Many thanks again to both referees for their comments. We are happy that referee 2 has accepted our manuscript for publication. Below, we address the comments that were given to us by referee 1. The referee's comments are in *italic gray*, our responses are in blue.
(Line numbers refer to the updated manuscript.)

This is my second review of the manuscript by Spaeth and Birner. The authors have addressed most of my minor comments; however, the comment regarding attribution analysis has not been properly addressed. In the revised version the authors introduce the FAR concept and apply it to attribute AO events to SSW (or SPV) events. I believe this is just another way to quantify the changed probability of AO events following SSWs, similar to relative probability increase (RPI). Furthermore, FAR analysis leads to confusing conclusions such as "approximately 50% of extremely negative AO states that follow SSWs may be attributed to the SSW". To make such statements one needs to block the SSW-AO link in these same SSW situations and see how many AOs would occur in such controlled experiments. This has not been done. Instead, what FAR indeed shows is the increased probability of AO following SSW, similarly to RPI. The difference between FAR (50% increase) and RPI (35-40% increase) is likely because the authors compare FAR to no-SSW situation while in RPI they compare SSW cases to all AO occurrence probability. Further, what the authors call "FAR among the population" looks like a decreased probability of AO during no SSW periods. In summary, I recommend the authors to modify the language and clarify that what they are doing are different ways to estimate increased probability of AO due to SSW occurrence. Without this I cannot recommend the manuscript for publication.

We appreciate the reviewer's concern about what can and cannot be inferred from the kind of attribution analysis we have done. Nevertheless, we are convinced that useful insights may be obtained by applying an FAR framework to SSW-AO coupling that go beyond of what simple probability increase measures provide.

First, let us summarize where we have adapted the manuscript in line with our arguments (see below):

- A paragraph was added at the end of the introduction that is aimed to clarify what is meant by "attributable to" (ll. 85ff): In particular, it is outlined that RPI and FAR may capture pathways other than the direct SSW-AO link.
- We are now more careful with terminology that would suggest causal relations (e.g., ll. 81-84, 393, 410ff, 416, 430, 441, 447, 479). Our first two research questions now read "By how much is the probability [...] increased following stratospheric polar vortex extremes?" (ll. 81ff; instead of: "By how much do stratospheric polar vortex extremes increase [...]"). In the abstract and in the answer of the third research question, we added a caveat concerning limited causal validity (ll. 20f, 518f). In the previous manuscript, we had summarized the results of FAR using the wording "may be attributed to". In the revised manuscript, we adjust the expression to "may be attributable to" in order to stress that attribution is only possible under certain assumptions. We chose not to use a weaker expression like "associated to" as this might suggest (the incorrect) interpretation to include fortuitous co-occurrences, which are indeed ruled out by our statistical estimates. Furthermore, the wording "may be attributable to" is terminologically consistent with FAR, therefore it implies on which analyses the conclusions are based on.
  To account for the restrictions that arise from a finite sample size and potential biases in the S2S models, we replaced verbs like "determine" and "quantify" by "estimate" (e.g., ll. 96, 233, 294, 365, 434).
- In section 3.5, we clarified that RPI and FAR are based solely on (conditional) probabilities and aim to quantify enhanced AO-occurrence following SSWs (ll. 79, 262ff). The explanations of RPI and FAR are slightly extended (ll. 251ff, 259).
- We extended the discussion of our results in the context of causality (ll. 534ff, 564ff). We explain that our results do themselves not allow strict causal statements. However, existing knowledge about pathways, involving the direct link between stratospheric and tropospheric extremes, but also teleconnections that act as common drivers, help to interpret RPI and FAR.

- further minor changes (see lines given in the discussions below and see differences file created via *latex-diff*)

Concerning the question of different statistical measures (RPI, FAR), we agree that they quantify, in different ways, the fact that AO- events occur more often following SSWs than without (see new ll. 79, 262ff). Our RPI definition does not include information about the "unexposed" part of the population, which is not well defined when applying a "forward" analysis: given SSW happened, what is increased likelihood for AO within certain time period following SSW (the only meaningful reference likelihood is base rate of AO occurrence, see ll. 242ff, 380ff, 396ff in manuscript). For our FAR estimates we take advantage of a "backward" analysis, where we change the time period in the event definitions: e.g., given SSW happened within certain time period, what is the likelihood for AO on the day following that period. This allows estimation of "unexposed" probabilities.

We are convinced that our FAR definition does provide a true estimate of the SSW effect on AO occurrence (=attributable risk), provided the following assumptions are met: 1) sufficiently large sample size, 2) unbiased S2S model, 3) common drivers can be ruled out. We believe that assumption 1) is certainly fulfilled and assumptions 2) and 3) are, although not completely, sufficiently fulfilled to the extent that a first estimate of FAR in the context of strat-trop coupling is feasible (as far as we're aware no previous such estimate exists).

We hope that future studies will be able to improve these estimates, in particular via addressing assumption 3). Uncertainties that arise from the influence of common drivers could be reduced by systematic conditioning (Kretschmer et al., 2021). Furthermore, "intervention experiments" can serve to isolate the SSW-AO pathway, which we indeed plan to perform and exploit ourselves.

Note that real world, "controlled" experiments would be needed to infer strict causality, but this is, of course, not possible. The same situation applies in epidemiology, where FAR was initially introduced. There the effect of exposure is impossible to assess from the fraction of population that has been exposed ("controlled experiments" are impossible as well). Instead an unexposed group is used to estimate the effect. FAR simply serves as a useful statistical measure to obtain an estimate of the effect. It turns simple conditional probabilities into more informative risk measures. Nevertheless, the usual caveats about any inferred causality apply (unaccounted-for common drivers, impossibility to "intervene" etc.) and we have revised the text further to be more clear about them.

Regarding comparison of our RPI and FAR, it is important to note that they refer to different event definitions. As pointed out by the reviewer, RPI compares AO occurrence following SSWs to climatological AO occurrence. FAR requires an "unexposed" AO likelihood, which we obtain by switching the definitions "SSW on day 0" to "at least one SSW within a time period" and "at least one AO within a time period" to "AO on the day after the (no-) SSW period". Therefore, the applied event definitions are themselves different, regardless of whether comparing AO occurrence to "climatology" (as in RPI) or to "without SSW" (as in FAR).
Second, and more importantly, we argue that our RPI and FAR aim to answer different questions, despite the fact that the two measures are indeed depended (because both rely on the increased likelihood of AO extremes following certain stratospheric events).
RPI quantifies the increased risk for an AO event following a SSW relative to its climatological occurrence. Hence, it ranges between -1 and 1/P(AO) - 1.
FAR is itself not a risk, but quantifies the fraction of events in addition to fortuitous SSW-AO co-occurrence. This fraction can take values from 0 to 1. Furthermore, FAR_p incorporates the SSW (=exposure) frequency, which is not the case for RPI and FAR_e.
Because of these differences in their definitions, a one-to-one relation between RPI and FAR is not possible and both measures require different interpretations (see, e.g., ll. 564ff).

Minor comments:
L11: expand ECMWF abbreviation

adapted.

L35-36: Add in "April"

adapted.

L169: change "quasi-deterministic forecast range of ten days" to "first ten days"

adapted.

L280: "process, indicating an AR1 process cannot reproduce the observed AO variability"
This statement does not make sense to me. You compare duration of AO in ECMWF with an AR1 process inferred from autocorrelation function based on observations.
How can you conclude from this comparison that AR1 cannot reproduce observed duration?

We added that the AO lag-1 autocorrelation agrees well between ERA5 (0.91) and ECMWF forecasts (0.88). The climatological occurrence of persistent negative AO phases is also very similar in the two datasets. The AR1 process shows, however, significantly more often short periods and less often long periods of negative AO, compared to both ERA5 reanalysis and ECMWF forecasts. We now write:

"[...], indicating an AR1 process cannot reproduce AO variability." (l. 206)

L313-315: The observed probability of AO-, given SSW, is not reported, but the confidence intervals are provided. Why not report the observed probability?

The observed probability for AO– given SSW fluctuates strongly with leadtime and is between ~0.55 and ~0.8, depending on the lag. The corresponding confidence interval is apparently also a function of lagtime. Due to the significant fluctuations, we do not want to draw too much attention on the reanalysis and therefore only provide a range of possible probabilities in terms of an overall estimated confidence interval.

L316: The observed baseline probability of AO-3sigma is not reported, what does it mean "modestly lower"?

We now provide the exact AO extreme base probabilities for ERA5: AO^{-3} = 0.06%; AO^{+3} = 0.02%. We added that the lower extreme probabilities in ERA5 are consistent with a negative kurtosis of the ERA5 AO distribution.

L345: Should not you multiply relative probability increase by 100% if you show it in Fig. 6 in percentages?

We added in the subtitle of Fig. 6 that the probability increase is shown in percent.

L348: Should not relative probability increase approach 0 in the limit of large t?

Yes, the limit of 1 referred to the fraction. To clarify, we changed to:

"The ratio is a function of the length of the time window t (see supplement Fig. S2). In the limit of large t, where the SSW influence becomes negligible, it is expected to approach 1, such that the relative probability increase approaches 0."

L392: In the denominator: P(noSSWwt)=1-P(SSWwt) adapted, thank you for spotting!

L449: Why only these two equations are numbered? Please number all equations.

adapted.

L478: the abstract report at least 35% increased AO- probability after SSW. Here, the number is 40%. You should be consistent.

adapted.

L504-505: Yes, causality in the stratosphere-troposphere coupling works in both directions, however, an AO event that occurs after an SSW can't cause the latter.

We have slightly modified the paragraph. The sentence was moved to a footnote and we added a note:

"It is important to keep in mind that the coupling is, in general, mutual and causality works in both directions (even though given events can of course not affect the past)."

L630: Allen et al. (2003) does not discuss FAR and is irrelevant to your study. The reference you need is: Allen M.R. (2003). Liability for climate change. Nature 2003, 421:892.

adapted, thank you!

Kretschmer, M., Adams, S. V., Arribas, A., Prudden, R., Robinson, N., Saggioro, E., & Shepherd, T. G. (2021). Quantifying Causal Pathways of Teleconnections, Bulletin of the American Meteorological Society, 102(12), E2247-E2263. Retrieved Jun 15, 2022, from https://journals.ametsoc.org/view/journals/bams/102/12/BAMS-D-20-0117.1.xml