I very much appreciate the opportunity to review this manuscript. This paper focuses on evaluating seasonal predictability of surface temperature and salinity and bottom temperature over the North America East Coast by using a dynamically downscaled model forecast system (MOM6-NWA12) and compared with the parent SPEAR model forecasts. Detailed discussions about sources of improved prediction skill from downscaling are included for the Northeast U.S. region, as well as discussions on the effect of long-term warming trends over this area. Besides, this paper also contributes a useful discussion on the ensemble size for reasonable prediction skill when predicting SST. This is a very important work with high quality contributing to the research field, and I only have a few minor comments on this work:

We thank the reviewer for taking the time to read the manuscript and provide a helpful review. Below we have pasted the reviewer's comments in bold, followed by our replies.

1. Some description about seasons in the Result section are confusing, not sure if authors are talking about initialization seasons or forecast seasons. For example, on lines 237-239, "the downscaled model has skill greater than persistence and SPEAR across a wide range of times, except in the winter…". It is not clear if "winter" here refers to the initialization month of December or forecast months in winter.

We see how this could be confusing. For lines 237-239, we have reworded it to "except for forecasts verifying in December". We have also clarified in this and a few of the following paragraphs whether the months or seasons we mention are the initial or verification months.

2. Lines 267-268, the Southeast U.S. LME, as shown in Figure 1, is not narrow compared to most other LMEs. I also question on its dominance by the Gulf Stream, as Gulf Stream is usually referred to the western boundary current north of Cape Hatteras (so north of the Southeast U.S. LME).

We reworded this to say that the Southeast US LME "has a narrow shelf and is dominated by the western boundary current".

3. Description of the forecast-observation mean bias (for Figs 5-7) could be more focused on those forecasts that have significant forecast-observation correlation coefficient (Figs 2-4).

In the revised manuscript we now mention that the SST biases in the SS and NEUS LMEs are improved in the downscaled model for forecasts verifying in autumn and early winter when the downscaled forecasts have the most skill.

4. Lines 278-284: authors could just write out the season name, instead of "first season", "last season", and "seasons 0 and 2".

In the revised version we have replaced "seasons 0 and 2" with "the first and third seasons".

5. Lines 294-295:
(1) “remaining three regions” -> “remaining four regions”?

Yes, we have fixed this to now read “remaining four regions”.

(2) “aside from the increased spread in the Southeast U.S.” not sure why it is “increased” when comparing with SS and NEUS based on Figure 9, please consider rephrasing this sentence.

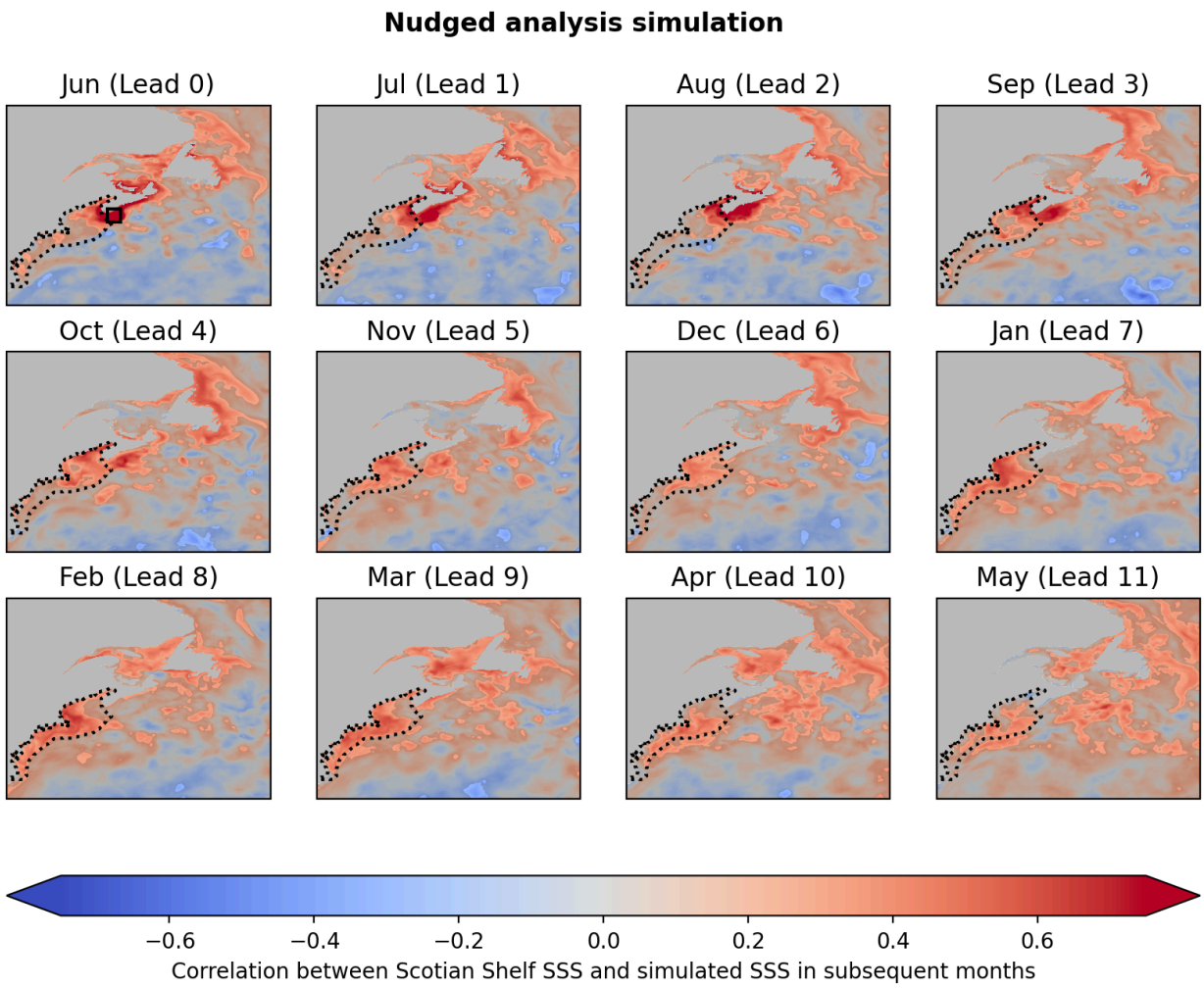
We agree that this was worded confusingly. It now reads “Differences between the two models are smaller in the remaining four regions, aside from NWA12 having higher spread than SPEAR in the Southeast U.S. and lower spread and RMSE than SPEAR in the Floridian region”.

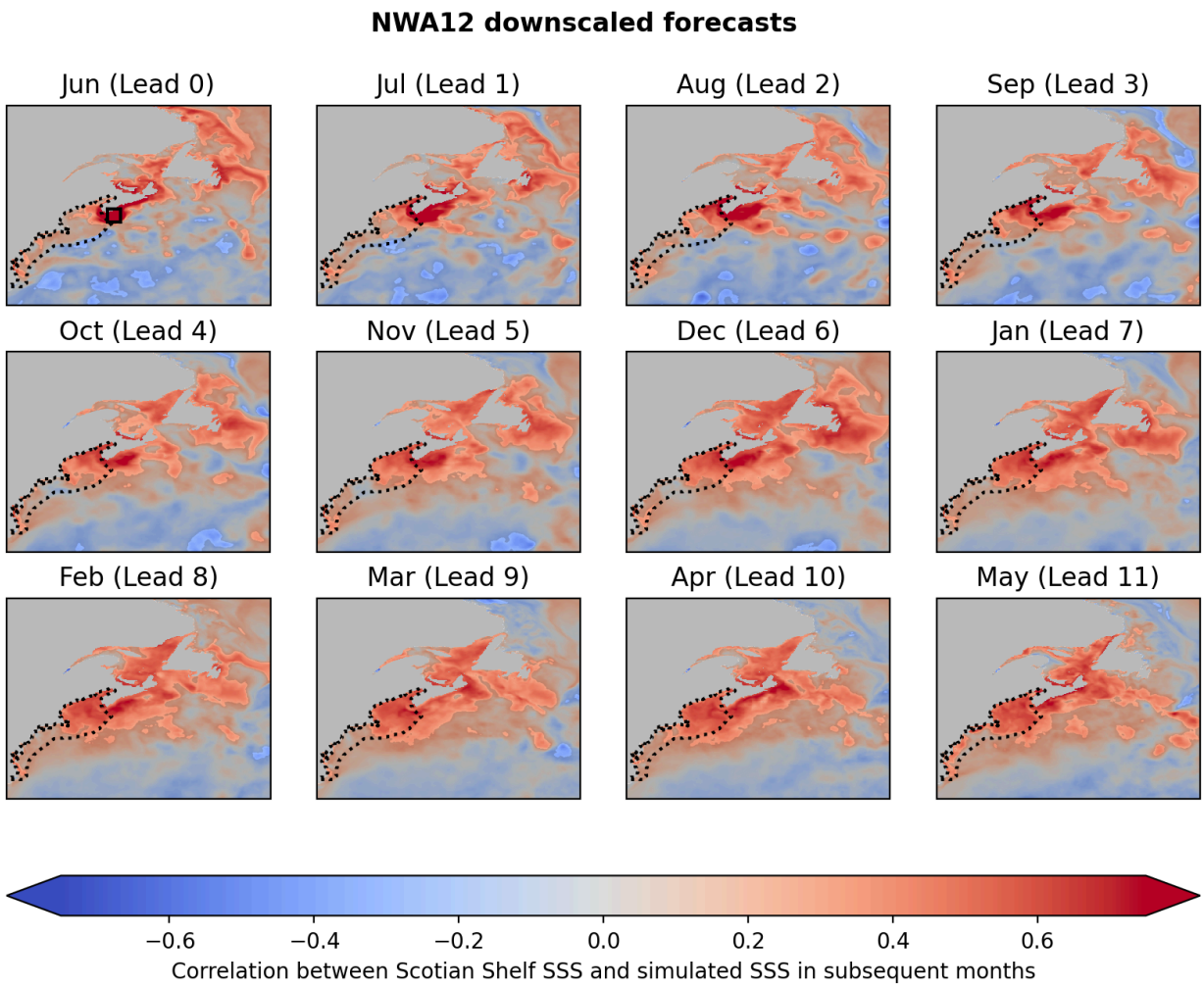
6. Line 340: “mid-Atlantic Bight” -> MAB

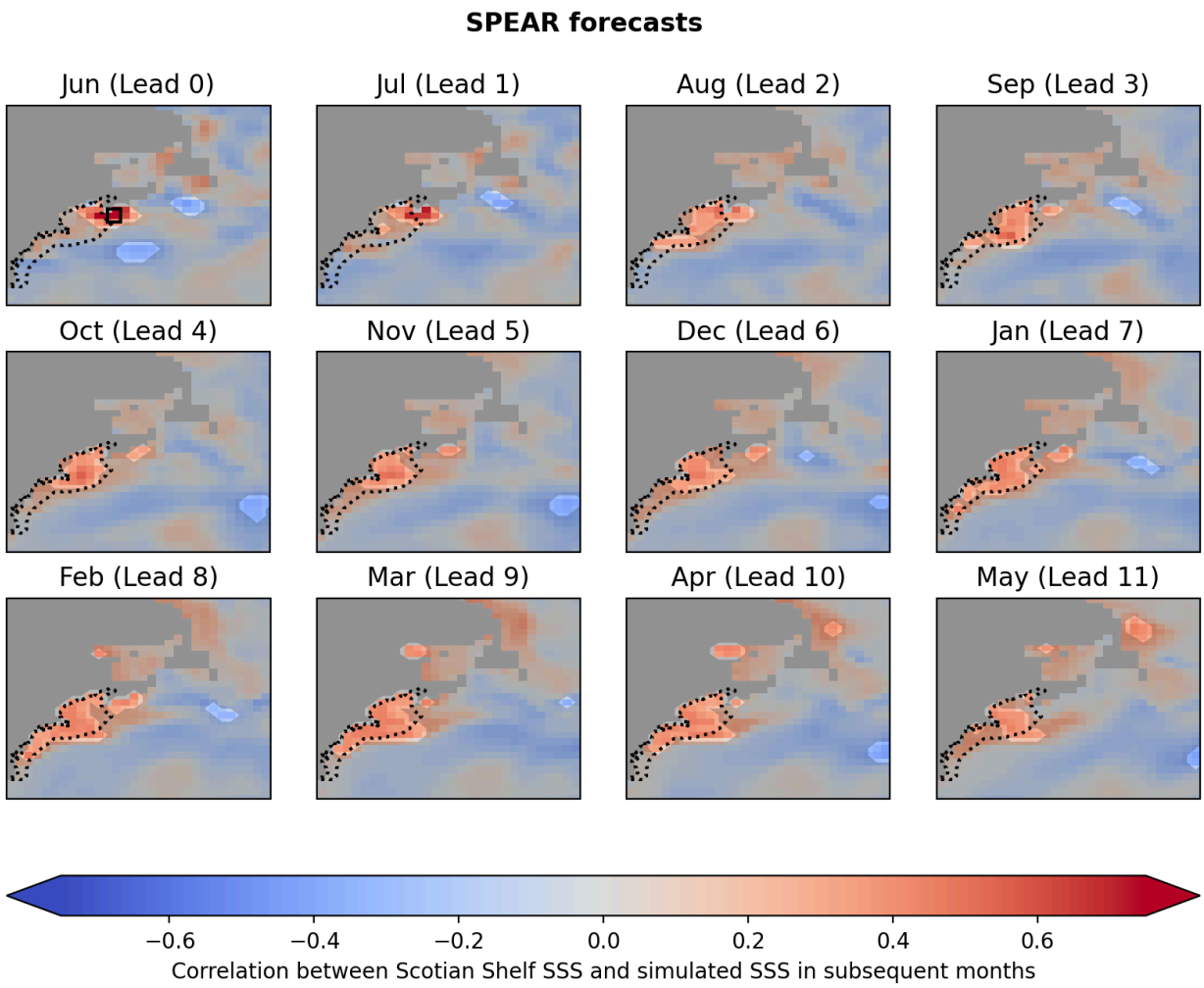
We appreciate this suggestion and have also replaced a few other instances of “Mid-Atlantic Bight” with MAB.

7. Figure 11: Please indicate correlation significance in the figure for each panel. Figure 11 shows the correlation in the GOM is minimum at Lead 6 but increases at Lead 7. Could you please explain it? Please consider adding the location of the Scotian Shelf box in Figure 1.
-and-
8. Figure 12-13: please consider adding correlation significance in each correlation map.

We have a semi transparent gray shading to the regions where the correlation is not significant at alpha = 0.1 in Figures 11, 12, and 13. A copy of these figures is included below. We have also added an outline of the Scotian Shelf box to the first panel of each figure.

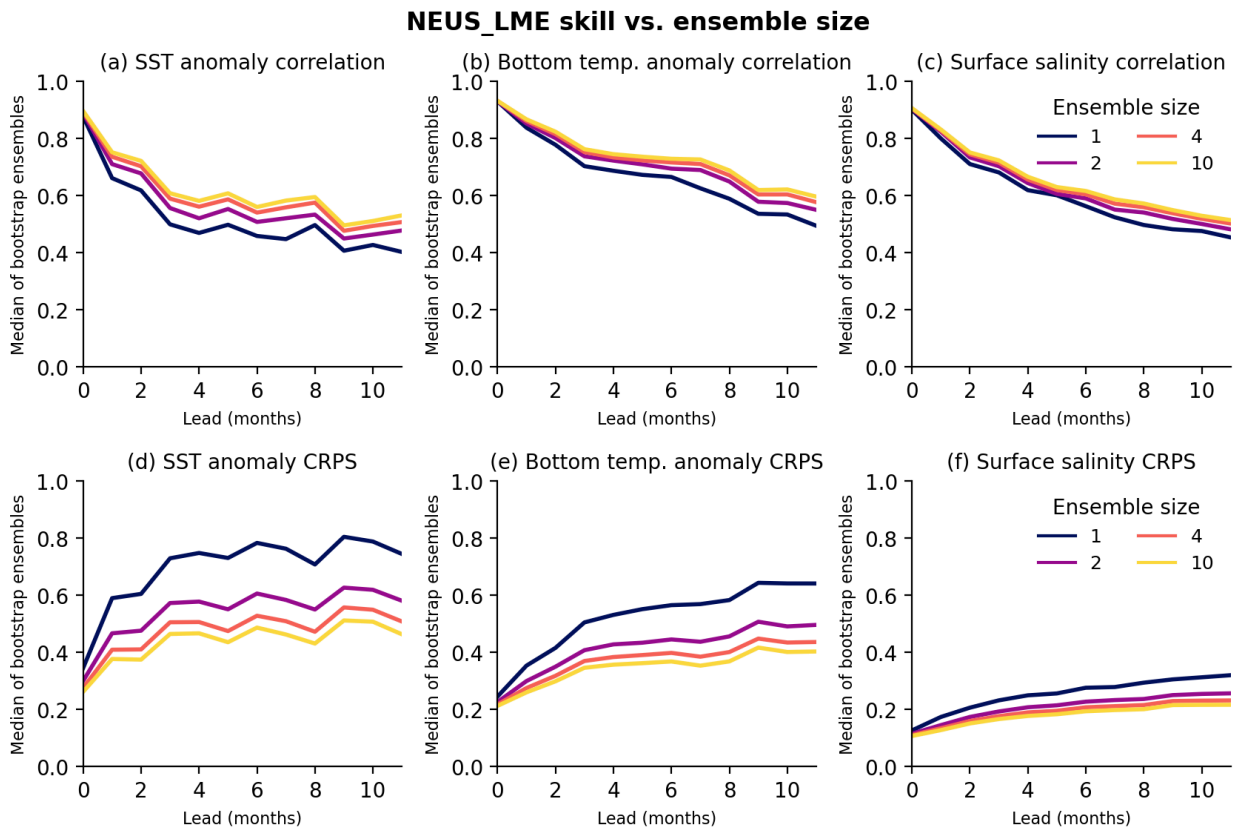

Nudged analysis simulation
Jun (Lead 0)
Jul (Lead 1)
Aug (Lead 2)
Sep (Lead 3)
Oct (Lead 4)
Nov (Lead 5)
Dec (Lead 6)
Jan (Lead 7)
Feb (Lead 8)
Mar (Lead 9)
Apr (Lead 10)
May (Lead 11)
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
Correlation between Scotian Shelf SSS and simulated SSS in subsequent months


NWA12 downscaled forecasts
Jun (Lead 0)
Jul (Lead 1)
Aug (Lead 2)
Sep (Lead 3)
Oct (Lead 4)
Nov (Lead 5)
Dec (Lead 6)
Jan (Lead 7)
Feb (Lead 8)
Mar (Lead 9)
Apr (Lead 10)
May (Lead 11)
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
Correlation between Scotian Shelf SSS and simulated SSS in subsequent months



9. Figure 20: Do predictions of bottom temperature also require approximately 4 ensemble members to provide a reasonable compromise between computational costs and prediction skill?

Yes, bottom temperature and surface salinity have similar patterns of skill vs ensemble size. To show this, we have revised figure 20 to include panels for bottom temperature and surface salinity (figure also included below). We have also revised the text to mention that the effect of ensemble size is similar for all three variables.


NEUS_LME skill vs. ensemble size
(a) SST anomaly correlation
(b) Bottom temp. anomaly correlation
(c) Surface salinity correlation
(d) SST anomaly CRPS
(e) Bottom temp. anomaly CRPS
(f) Surface salinity CRPS
Median of bootstrap ensembles
Lead (months)
Ensemble size
1
2
4
10