The authors thank the reviewers for their feedback, we feel it has improved the quality of the manuscript. We have provided inline responses to each reviewer comment, as well as a reference to where such changes can be found in the revised text.

RC1:

While I have no concerns about the methodology as such, I think that the exploration of the physical causes of bimodality remains a bit vague and more could be done to pinpoint the actual causes of the bimodality. Some suggestions in that regard are given below. Overall the paper is well written and figures are readable. However, the language is sometimes unclear and the method could be better explained. Also, there is a lot of jargon, which may be understood by specialists in subseasonal-to-seasonal prediction, but not the general readership of WCD. Since the study is certainly of interest for a broader readership in the climate dynamics community, I suggest the authors reduce the use of specific jargon and try to better explain certain concepts.

We have made substantial efforts to reduce the jargon and have tried to be more explicit in defining the jargon that is still present. Furthermore, we more concisely and explicitly present specific definitions that we introduce for the clustering methodology/ tracking algorithm (Section 2).

We have also added several aspects of our analyses to better identify the physical causes of bimodality, though see our response to the first comment of reviewer 2 where we discuss the scope of our efforts.

General comments: 1. What do you mean by the term "atmospheric events"? This term is used throughout the paper, but it remains unclear what exactly you mean by it. Do you think of synoptic events that might cause some ensemble members to follow a very different trajectory in phase space than most others, hence causing bimodality, or rather slower processes such as the continuous interaction of the atmosphere with the boundary conditions? Reading further ahead, I assume it is the latter. I think it is necessary that the term "atmospheric events" is precisely defined already in the introduction.

In the revised text we remove the term 'atmospheric events'. This was originally meant to be a vague term to represent atmospheric phenomena that may lead to bimodal events without necessarily knowing what form these will take. However, we have rewritten the manuscript such that we introduce the term 'bimodal events' earlier and make it clear that we are trying to understand if (and why) they exist (Introduction).

2. The geopotential height patterns are reminiscent of Rossby wave breaking events in this region causing cold surges in Brazil, see Sprenger et al. (2013). Hence, I was wondering whether the bimodality is related to the presence and absence of Rossby wave breaking? Considering Rossby wave breaking may give a more direct physical linkage to the causes of the bimodality. Sprenger, M., Martius, O. and Arnold, J. (2013), Cold surge episodes over southeastern Brazil – a potential vorticity perspective. Int. J. Climatol, 33: 2758-2767. https://doi.org/10.1002/joc.3618

This is a very compelling argument, however, beyond a much more extended analysis, which is beyond the scope of this manuscript, it is difficult to confirm this hypothesis. In the revised manuscript, we have mentioned this as a possibility and include the reference (Section 3).

3. The relationship of bimodality in the Southern Ocean region to sea ice is interesting and appears plausible through the different heat capacities of water and ice, resulting in a damping effect on temperature variability in the former case. The explanation of why differences in sea ice state occur, however, remains overly vague. Generally, sea ice in this region reacts strongly to persistent wind anomalies, which can push the ice edge far away from its climatological position. Hence, It appears to me that it is not single synoptic events that cause the bimodality, but rather the accumulated effect of anomalous winds (for example through several cyclones passing through the Amundsen and Bellingshausen Seas) over one or two weeks, hence linking these events to longer timescales. Could the authors look into circulation anomalies in the preceding weeks?

This is a good point and the fact that the mean persistence of these events is only a couple of days to begin with, may indicate that dynamic (or thermodynamic) influences from the differing sea ice fields is not the main cause. We address this by looking at the sea ice fields of the two modes of several case studies to examine if they're identical or not. We find that while the sea ice fields do differ between the two modes (and bimodality develops along this difference), bimodality also exists away from this sea ice edge. It seems bimodality may in part be caused simply by the direction of low level winds over the sea-ice-to-ocean transition. We provide a more extended discussion of these processes in Section 3 of the revised manuscript.

4. Could explosive cyclogenesis play a role for the North Atlantic events? This region is well known for the frequent occurrence of bomb cyclones and Fig. 14b suggests the presence of a deep low in mode 2 which is absent in mode 1. Exploring this might give you a more direct physical linkage between the SST gradient associated with the Gulf Stream, which is known to play a crucial role for the rapid intensification of cyclones, and bimodality.

Similar to the Andes region, while this is a possibility it is difficult to state this for certain without a much more extended analysis. In the case studies that we examined in detail, explosive cyclogenesis does not seem to be at least the main driving factor, however, these events may certainly lead to some of the developed bimodality. We include this process as a possibility in the revised manuscript (Section 3).

RC2:

Major comments: 1. The methodology used in this study is perhaps unnecessarily complicated and uses lots of arbitrary thresholds and other subjective elements. This undermines the robustness of the study. This is beyond something that is not elegant – it is cumbersome and hard to defend. I recommend that the Authors scrutinize their methodology and eliminate unnecessary conditions or base their choices on more objective conditions. Examples include: l.111, 116-117, 128, 136- 137, 154 (subjective choice of 3 regions), 160, 166, 179-185. Some of the arbitrary choices are necessitated by earlier subjective conditions etc. Several of these choices appear to be made so a desired outcome is achieved (e.g., l.160, 179- 185), undermining the generality of some of the findings. For further details, see detailed comments below.

We thank the reviewer for their critical and constructive comments. We have substantially rewritten the manuscript in an effort to clarify the methodology, the scope of our study, and the main results.

We agree that our methods suffer from systematic selection biases—specifically towards larger events. However, the point of this methodology is not to accurately capture all bimodal events, but rather to demonstrate that they exist in the first place and to begin to connect them to several large-scale phenomena as a basis for future study. To carry out a systematic study of all bimodality present in these forecasts would be far too large a task for a single paper. We have reconstructed some of the text to try to more explicitly state this goal (Introduction and Section 2). A follow up study would be necessary to fully characterize, for example, if or what mesoscale events may lead to bimodality. Furthermore, we have removed some of the filtering that we perform on the clusters to streamline the methodology (with minimal change to results).

2. The study presents results suggesting bimodality in the distribution of ensemble members. This condition, however, is never established objectively. I recommend that the Authors consider developing an algorithm to test whether the results they present are statistically (in)distinguishable from surrogate data generated using a null hypothesis that ensemble forecasts are normally distributed. Are results over the entire sample (i.e., over selectively chosen data) indistinguishable from randomly generated samples from a normal distribution? Without such a test, the results are not convincing and would be appropriate to present only as a short note.

We have already performed such an analysis in the initial study of bimodality (see Bertossa et al. 2021). Bertossa, Cameron, et al. "Bimodality in ensemble forecasts of 2 m temperature: identification." *Weather and Climate Dynamics* 2.4 (2021): 1209-1224.

3. In my view, no major conclusion appears to emerge from Section 3. The discussion is mostly descriptive, with no insight into the dynamics leading to the emergence of bimodality. As the results in Section 3 may be limited in significance, I find the number of figures (10) excessive. I recommend that the Authors significantly reduce the number of figures and resort to using short verbal statements to describe less important results. Also, as some of the results for the three regions studied are similar, results for the three regions could be compared and discussed together, shortening the text.

We feel that the combination of the summarizing statistics and the case studies are necessary to explore the properties of events that lead to bimodality. This is a completely unexplored topic and so we intend to present a broad overview of each region. However, the additional analysis of the sea ice field, for example, has allowed us to refine our insights and shorten the text. Furthermore, we have shortened discussion within each region and added a general 'discussion' section in which similarities between regions are discussed, removing redundancies.

4. The Discussion in Section 4 is interesting but appears somewhat preliminary. A more logical organization of the material and a shorter presentation would improve this section.

We agree that with the present findings, much of this is speculative. We have removed this section and only place a shortened version in the conclusions section.

5. Apart from its last paragraph, material in the Conclusions section also appears to be disorganized and premature. As the manuscript mostly presents hypotheses without objective validation, unless more evidence is added, some of the statements (e.g., l.420-421 should be tempered accordingly. Depending on how the Authors respond to the reviews, this section also needs a major revision.

We have revised the conclusion section significantly and feel that with the new analysis, our results are significant and better supported by the analysis presented. However, we also reemphasize that the objective of our manuscript is more to identify these bimodal events rather than exhaustively characterize their behavior.

6. The quality of the presentation is mixed. There are some parts in the Introduction that are well written. The bulk of the manuscript, however, needs significant improvements before possible publication. See detailed comments below for some examples. Some suggestions: a) Clearly define all terms used in the study. Examples include: size, occupancy, points, forecast in the description of the methodology. Some terms may be poorly named, using commonly used words with a different meaning. In such a case, clear definitions are especially important. b) Figures should be defined in their captions but the discussion of the results should appear in the text. c) Replace hypothetical examples used in some figures (e.g., Fig 2) with examples from the actual data.

In the revised text, we more concisely and explicitly define terms in the methodology section and limit defining figures to just their captions rather than in-text. The authors feel that c) is important since it is intended to give a clear example of the clustering process and definitions. The actual data is less idealized and harder to present in such a simple manner.