Delta\phi = \frac{\partial\phi}{\partial DIC}\Delta DIC + \frac{\partial\phi}{\partial TA}\Delta TA + \frac{\partial\phi}{\partial Salt}\Delta Salt + \frac{\partial\phi}{\partial Temp}\Delta Temp \tag{11}$$

Where $\Delta\phi$ represents the time change in the calculated carbonate parameter (pCO_2, Ω_{arag}, or $[H^+]$), and the four variables on the right-hand side of the equation account for the contributions of DIC, TA, salinity, and temperature respectively. The partial derivatives are calculated through small perturbations using CO2SYS (Lewis and Wallace, 1998; Sharp et al., 2023). We employ the Taylor series decomposition for both the entire 1970-2022 timeframe, in addition to the CFSR 1998-2022 timeframe (Fig. 15). This decomposition further highlights that the OA trends are driven by increasing DIC, particularly for bottom waters. Surface carbonate trends are also driven to a lesser extent by decreasing TA over the 1970-2022 timeframe, though this effect is somewhat diminished during the more recent 1998-2022. On this timeframe, warming temperatures emerge as a driver for surface and bottom pCO_2 and $[H^+]$, though still lower in magnitude than DIC."

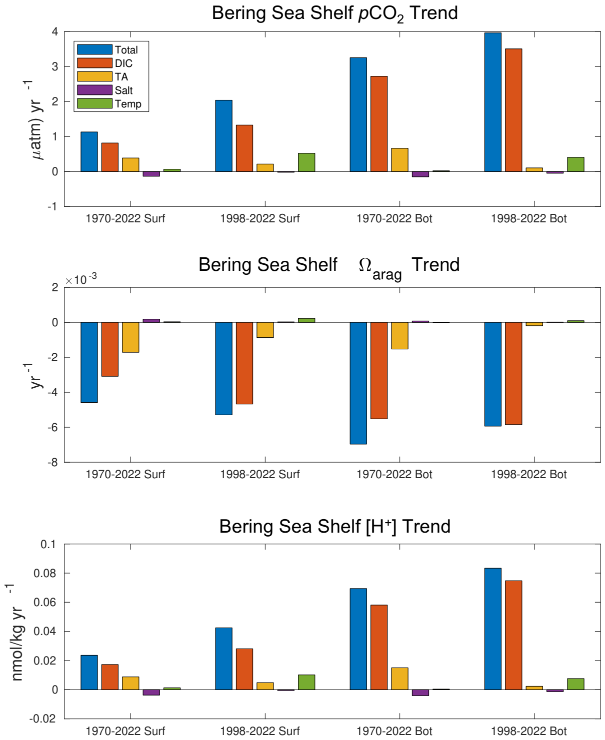


Figure 15: Taylor series decomposition of trends in pCO_2 (top), Ω_{arag} (middle), and $[H^+]$ (bottom) for surface and bottom waters over the 1970-2022 and 1998-2022 timeframes.

Minor comments:

I suggest tone done all the modeling results, considering that an ocean-BGC model rather suggest than demonstrate the OA patterns. For example, in the abstract: "surface Ωarag decreases by -0.043 decade-1 and surface pH by -0.014 decade-1" => "model results suggest that Ωarag decreases by 0.043 decade-1 and surface pH by -0.014 decade-1"

RESPONSE: Thank you for the suggestion, we have toned down some of the language for underlying results through the text and the abstract.

168-171: Could you provide the spatial resolution of the Northeast Pacific model and the CFS products used for the boundary conditions?

RESPONSE: We have added the spatial resolutions for both models.

174: How did you estimate the empirical climatological profiles for iron?

RESPONSE: We have added the following description to the text

"Water column iron concentrations are nudged towards empirical climatological profiles, which use an analytical function based on Seward line data in the Gulf of Alaska for coastal regions (Hinckley et. al, 2009). On-shelf values are set to 2.0 mmol/m^3 at the surface and 4.0 mmol/m^3 at depth, and this gradient transitions linearly to 0.01 mmol/m^3 at the surface and 2 mmol/m^3 at depth in water depths greater than 100m."

393: "variables are comparable" what do you mean?

RESPONSE: That the trends for all three variables at the surface are similar. For clarity, we have changed to "similar" rather than comparable.

415: "seasonally occurring" => provide the season name.

RESPONSE: Here, seasonally was not referring to a specific season. We have removed for clarity.

455: I suggest using the same x-axis range (1970-2022) for the two panels, independently that the M8 station data extend only until the 80s.

RESPONSE: Thank you for catching this, the x-axis was mislabeled as it does indeed extend from 1970-2022. We have corrected the axis label.