Response to Reviewers - 2nd revision

Dear Editor,

We would like to thank all reviewers and the editor for their reviews of our manuscript and their insightful comments. Please find our detailed responses to the reviewers' comments and suggestions below. The changes have been included into the manuscript (indicated in **bold**). All line numbers refer to the new (annotated) version of the manuscript.

Sincerely,

The authors

Reviewer 1

Minor comments:

Figure 3 There are currently two black boxes which is confusing throughout the text. I suggest distinguishing between them, for example, using another color or a dashed line for one of them.
We thank the reviewer for their comments. We changed the color of the European box to red. The North Atlantic box is marked in black.

L198 As you now have boxes plotted in all of the subplots it's enough to say 'Figure 3'. However, it is good to specify here which box you're referring to (see my comment above).
We added multiple specification of which box is refereed to.

Figure 8 caption: although you mention the coordinates, it's better to specify here which box you're referring to once you've distinguished the boxes (see my comment above).
Corrected.

Figure 10 Please mention in the caption that this is a 28d average
We added this information to the caption.

L342 Although you mention using Z100 in the Data and Methods, I believe that this is the first time you mention Z'100 (Z100 anomalies) so it makes sense to specify it here.
We added an explanation for Z'100 acronym in section 3.7 and removed any unused acronym from the Data and Methods section.

L358 please fill in the Figure number instead of '??'
We corrected the missing figure reference.

L370 typo 'to be associated'
Corrected.

L373 probably 'to SSW events'
Corrected.

L452 I think that you forgot to add the link to Github in the Code availability section.
Thank you for this comment. The previous version inadvertently omitted the link to the Github site. We have updated the "Code availability" section with the correct link.

Reviewer 2

Overall Opinion:

This is my second review of the manuscript. The prior version of the manuscript lacked two key elements that I thought made it unsuitable for publication: (1) The study focused on only one model, whereas the database has several other models with hindcasts that could be explored for the same phenomenon; and (2) the paper lacked any dynamical insight into why the results were the way they were. The authors responded that doing the analysis for more than one model would be too laborious and that the ECMWF model was a well-trusted subseasonal model. I kind of agree with this thought, though I think the paper would be strengthened with more than one model analyzed for the study. But, I won't hold up publication of the study based just on that issue. For the second point, the authors expanded their analysis to look at the nature of downward propagation and the tropospheric circulation after a strong or weak vortex event. Specifically, the authors contrast between ensemble members which correctly predict the anomaly of the North Atlantic cyclone frequency following a strong/weak polar vortex event and those that do not. However, I don't think the methodology used actually addresses the dynamical interpretation of the results that the authors intend.

So, taken together, I would consider the paper ready for publication but only after major revisions, particularly to the dynamical interpretation portion of the paper.

Major comments:

1. Sampling Issues and Figure 4. The authors continue to note that sample size is a concern for many of their results, and I agree with this point. However, one area this is not addressed enough is in Figure 4. In particular, the sample size between reanalysis and reforests for the SSW and strong vortex events is about a factor of 10 different, which makes comparison of the probability distribution functions (PDFs) very difficult (in fact, I question the reproducibility and representation of a distribution of a variable with only 14 samples). For example, the authors could repeatedly

sample 14 random cases (with replacement) from the reforecasts to make a comparative PDF with reanalysis. I think in its current form, it is hard to argue about statistically significant differences in these distributions, whether comparing the cases or comparing reanalysis to reforecasts. The authors may want to pursue alternate strategies to strengthen this argument in the paper.

We thank the reviewer for this comment. Figure 4 shows distributions of cyclone frequency anomaly in the central North Atlantic for reanalysis and re-forecasts for SSW and strong vortex events. A statistical significance test, a two-sample Kolmogorov-Smirnov test, has been performed between two fitted distributions. However, we are aware that it is a difficult task to do a comparison of the two distributions given their different sample size (14 for the reanalysis samples, and 140 for the probabilistic forecasts). Therefore, and given the current structure of the manuscript, Figure 4 has been removed from the manuscript. Instead of further statistical comparison between reanalysis and the reforecasts, in the revised manuscript we extended the discussion on the dynamical causes for successful or unsuccessful prediction of the canonical surface response after stratospheric events (section 3.7) using a relatively larger sample size (100 samples) provided by the reforecasts. One of the new additions includes a new figure (Figure 11) added to this subsection.

2. Dynamical Interpretation. I like the efforts that the authors made in trying to bring some dynamical insight into their model evaluation study of Atlantic storm track changes due to stratospheric polar vortex variability. However, I am not convinced that the analyses shown actually accomplish this effort. In particular, the authors use mean sea level pressure (MSLP) to represent the tropospheric circulation changes after "successful" and "unsuccessful" forecasts. However, storm tracks are defined in this study using MSLP. So, it is a bit circular to argue that differences in MSLP ("the tropospheric circulation") are the leading reason why there are changes in storm tracks (which are determined by MSLP). I like the use of lower tropospheric winds and even looking at the lower stratosphere (Z100). But, the authors should reconsider how they measure the tropospheric circulation and consider other variables for that other than MSLP.

We thank the reviewer for this comment. MSLP anomalies, shown originally in figure 11, represent surface circulation anomalies and provide different insights than the cyclone frequency. In particular, MSLP anomalies consist of both the high and the low pressure anomalies of the synoptic-scale flow, while cyclone frequency is primarily associated with low pressure systems (in the Northern Hemisphere). Thus, MSLP and cyclone frequency can represent different quantities.
Following the reviewer's suggestion, we changed the variable that we use to represent tropospheric circulation, and we now use zonal wind at 850 hPa (U'850). We added an analysis of U'850 in successful/unsuccessful forecasts (Figure 9). In addition, we

replaced MSLP with U'850 in an analysis of the time evolution of lower tropospheric anomalies (figure 10), thus using the same variable to represent tropospheric circulation anomalies in both figures.

Minor comments:

1. Lines 14-18. This sentence is long and confusing to understand. Please revise.

We revised this sentence in the Abstract, as follows:

"However, although the response of cyclone frequency following SSWs with a canonical surface impact is typically well-captured during weeks 1-4, less than 25% of the reforecasts manage to capture the response following SSWs with a 'non-canonical' impact. This suggests a possible overconfidence in the reforecasts with respect to reanalysis in predicting the canonical response after SSWs, although it only occurs in about two thirds of the events".

2. Lines 74-75. The acronym "ERA5" already contains the word "reanalysis" in it. So, it is redundant to say "ERA5 reanalysis."

Corrected. We replaced "ERA5 reanalysis" with "ERA5" throughout the manuscript.

3. Line 228. Please move the comma from after "ERA5" to after "reforecasts."

Corrected.

4. Lines 313. There is no need to define "MSLP" again here. Also, I think it is unnecessary to introduce another acronym into the paper for the 100 hPa geopotential height anomalies. Instead, the authors can just use the already-defined acronym for 100 hPa geopotential height anomalies (Z100). Can you just write "Z100 anomalies?"

We removed the repeated MSLP acronym in line 313, and in the Methods section (line 75) we removed other acronym (U10, U850 and Z100) that are not used throughout the revised manuscript. Instead, we added acronym to U'850 and Z'100 as these are used multiple times in section 3.7.

5. Lines 338-340. This sentence structure (with the parentheses) is no longer favored in journal articles for readability and understanding. Please rephrase as two sentences or in another way. Same comment for Lines 411-412.

We thank the reviewer for this clarification. We rephrased both of these sentences to improve their readability and understanding.

6. Line 353. It looks like there is a missing figure reference here ("??").

We corrected the missing figure reference.

7. Lines 362-363. I am unclear what "larger natural variability" means, particularly in reference to model reforecasts. What does "natural variability" in a simulated atmosphere mean?
Originally, the term natural variability in this context referred to variation in climate parameters of the simulated atmosphere caused by nonhuman forces. To clarify this meaning, we replaced the term "larger natural variability" by "larger variability".

8. Line 364. "...found to b associated with a..." —¿ "...found to be associated with a..."
Corrected.

9. Lines 445-448. The authors previously mentioned that they were going to have a Github site to make their data publicly accessible. This site is not listed here - please add the information for completion.
Thank you for this comment. The previous version inadvertently omitted the link to the Github site. We have updated the "Code availability" section with the correct link.