Reviewer 2

This study examines the climate system response to three different emissions scenarios in one climate model. One scenario (RCP4.5) reaches 2.0C of global temperature rise above pre-industrial levels at 2100, while the other two scenarios (stabilization and overshoot) aim to reach 1.5C warming via different emission pathways by 2100. The study documents the response of global mean temperature, sea level, AMOC, and sea ice.

The manuscript addresses an important topic – the effects of different pathways to 1.5C or 2.0C warming on the climate system. It well written, well organized and the figures are clear. I believe it offers several important new messages, but needs more evidence to back these up properly. I recommend it be accepted subject to moderate revisions (no option for this in the list of recommendations).

We thank the reviewer for a supportive, insightful and thorough review, which helped us to improve the manuscript. We respond to each point in blue text interwoven with the reviewer comments below. All major additions to the manuscript are also highlighted in blue text.

Comments:

1. The manuscript currently documents the responses of a range of variables to do with "Climate, ocean circulation, and sea level changes" (AMOC, sea ice, surface temperature sea level, mixed layer heat budget for the NAtl), which is fine, but it ends up feeling a little "light" or qualitative. Many of the statements in the manuscript could easily be backed up by using the model output. It seems that sea level rise is one of the main points the authors are emphasizing? Or perhaps the North Atlantic warming hole? If so, the manuscript would benefit from some reworking to solidify these aspects. For example:

We have added additional quantifications to the text and have reworked aspects of the text. Details follow below in response to specific suggestions.

– Pg 8, paragraph starting on L5: I'm not sure I fully understand the point being made about the North Atlantic warming hole. The warming hole is "stronger" (i.e., colder) in the stabilization and overshoot scenarios than in RCP4.5. However, a) the heat budget perturbation is similar for all three scenarios while b) the AMOC weakens more for the overshoot and RCP4.5 than for the stabilization. Furthermore, the authors speculate that a "slowdown in advective heat supply may suggest a role for the horizontal gyre circulations" (should be for all three scenarios since the heat budget changes are similar for all three, but not entirely clear to me)? In the end, I'm still not sure why the warming hole is stronger in the stabilization and overshoot scenarios, though I could speculate. All the relevant pieces of information seem to be here, they just need to be linked and interpreted together. Perhaps some zoomed in figures of the North Atlantic would help, including showing the actual changes in the gyres (Sverdrup transport) or surface heat fluxes if these are indeed important.

The warming hole is not substantially different between the different scenarios, as we try to emphasize more clearly in the revised text. Figure 3 is meant to show that the temperature tendency is the result of a very small residual of large, compensating physical processes: namely, the advective supply of heat and its loss through air-sea exchange, with a subdominant contribution from mixing of heat into the region. As the circulation changes, both the advective supply of heat and the air-sea exchange of heat is reduced by more than 75% in both scenarios, while the mixing supply of heat increases slightly. The ultimate rate of cooling is determined by the balance of these processes, with the cooling due to the slowdown in advective heat supply "winning" by a very small margin. The balance of these processes is quite similar among the models and we did not mean to emphasize the slightly cooler temperatures. Rather, our intention was to find a causal explanation for the cool region, since the cause of subpolar temperature anomalies is a source of much debate [*Clement et al.*, 2015; *Bellomo et al.*, 2017; *Zhang*, 2017; *Caesar et al.*, 2018]. We agree that the breakdown of the heat transport into horizontal gyre versus overturning circulation is another interesting direction, but would argue that this is a separate line of inquiry, that has been recently explored [*Piecuch et al.*, 2017]. The decomposition here must be done in isopycnal space since isopycnals slope both zonally and meridionally in the subpolar North Atlantic. So, the horizontal transport can cross isopycnals and be part of the overturning circulation, as defined in density space.

- Pg 7, L13-14: Could the authors show the response of the total ocean heat content and TOA radiative imbalance in the three scenarios (since these determine the steric sea level rise)?
Figure 1 of this response document (below) gives the total ocean heat content change (in joules) over the 21st century, which is essentially the same as the time integral of the TOA radiative imbalance, given the small heat capacity of the atmosphere and the lack of interactive ice sheets in the model. These curves show almost precisely the same pattern as the global steric sea level rise, since the lion's share of steric sea level rise is due to the thermosteric effect. Even though the equation of state is non-linear, there is nothing notable to distinguish this time series from the steric sea level (Figure 1e). Therefore, we have not added this figure, but note the similarity in the time evolution in the text (lines 22-25 on page 7).

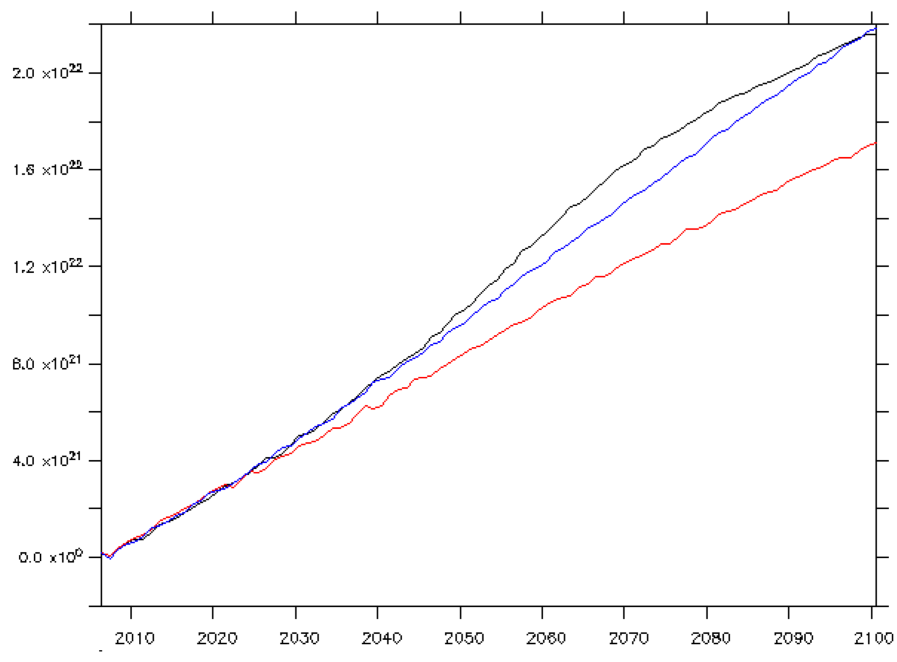

Figure 1: Global ocean heat content anomaly (relative to 2006) for the ensemble mean of Stabilization (red), RCP4.5 (blue), and Overshoot (black) scenarios.

– Fig. 2 and 4 and the global mean temperature curves suggest some interesting differences in

ocean heat uptake between the three scenarios. It would be nice to see what is happening, e.g., some maps of ocean heat content, to show where the ocean ends up sequestering heat.
We now provide heat uptake maps in the Supplementary information, with a brief discussion on page 7: "As has been noted in previous studies with this model, the heat uptake per area in the Atlantic outpaces the other basins, with this pattern arising in part due to the overturning circulation and its slowdown [e.g. *Winton et al.*, 2013]. Notably, the Atlantic heat uptake in RCP4.5 and Overshoot are slightly faster than the uptake in Stabilization, even after removing the global mean uptake rate."

– What are the contributions of the global steric term and local steric term to the differences in sea level rise between the three scenarios. Figure 6 suggests that both contribute. This could be quantified for the various locations.
Figure 6 shows the total local sea level rise (solid line), which is the sum of the local steric effect, the global steric effect, and the local change in ocean mass (due to surface fluxes and column-integrated mass convergence). The dashed lines show just the sum of the two local terms, without the global steric effect. We believe this decomposition gives a good indication of the changing ocean dynamics that leads to local sea level differences. Further decomposition of the local term into components might not be as revealing as one would hope, since the local steric effect will be influenced by dynamical shifts that transport water of different temperatures around the ocean. Further, it is not trivial to meaningfully separate the various terms in the sea level equation (see Griffies and Greatbach [2012] for a thorough exploration of sea level in ocean models). Specifically, we could further separate the local term into local steric effect and the local change in mass, but then we would have to correct for a spurious mass source that arises in Boussinesq models (see Appendix D of Griffies and Greatbach [2012]). For these reasons, we continue to rely on Figure 6 to explain the sea level differences among the scenarios.

2. It would be useful to show some measure of how strong the signals in the 2096-2100 averages are compared to the ensemble spread (even though these are small ensembles), and internal variability (from the time series in Figure 1, it is clear there is quite a lot of interannual to decadal variability in the responses).
Thank you for this suggestion. We have added the ensemble standard deviation for the final 5 years (2096-2100) to Figure 6, which makes our point more clearly that the stabilization and overshoot pathways can create large differences in some regions (Washington, Boston, St. Johns) and almost no difference in others (Charleston, Miami, Brest). In the latter three cities, the ensemble standard deviations overlap, an indication that internal variability is larger than the scenario-driven difference in sea level at these locations.

3. To provide some context for this single-model study, it would be useful to have some idea of how GFDL-ESM2M performs relative to other CMIP models in the RCP4.5 scenario. This is done on pg 10 for the ACC, but it would be nice to see similar discussion for variables/features that are the main focus of the manuscript.
We have added brief discussions of the following points, aimed at contextualizing GFDL-ESM2M relative to other CMIP models:

- Bottom of page 6: We note that the transient climate response in GFDL ESM2M is at the lower end of the CMIP5 model suite.

- Middle of page 7: Model differences in surface pH changes are small [*Frölicher et al.*, 2016], so we would expect that the surface pH evolution to be robust across models.
- Top of page 8: The decrease in AMOC is relatively large in comparison with other models, because mean state is also at the higher end.

Other comments:
1. Figure 1: the text labels are quite small, especially in panel b, and the thin lines for individual ensemble members are too faint.
The text labels have been enlarged. The ensemble members in panel b are just very close together, and therefore hard to distinguish. They are clearer in other panels, in which the ensemble variability is greater.

2. Equation 3 seems to be missing a dz in the integral.
Thank you for catching this omission. We corrected it.

3. Pg 6, L12: arises -> arise
Fixed.

4. Pg 7, L10-11: "... the pathway to a given forcing does not factor prominently in the annual mean, global average surface air temperature change in response to that forcing." This statement seems a bit odd to me given that the stabilization and overshoot emissions pathways were constructed with an aim of achieving 1.5C warming at 2100. Maybe it's just that I have misunderstood the description of how the scenarios were designed, in which case the authors could try to clarify this instead (pg 3 last paragraph).
It is very difficult to construct scenarios in an AOGCM that match exactly the desired temperature in a given year. Thus, this statement was meant to highlight the success of the method used to construct the pathways, as well as make clear that the nonlinearities that arise under these scenarios do not cause major deviations from the expected global mean temperature pathway.

5. Figure 5: What is "vertically-averaged" sea ice concentration?
There are 5 vertical layers of sea ice. The concentration over all five layers sums to one if the grid cell is covered in sea ice for the whole year over all five vertical layers. We agree that "vertically-averaged" sea ice concentration is confusing, and therefore changed this graphic in the final version and its corresponding description. The revised figure shows the change in vertical sum of the concentrations (with a maximum value of 1) between the preindustrial 2096-2100 year average of the ensemble mean and the preindustrial control.

6. The implications of these results for ocean acidification is mentioned in the conclusions. This seems to be a very important point. Would it be worth including this in the results, with some analysis of the model output to back this up?
As in response to Reviewer 1: We added a panel to Figure 1 to show the pH changes over time in the surface layer, and a deeper layer (1573 m). We also added the following text to the manuscript to describe the changes:

The higher atmospheric carbon concentrations in the overshoot pathway relative to the stabilization pathway temporarily cause more ocean acidification (i.e. lower pH values) at the surface (Fig. 1f). At the point that atmospheric CO_2 in overshoot matches nearly stabilization (2100), the surface pH is still lower in overshoot, but is rapidly approaching the stabilization surface value. At depth (i.e. 1573m), both the emergence of greater acidification as well as the turn-around in pH in the overshoot scenario are delayed by about 20 years relative to the trends in atmospheric CO2, because transport and mixing between the surface and depth is relatively slow [Gebbie et al., 2012]. As a result, ocean pH is lower in the overshoot scenario than in the stabilization scenario at the end of the 21st century, as expected from the carbon budget (Fig. 1b). Based on the rate of change at 2100 in our simulations, we expect this difference in pH to persist for a few decades after the crossover point in atmospheric CO_2, as the ocean slowly releases the excess CO_2 to the atmosphere