Manuscript Title: Stratospheric influence on the winter North Atlantic storm track in subseasonal reforecasts

Authors: H. Afargan-Gerstman, D. Büeler, C. Ole Wulff, M. Sprenger, D. I. V. Domeisen

Recommendation: Reject but resubmit

Summary

The study examines changes in the North Atlantic storm track following either strong or weak stratospheric polar vortex conditions in both reanalysis and the ECMWF subseasonal model reforecasts. Using reanalysis, the authors illustrate that the storm track changes often follow the expected canonical response to each polar vortex state; i.e., an anomalously poleward-enhanced storm track following strong events and an anomalously equatorward-enhanced storm track following weak stratospheric polar vortex events. However, this is not always the case, and the operational models struggle to capture these contrary cases. Additionally, there is no significant change in the distribution of cyclone frequency after weak vortex conditions in the reforecasts or in reanalysis when compared to winter climatology. The only significant shift in cyclone frequency is detected following strong vortex states. In terms of predictability from the ECMWF, there is somewhat better predictability for getting the sign of the cyclone frequency anomaly correct following the strong vortex cases versus the weak ones, even in the non-canonical cases. Of course, the hit rate decreases with lead time (as might be expected), but it could still be useful for the strong vortex cases for a forecasting application.

Overall Opinion

The study presents some interesting results on how the state of the stratospheric polar vortex can offer some improved forecast skill for cyclone frequency across the North Atlantic (both intensity and spatial area). I recognize the work that the authors put into many of these analyses and appreciate some of the results, including the difference in forecast skill of storm tracks between strong and weak vortex states. Unfortunately, I do not find the results particularly novel; moreover, they ack the robustness and completeness that could make this work much more useful for the forecasting community. As such, I find the manuscript slightly incomplete in its presentation - there is more that would need to be completed to bolster the main messages of the work. Hence, I regretfully recommend that this paper not be published in its current form but instead sent back to the authors for further analysis and resubmitted at a later time. More detailed reasons for my decision are below.

Reasons for Rejection

First, the study examines only one subseasonal model (ECMWF) and therefore lacks a generalized view of how other leading subseasonal prediction systems reproduce the stratosphere-North Atlantic storm track relationship. The fields from the reforecasts of the other models are readily accessible and possible to be analyzed and compared/contrasted. I am not necessarily advocating using *every* model, but I think adding a few more will be very useful and strengthen the message.

Next, I found aspects of the methodology confusing. In the methods section, the authors mention that they use the ERA-Interim reanalysis product for determining the state of the

polar vortex (strong vs weak) but then use ERA5 for their analyses. Determination of events in the ERA5 dataset is very straightforward. So, to be consistent, the authors should use one reanalysis only throughout their work. Next, since the ERA5 is used to initialize the ECMWF reforecasts, and since the two share aspects of their modeling components, independence in the comparisons is hard to justify. Again, this aspect limits the applicability of the results of this work to other forecast systems and reanalyses. Next, the authors also comment frequently on the limited sample size from ERA5 for their results. This facet factors into their significance testing and other conclusions (e.g., Fig. 5). If sample size is too small, why should we trust the results? I am not saying that the limited sample size is a game-ender for the paper (trust me - this is a constant issue with my own work!). But, to use this concern over and over again in the manuscript as a caveat raises questions as to whether or not the findings are just an artifact of a short sample size.

Finally, I was disappointed that the paper did not investigate any physical reasoning for why the storm tracks change as they do in reanalysis vs the reforecasts. The authors mention a few times that their results are "consistent with" previous studies, which is good. But, the reforecasts and their multiple ensemble members offer a fantastic opportunity for the authors to address the "why." They could explore changes in wave fluxes, baroclinicity, jet stream dynamics, etc. and provide an idea of why the stratosphere is influencing the storm tracks the way it is. I think this is a missed opportunity with this paper, thus making it contribution less novel than it otherwise could be.

For these reasons, I must unfortunately recommend the paper not be published in its current form. Instead, I would suggest the authors consider my suggestions above, expand their study, and then resubmit the paper with these improvements. I think this avenue would be the most beneficial for the authors and for the subseasonal community.

Other Comments

1. **The acronym "SPV."** The use of this acronym is confusing - it is normally used to mean "stratospheric polar vortex" in many other papers. Furthermore, I don't find that the acronym is necessary in the work - "strong vortex events" is clear enough and not overly long. I recommend that the authors reconsider using this acronym.

2. **Lines 154-155.** I don't understand this sentence. How is the "response in ERA5"… "stronger in ERA5?"

3. **Line 221.** Either the results are statistically significant or they are not - they cannot be "partly significant."

4. **Lines 281-283.** Is this a "result" or "finding" that is unique to this work? I think that finding has already been shown in many past works and is also based on the fundamentals of what the jet stream is.

5. **Lines 291-293.** How would the authors propose to increase the sample size to meet their objective of determining the robustness of the results? (See my comments above as well.)

6. **Figures 2 and 3.** How is significance tested exactly for the reforecasts? What is the null hypothesis?

7. **Figure 7.** "Successfully" is spelled incorrectly in the y-axis labels of panels (b) and (d). Also, it is unclear what an "increase of cyclone frequency anomaly" means. Is it that it is a positive anomaly, or that the anomaly actually gets more positive over some time?

8. **Code and data availability.** The authors have not provided a public-accessible repository where their code is available. Please set up a Github and place your code on there for transparency and accessibility.