Response to Reviewer 1

We thank the Reviewer for their thorough review of our work. Many of their suggestions for analyses greatly strengthened the paper.

1. Major Comments
    1.1. Reviewer 1's first major comment was with regards to a lack of overall coherence to the paper due to its evaluation of two distinct model configurations, CESM2 LENS and CESM2 S2S. We have addressed this concern by expanding our discussion of the structure of the paper in the introduction to better explain how the two sections relate. We also added a clearer transition between the two sections in the Results, reiterating the connections between the two, and editing the text as a whole to reiterate how each piece supports the goal.
    1.2. The second major concern was with regards to the "physical reasoning behind the underrepresentation of sea ice loss during VRILEs". We have added a discussion and two corresponding figures regarding the mean atmospheric state in the five days before VRILEs occur. In short, VRILEs tend to be preceded by strong pressure gradients caused by a surface cyclone interacting with a preexisting high pressure system, such as the Beaufort High. These systems drive strong polarward winds that are thought to interact with the sea ice both directly and indirectly by facilitating poleward atmospheric heat transport, inducing vertical ocean mixing, and driving large ocean waves. Adding this discussion also helps address Reviewer 1's first major concern, lack of coherence.
        1.2.1. We also replaced Figure 6, which showed the results from the CESM2 S2S reforecast VRILE cases, with a separate figure for each case. These figures now include plots of the mean atmospheric state for the 5 days before the VRILE and the difference between NSIDC and CESM2 S2S sea ice concentration at the day the reforecast was initiated and on the day of the VRILE. This allows a more indepth discussion of the physical processes driving short timescale sea ice loss.
2. Specific/Minor Comments
    2.1. Line 4: It is unclear what the authors mean by "climate models underrepresent sea ice processes". Please be more specific about what processes this refers to and what "underrepresent" means here (does it refer to models not melting or growing sea ice fast enough?)
        2.1.1. Yes. Edited for clarity.
    2.2. Abstract: A brief overview of results from the evaluations of sea ice variability unrelated to VRILES is missing from the abstract and should be included (i.e., only VRILE analysis results appear in the abstract).
        2.2.1. Expanded our discussion of CESM2's interannual variability more generally.
    2.3. Lines 22-24: The timing of the first ice-free summer should not be used to judge the skill of a given climate model since an ice-free summer has not yet occurred

in reality. Projections of the timing of the first ice-free summer are produced from climate models and have uncertainty associated with them. Therefore, a model that has an earlier or later ice-free summer projection than another model (or consensus of models) is not necessarily deficient because we don't know when the first ice-free summer will actually occur. Thus, I recommend rewording or clarifying the second part of this sentence accordingly.

2.3.1. We removed the comment on ice-free summers to better emphasize the main point of older models not capturing observed climate trends.

2.4. Lines 31, 172-3, 188, 268: The citations on these lines should appear entirely within parentheses.

2.4.1. Corrected

2.5. Lines 77-79: If there are daily values from each individual ensemble member, shouldn't it be possible to average all the members together to obtain a daily ensemble mean?

2.5.1. Yes, this is in theory possible. However, constructing such a dataset exceeded our available data storage and memory limits. The sea ice data alone is about 18 TB.

2.6. Line 83: I had to read this sentence several times to fully grasp what is meant by the model year in CESM2 LENS not directly relating to the calendar year. I recommend rewording this part to say something like "Since the sea ice extent in CESM2 LENS diverges from the observed sea ice extent, we compare the observed sea ice extent from NSIDC between 1980 and 2020 to earlier periods in the CESM2 forecast when the modeled sea ice extent is more comparable to the observations during this period." I also felt this sentence should be moved later in this paragraph, to just before the description of the specific periods in the model being compared to NSIDC (~Line 90).

2.6.1. Clarification added and moved to later in the paragraph.

2.7. Line 119: I would replace "dynamical cores" here with "atmospheric model configurations" because the differences described here at not actually differences in dynamical cores (dynamical core refers to the numerical solver of the Navier-Stokes equations, not the physics packages or model grid configuration).

2.7.1. Corrected.

2.8. Lines 122-123: Do we know why the deeper atmosphere and interactive chemistry in WACCM6 results in better ice forecasts?

2.8.1. I suspect that this is related to the representation of clouds and their effect on latent heating and surface radiative forcing, but to my knowledge this is an open area of research and beyond the scope of this study.

2.9. Line 127: It's unclear what is meant by "does not provide enough time to evaluate the VRILE that occurred on 3 July". Please provide a clearer explanation of what is meant here.

2.9.1. Clarification added. We simply meant that the atmospheric and ocean processes that lead to VRILEs need more than one day to act on the ice to cause a VRILE, making evaluating the 2 July reforecast unenlightening to our overall analysis.

2.10. Line 136: Please add the phrase “with differences” before “becoming” on this line.

2.10.1. Added

2.11. Section 2: A description of the basic aspects of CESM2 LENS (akin to the description of CESM2 S2S on Lines 112-118) is missing from this section. At minimum, the text should describe when the CESM2 LENS simulation is initialized (I assume 1850?) and what anthropogenic forcing scenario is used.

2.11.1. Added

2.12. Line 150 (and elsewhere): Please replace all instances of “inner-quartile range” with the corrected name: “interquartile range”

2.12.1. Corrected

2.13. Lines 166-168/Figure 4: Could the reduced variability of SIC within the marginal ice zone regions in CESM2 compared to NSIDC simply be due to these regions being ice-free in CESM2? This seems plausible given the large negative ice extent bias in CESM2, which would likely result in more ice-free areas at these lower latitudes. I recommend checking what the actual SIC values are in these regions in CESM2. It may also help to also include the 0.15 SIC contour from CESM2 on the panels in Fig. 4 so the reader has a better sense for where the ice edge is in the simulation (not just in the observations).

2.13.1. We have played around with presenting the differences with both ice edges and 0.15 SIC contours or showing the SIC standard deviations of each dataset individually, and it isn’t qualitatively different than the version in the paper, but it does get visually muddy when we start adding additional contours.

2.14. Line 170: Change “have positive values” to “there are positive values” to make this sentence grammatically correct.

2.14.1. Corrected

2.15. Fig. 5: The caption says CESM2 VRILE sizes are indicated by blues, but when I zoom in on the smaller bars that presumably correspond to CESM2 in this figure, I see black, purple, red, orange and yellow segments, but no blue segments. I recommend either adjusting the color guidance in the caption to better correspond to what is shown in the figure, or changing the figure colors to truly have shades of blue for CESM2.

2.15.1. Changed to “dark colors”

2.16. Fig. 6: This figure should include another row showing the 5-day change in SIC in NSIDC to illustrate what the observed ice loss associated with the VRILE was.

2.16.1. This figure has been replaced by four separate figures, one for each of the VRILE case studies. These new figures include the 5-day change in SIC in NSIDC and the 5-day mean of mslp and surface winds.

2.17. Lines 200-201: It’s not clear to me that the ice biases in a long running CESM2 LENS forecast should have any relationship to biases in short-term forecasts from CESM2 S2S, given the significant differences between the two model configurations (free running climate simulation vs. initialized short forecast). This gets back to my major comment 1 above.

2.17.1. We removed the line in this discussion that mentions CESM2 LENS's overall sea ice bias and added significantly more discussion of what insight we hoped to gain into CESM2's sea ice model's (CICEv5) physics from analyzing these reforecasted VRILEs to Section 2. More detail is in our response to Reviewer 1's major comments.

2.18. Lines 231-234: Have the authors checked if the areas corresponding to the VRILE are actually open water (ice-free) in CESM at the beginning of the reforecast? The negative differences in Fig. 6d, along with what were known to be low sea ice concentrations in the Beaufort Sea at this time, suggests CESM may not even have sea ice here, which would change the interpretation of the results for this case.

2.18.1. Yes. Plots showing the difference between the initial sea ice state in the reforecast and NSIDC have been added for clarity.

2.19. Line 245: Change "unseasonal" to "unseasonable".

2.19.1. Corrected