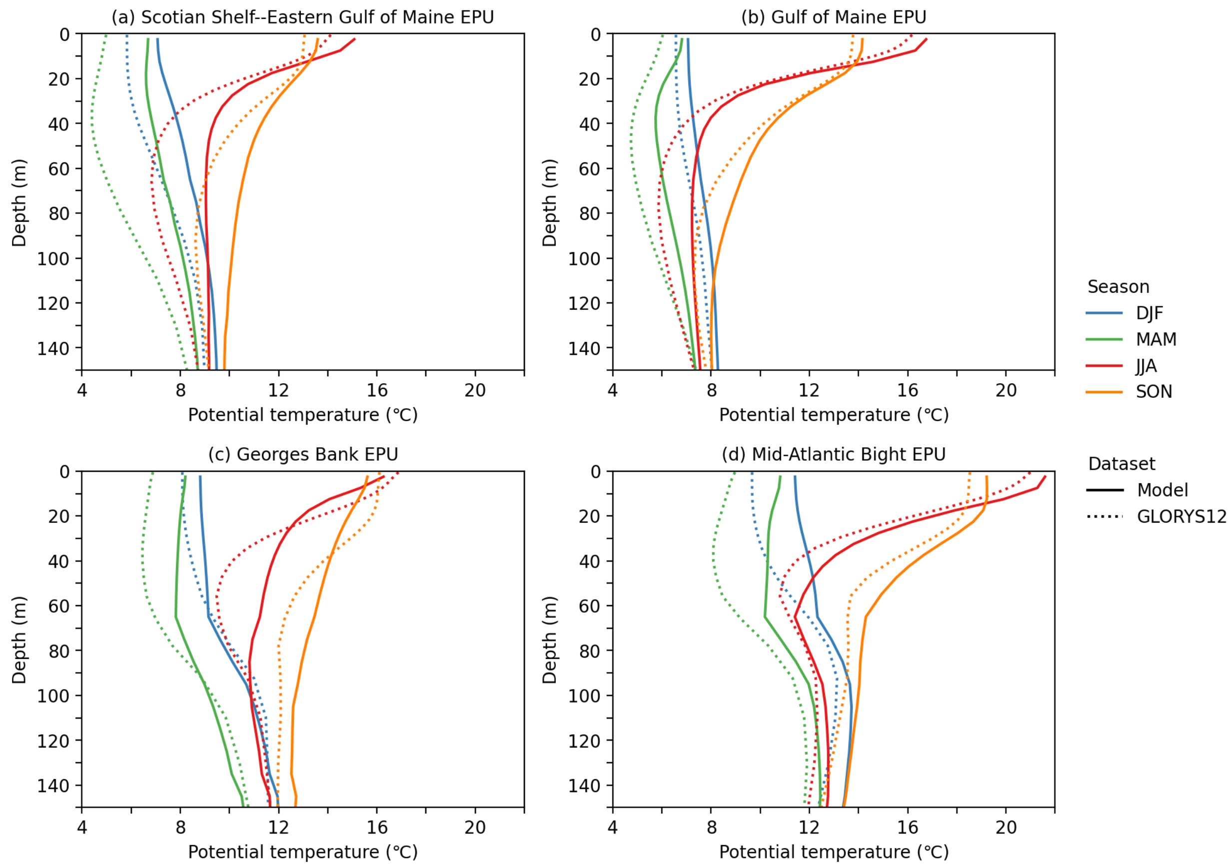
Summary:

We thank the reviewer for providing a detailed and well-organized set of suggestions to improve the manuscript. Below, we have restated each review comment in bold text followed by our reply.

General comments:

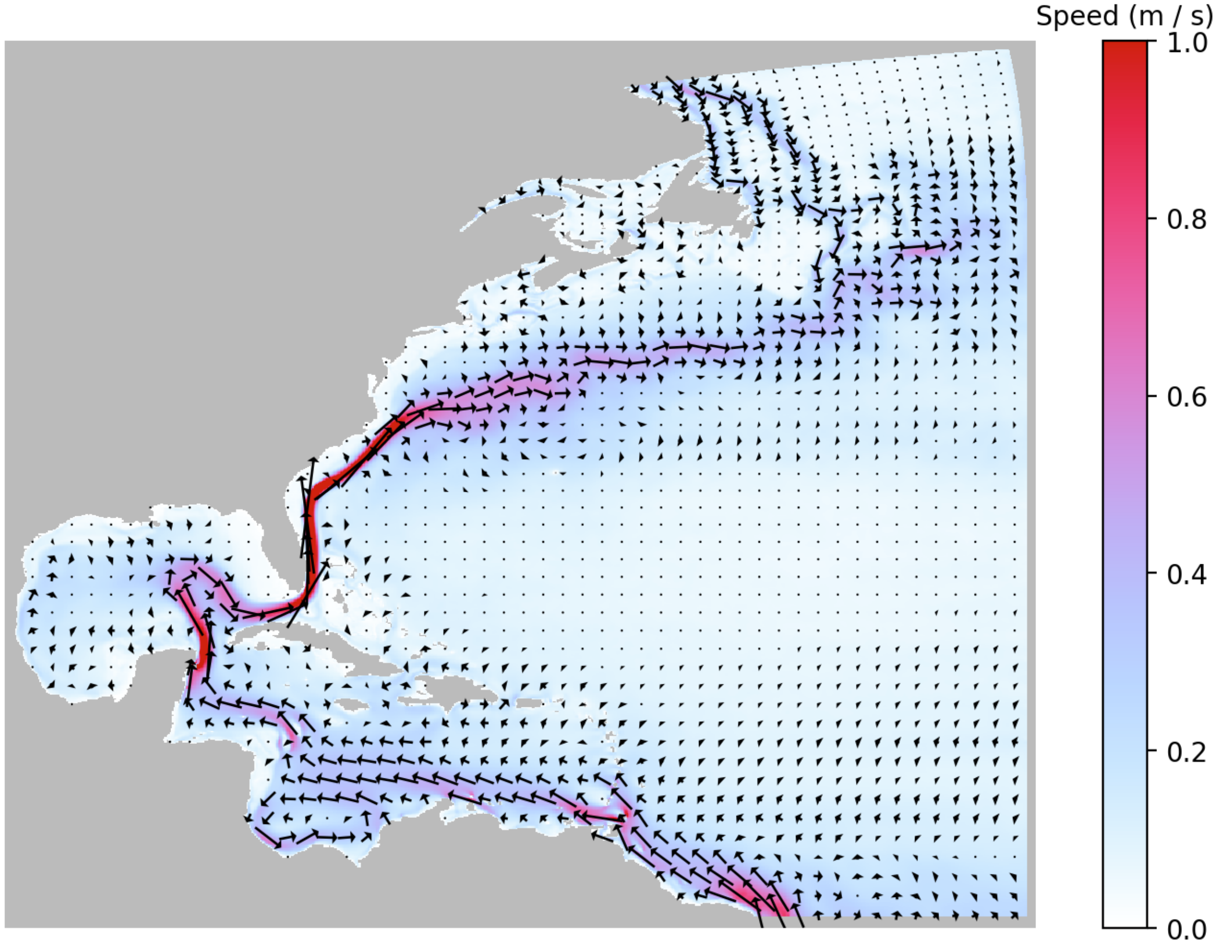
1. Water masses – Vertical sections. The comparison metrics used are very informative. However, in my opinion, it will be helpful to add some comparisons between the model and observations/reanalysis in terms of key vertical sections (or profiles) to better demonstrate the representation of different water masses in the model. For example, I suggest adding such comparisons for the winter and summer temperature and salinity climatologies for either: (i) sections of depth versus distance form the coast to beyond the shelf break, for maybe the Gulf of Maine+Georges Bank+slope, the MAB+shelf break and the Gulf of Mexico; or (ii) representative vertical profiles (perhaps averages) over the 4 EPUs. Consider doing the same for phosphate (or nitrate) and perhaps DIC.

We have inserted a new figure (now Figure 18) that shows vertical profiles of seasonal temperature climatologies for the 4 EPUs. We added only temperature because it has stronger seasonal variation than salinity, and we don't have reliable climatologies of biogeochemical variables for these regions (World Ocean Atlas would have data on these variables, for example, but it is not accurate at these fine coastal scales).



2. Main circulation features. In my opinion, it would beneficial to have a visualisation of the model mean circulation. I suggest adding a figure/map that compares the surface (or the vertical averaged) currents in the model to the reanalysis. The authors can decide how best to visualise this (perhaps using arrow plots with a background colormap for speed).

We have included below a figure of the model mean surface current speed and direction. This figure shows that the model includes the well-known large-scale features, including the Loop Current, Gulf Stream, inflowing Labrador current, and counterclockwise circulation around the Gulf of Maine. One feature that appears poorly simulated is the Mid-Atlantic Bight shelfbreak jet.



We have not added this visualization to the manuscript for several reasons. First, the geostrophic currents can already be visually approximated from the evaluation of mean sea surface height. Second, quality long-term observations to compare to are lacking. Even the GLORYS12 reanalysis does a poor job at reproducing currents near the coast; for example, it does not have the cyclonic circulation in the Gulf of Maine. Third, during the model simulation we saved fewer outputs for velocity, and some of these diagnostics like the mean speed are affected by the presence of tides. For the figure above, for example, we had to use monthly mean surface u and v velocity components, which will bias the results towards lower average speeds.

3. Streamlining (this is just a suggestion): The introduction is very informative; however, in my opinion, the discussion for the ocean conditions and variability in the North West Atlantic (lines 24-90) is a little too descriptive/long and could be condensed to better highlight the shortcomings due to limited availability of skilful high-resolution regional predictions and projections. I suggest streamlining lines 24-90, but this is just a suggestion and up to the authors' preference.

We appreciate the suggestion. We hope to follow this manuscript with studies on applications of the model, as well as eventual papers on continued model development and improvement. We have thus elected to leave the introduction as is, with the intention to provide a fairly comprehensive overview that sets the motivation for the present and future papers and can be referred to in the future.

4. Figures and latitude longitude (this is just a suggestion): I understand that the map-figures do not have latitude and longitude so as to preserve space and make them look more compact. However, adding latitude and longitude to all of the map-figures would help readers who are not familiar with the region to easily follow the features shown in these figures; particularly in figures that involve zooms in specific regions (e.g., Figures 6,13, 14, 20, 21, 23).

We have added lat/lon ticks and labels to all of the figures that zoom in to specific regions (Figures 6, 7, 13, 14 and 20, 21, 23 (now 21, 22, 24)).

5. Seasonal mean estimates. The seasons for temperature, and chlorophyll (Figures 3, 12, 23) are defined as: (i) December-February, (ii) March-May, (iii) June-August, (iv) September-November; while the seasons for nutrients (Figures 10 and 11) are defined as: (i) January-March, (ii) April-June, (iii) July-September, (iv) October-December. I was wondering if there is a reason for using different months to define the seasons for nutrients. If so, please explain it in the text. If not, it might be better to keep the definition of the seasons consistent for all the fields.

The different definition of the seasons used in the nutrient figures originates from the World Ocean Atlas dataset. WOA provides data in seasons starting with JFM, whereas in most cases we prefer to use the standard meteorological seasons starting with DJF. (World Ocean Atlas does provide a monthly climatology, which could be averaged to standard seasons; however, considering the sparseness of the observations in some areas,we prefer to use the WOA seasonal climatology which will have more data available for the WOA interpolation and smoothing steps). We have not changed the analysis, but have now highlighted in the text the different definition of the seasons and the reasoning why.

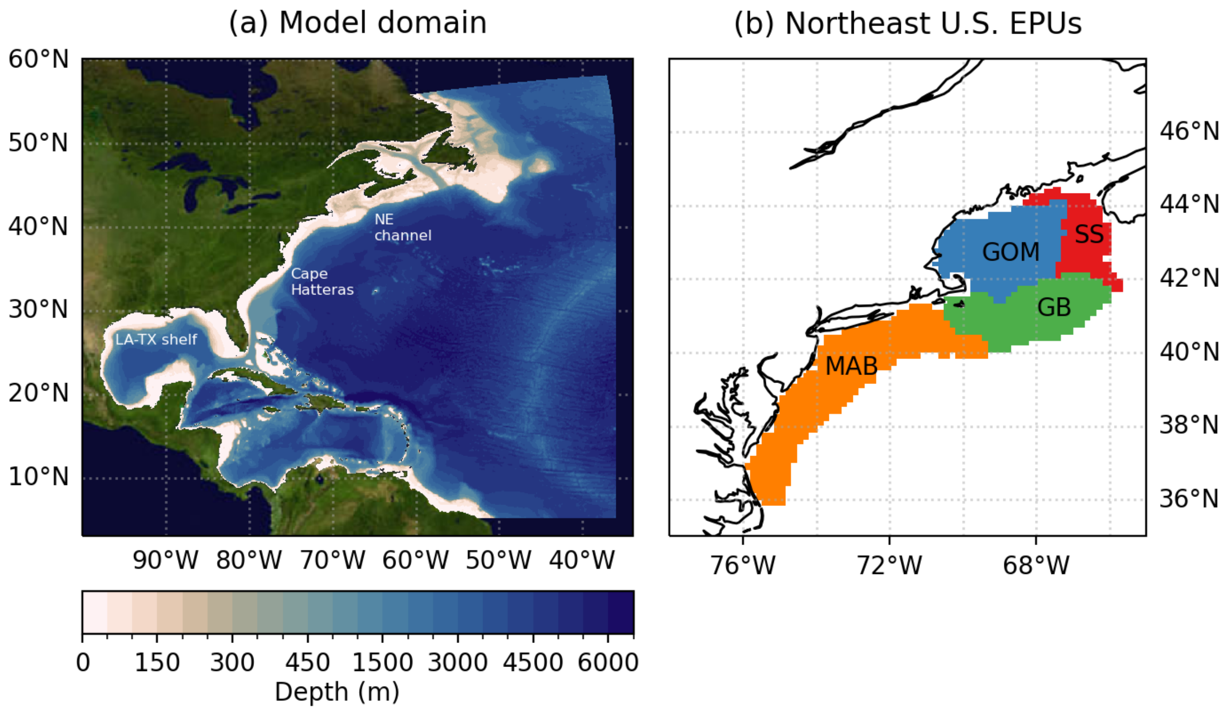
Specific comments:

6. Line 133. I am not sure what "… coastwide extend to address the prominent cross-boundary issue expected under climate change" means here. Maybe it refers to along-shelf propagating signals, such that for example you have a large domain covering the whole US East coast? If yes, in my opinion, your domain is still affected by cross-boundary issues, as the Labrador current is not resolved in the domain but rather prescribed at the north ocean boundary. Please consider clarifying/re-writing. (This is not a criticism as all the regional models are subject to this, but more a request for clarification.)

Our intent was to refer to issues occurring across fisheries management regions, Exclusive Economic Zones, and other geopolitical boundaries that our model's fairly large domain covers (for example, a favorable environment for a fish shifting from U.S. to Canadian waters). However, we realize that our wording was not clear and thank the reviewer for pointing this out. We have tried to clarify by now stating that the model “covers a large "coast-wide" domain to address the prominent climate impacts expected to extend across fisheries management regions and other traditional geopolitical boundaries”.

7. Figure 1. Consider adding the names of some key regions in figure 1.a: e.g., the Northeast Channel, Cape Hatteras, Texas, Louisiana, Florida, Gulf of Mexico, Gulf of Saint Lawrence etc.

We have added annotations for three potentially less-known features: the Northeast Channel, Cape Hatteras, and the LA-TX shelf.



8. Lines 209-213. There is an emphasis on obtaining reliable solutions without applying restoring (e.g., surface salinity restoring). In my understanding, the simulation covers about 25-30 years (+ spin-up), and I was wondering if 25-30 years is a short time period in terms of the model developing a significant drift. Hence, I am not sure if the absence of restoring in the 25-hindcast run is indicative of the model’s performance in longer simulations (e.g., climate projections), in terms of the emergence of significant drift; and if it will actually be beneficial to include a strategy for accounting for this drift in regional climate projections with your model. Maybe I have misunderstood something, but I suggest adding a brief discussion to clarify why it is expected that there will be no need for restoring to account for any drift in your regional ocean-only model under long-term simulations (longer than 25-30 years).

This is a fair point. During development of the model, our experience was that errors that noticeably impacted the solution typically manifested as a drift away from reality within about 10 years or less. However, we acknowledge that it is possible that smaller drifts remain that are not noticeable in our 27-year simulation but could become relevant over climate-scale simulations. We think this is not a critical point (although we are pleased that the 27-year run does not appear to need restoring, MOM6 does have the ability to restore and it could be added in the future if longer runs show that it is necessary), so we have de-emphasized it in the text (deleting "it is important to note" and "deliberate") and added a note that "it will, however, be necessary to confirm that the model remains reliable over simulations longer than the 27-year run examined here". Finally, we will note here but not in the text that for work in progress we have run historical experiments over 60 years long without restoring and still obtained stable simulations.

9. Lines 228-229. Please can you clarify how the rivers runoff salinity and temperature is treated (e.g., are you prescribing/assuming a constant 0 PSU salinity for river runoffs, or observed values?).

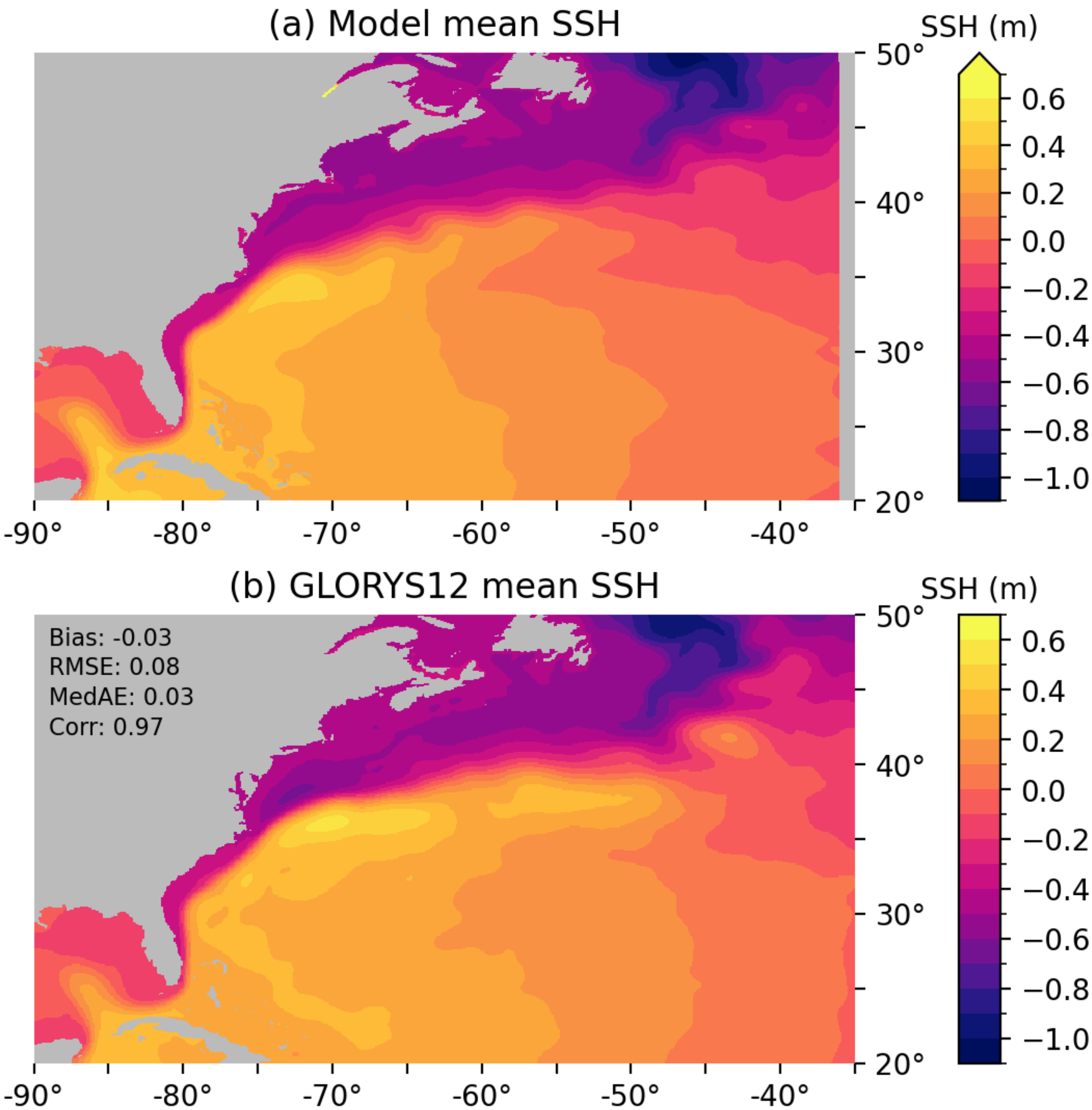
We have added this sentence: "River discharge entered with zero salinity and a temperature equal to the surface temperature of the discharge grid cell".

10. Line 404-405: I am not sure I understand what the feedback of biogeochemistry to tides will be in you model? (So I am not sure why there is expected to be even a negligible feedback). Please, can you clarify what this small feedback would involve (at least to the reply, as I was confused).

We have deleted the mention of the expected negligible feedback on tides and now just state that biogeochemistry was not included for the tide estimation simulation. "Negligible" was an understatement here and we apologize for making it unclear. The only feedback from the biology to the ocean physics in the model is the ability of the simulated chlorophyll to modify light absorption and heating in the water. In principle this could affect the stratification and the generation of internal tides, but it is improbable that this would produce a meaningful difference in the tidal analysis.

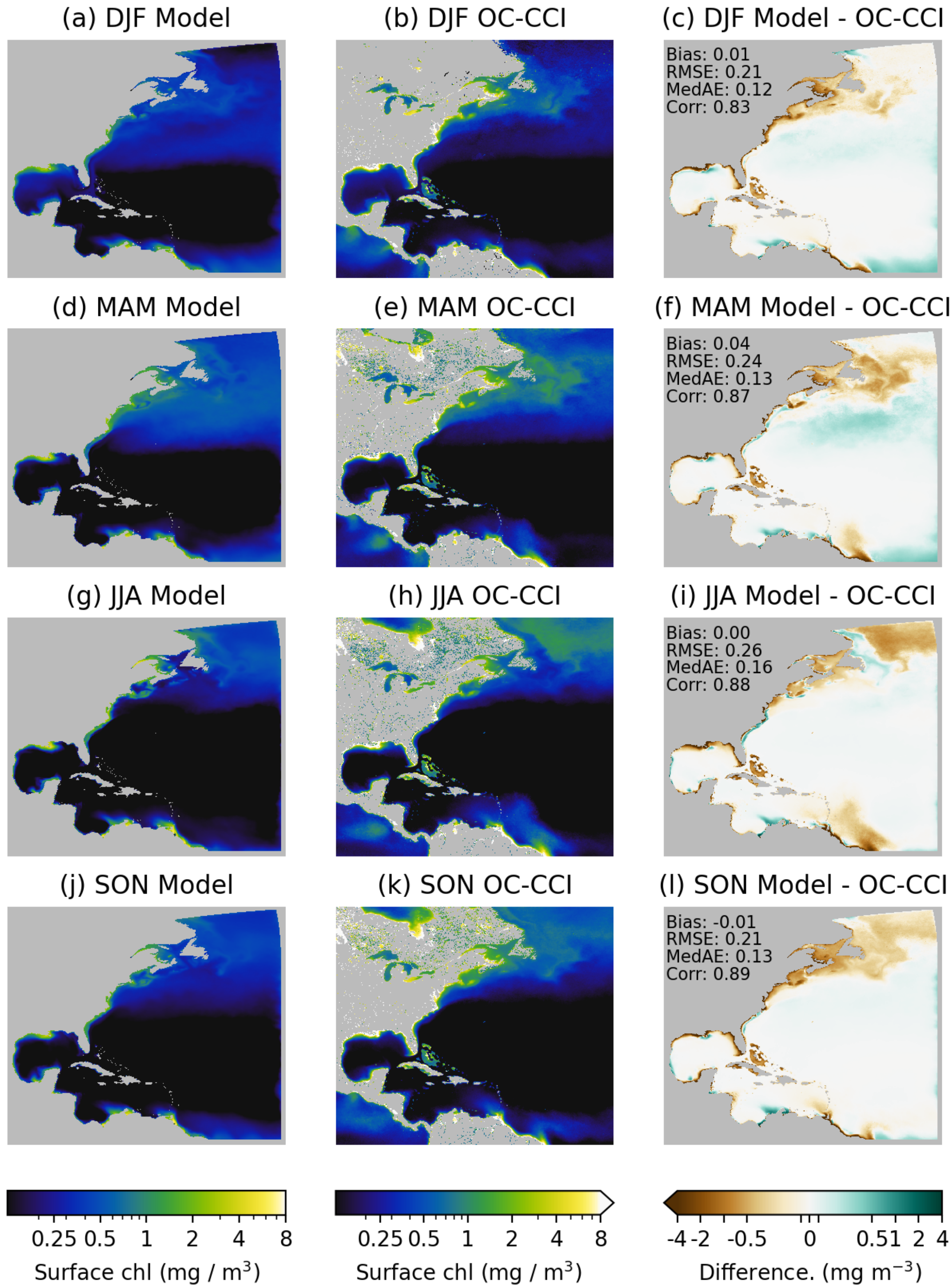
11. Lines 590-592 and Figure 7. To me, based on the 0.4 m shift in the colormap, it appears that the absolute dynamic topography and the model SSH have a difference/bias in magnitude. Is this maybe associated with the estimates of absolute dynamic topography and the geoid (I am not an expert on this so I am just curious about it)? I suggest, for clarity, to add the difference between model and observed SSH and the equivalent metrics as in the other figures (bias, RMSE, MedAR and Corr). This would help to better understand the magnitude and significance of the difference between the two datasets.

Yes, this is likely associated with a difference between the geoid used in the reanalysis data and the satellite product. To reduce confusion, and because the reanalysis assimilates satellite data and has essentially the same long-term mean spatial pattern, we have removed the satellite product from Figure 7. We added a note in the text that “The reanalysis assimilates satellite altimetry data and has a nearly identical long-term mean, aside from an apparent offset in the reference level”. We also added the skill metrics comparing the model to reanalysis mean SSH to Figure 7.



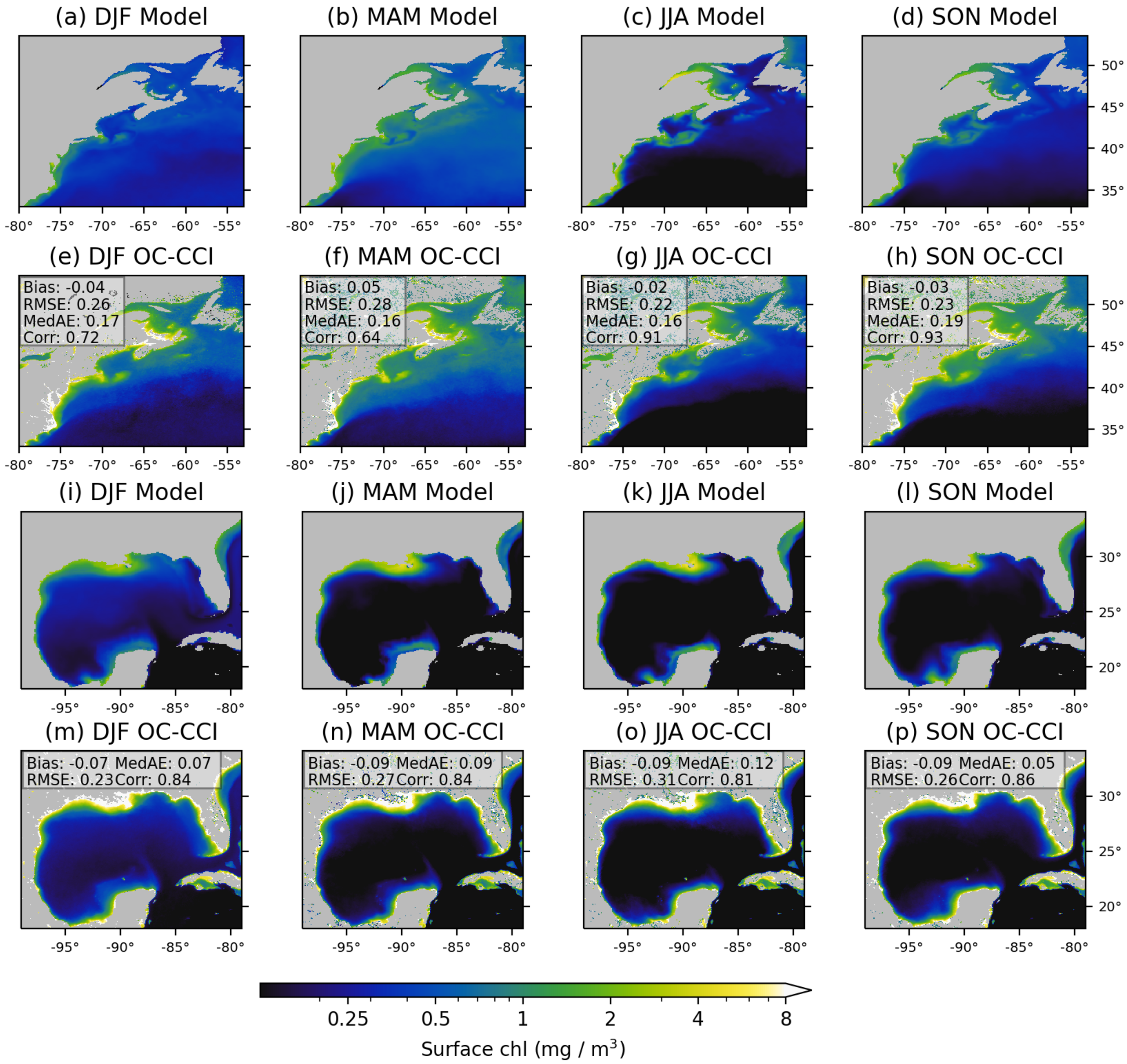
12. Figure 12. I suggest that you add the difference between model and OC-CCI satellite in the figure, as it is a little difficult to compare by eye where the model overestimates or underestimates the surface chlorophyll.

In the revised version we have added the suggested panels that show the difference.


(a) DJF Model
(b) DJF OC-CCI
(c) DJF Model - OC-CCI
Bias: 0.01
RMSE: 0.21
MedAE: 0.12
Corr: 0.83
(d) MAM Model
(e) MAM OC-CCI
(f) MAM Model - OC-CCI
Bias: 0.04
RMSE: 0.24
MedAE: 0.13
Corr: 0.87
(g) JJA Model
(h) JJA OC-CCI
(i) JJA Model - OC-CCI
Bias: 0.00
RMSE: 0.26
MedAE: 0.16
Corr: 0.88
(j) SON Model
(k) SON OC-CCI
(l) SON Model - OC-CCI
Bias: -0.01
RMSE: 0.21
MedAE: 0.13
Corr: 0.89
0.25 0.5 1 2 4 8
Surface chl (mg / m^3)
0.25 0.5 1 2 4 8
Surface chl (mg / m^3)
-4 -2 -0.5 0 0.51 2 4
Difference. (mg m^{-3})

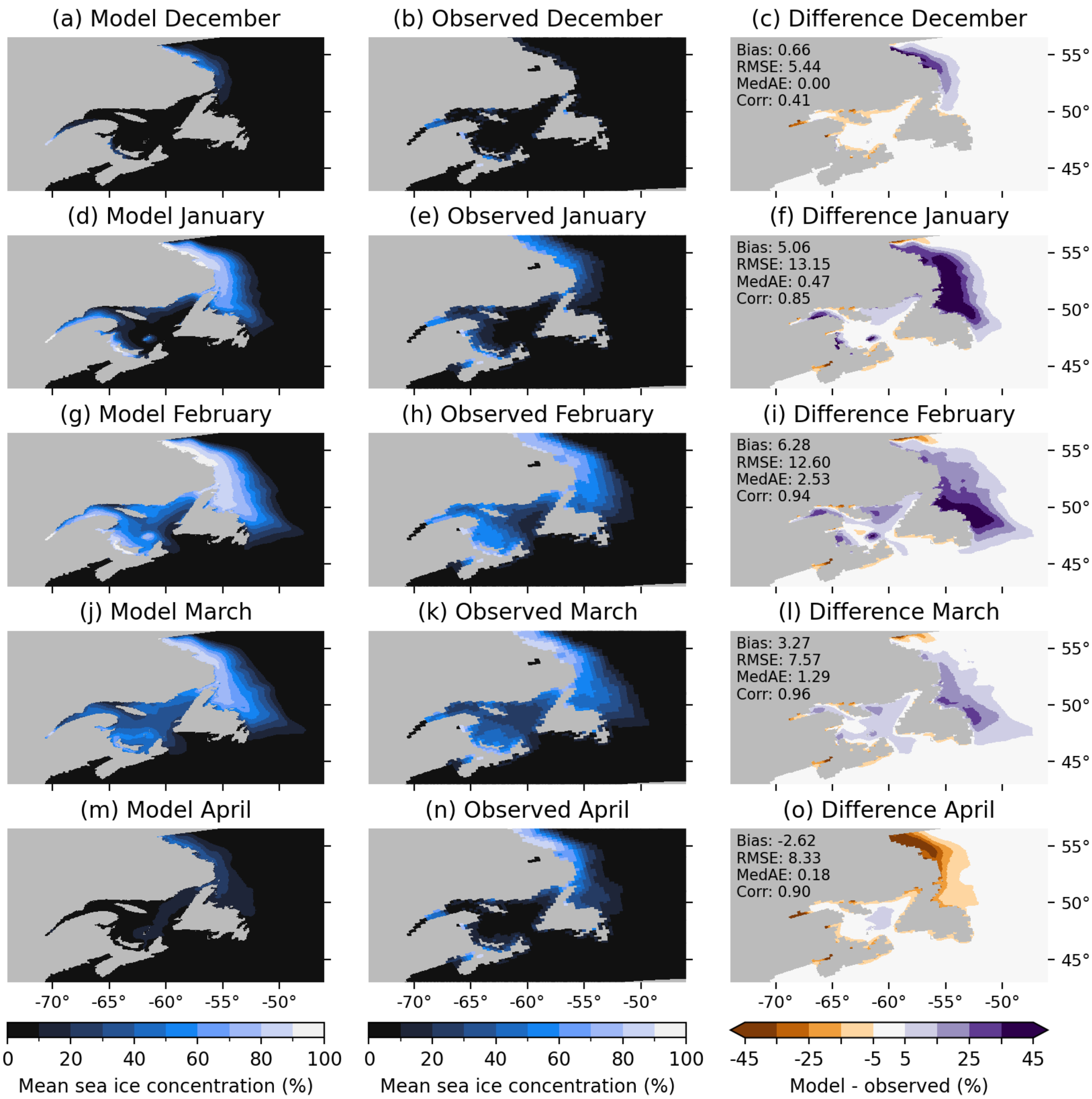
13. Figure 13. If I understood correctly, this is a zoom-in of figure 12 (maybe mention this in the Figure 13 caption). I suggest adding the bias, RMSE, MedAR and Corr metrics in the figure for the two different regions.

Yes, this figure is a close-up of Figure 12. We have now stated this in the caption, and added the suggested skill metrics to the panels as well.



14. Lines 727-728, Figure 21: It is a bit difficult to compare by eye the observations and model sea-ice concentration in Figure 21. Please consider adding the difference between observed and model in a third panel, as well as the associated metrics (bias, RMSE, MedAR and Corr).

We have added panels for the difference and included the skill metrics.



15. Lines 731-733. Please consider highlighting in the text that this is for the Gulf of St Lawrence, for example " … in sea ice coverage in the Gulf of St Lawrence, with correlation …).

We have updated this to read: "The time series of monthly sea ice extent in the Gulf of St. Lawrence (Fig. 22) shows that the model captures nearly all of the year-to-year variability in sea ice coverage in the Gulf…"

16. Typo, Lines 87-89: decadal time scale is repeated, maybe consider removing the "At decadal timescales" in the beginning.

Fixed in the revised version.