Review of "Using large ensembles to quantify the impact of sudden stratospheric warmings on the North Atlantic Oscillation" by Bett et al.

By Erik W. Kolstad

First, I'd like to congratulate the authors on a very readable and scientifically sound paper. I enjoyed reading it, and I think it will be an important contribution to the field. It is very useful to see that the observed SSW frequency and post-SSW NAO signature (in reanalyses) is within the range of model's (natural) variability, and it was interesting to read about the (lack of) precursors. The two-thirds frequency of negative NAO after SSWs seems to stand up to scrutiny. It's also encouraging to see that the downward progression of anomalies, introduced about 20 years ago, still holds. Further, I'm a fan of your method (using large ensembles for climate variability studies), and I'm glad to see that it seems to be gaining traction, with several recent or upcoming papers.

Here are some mostly minor issues that should be addressed before the paper can be accepted.

In the introduction, you only mention the UNSEEN paper when you discuss previous studies. I think you should also mention some of the many other, including earlier, papers which have used a similar method (e.g., van den Brink et al., 2004, 2005; Breivik et al., 2013; Weaver et al., 2014; Chen & Kumar, 2017; Kent et al., 2017; Kelder et al., 2020; Wang et al., 2020; Spaeth & Birner, 2021; Brunner & Slater, 2022; Monnin et al., 2022).

I'm puzzled that you used ERA-Interim and not ERA5. I guess this is the reason that you stopped in 2018/19, whereas with ERA5 you could have used the last few winters as well. This should be justified. It's hard to see why you made this choice.

I think you should comment on the periods, which don't overlap exactly (1979/80 to 2018/19 for ERA-Interim and 1993/94 to 2015/16 for the hindcasts). There was along lull in SSWs in the 1990s, and for the first part of this period you don't have hindcast data. How many SSWs are there in ERA-Interim per year between 1979/80 and 1992/93, and from 2016/17 to 2018/19, compared to the frequency during the period for which the data overlap, and how might this influence your results?

What about detrending? You don't consider temperature, and I guess the PMSL and GPH trends might be negligible, but does it merit at least one sentence (i.e., why you don't detrend the data)?

I also have an issue with your definition of the NAO, although I know that some of you have used a similar index several times before. It seems strange not to standardize the northern and southern regions separately before you take the south-north difference. The variance in the northern region is higher than in the southern region and probably dominates your NAO index.

Figure 1: I struggled a bit to understand what was shown here. I think you should explain the fraction in panel a. The way I understand it this is the number of winters with at least one

SSW out of the 40 resampled ones, and then the count on the y-axis is the number of resampled time series in each bin (of 0.05 width). Please explain more thoroughly, so that the reader doesn't have to guess what the figure shows. Once you understand panel a, panel b is easier. Panel c and d though, are tougher. What I *think* it means is as follows. Panel c shows the 30d NAO anomaly across all the 545 SSWs in the hindcast winters, independent of resampling. In Panel d, you've first computed the mean SSW anomaly across all the 40 winters in each of the 1000 resamples, and then you show the distribution of these 1000 mean values. Please explain more thoroughly. (You should also consider using a dashed line for either the black or the red vertical line to avoid black and white and color-blindness issues.)

Harking back to the lack of references to similar papers in the introduction, after you do cite some of them, perhaps you should also discuss how your results agree or disagree with their results?

Other minor issues:

1. L74: Define "SPV".
2. L75: Define "PMSL".
3. L114: What does "standard deviation" mean here? Window?
4. Are you comfortable with using "tercile" do describe the data which is separated by the terciles? Strictly speaking, the "tercile" is the 1/3 quantile itself. I'd use "lower third" instead of "lower tercile", but this is probably a matter of taste.
5. L324: Replace "climate" with "conditions"?
6. L340: Would it be better to use "determinant" instead of "determiner"?
7. L414: Something went wrong with the dash in Andrew Charlton-Perez's name here.

References

Breivik, Ø., Aarnes, O. J., Bidlot, J.-R., Carrasco, A., & Saetra, Ø. (2013). Wave Extremes in the Northeast Atlantic from Ensemble Forecasts. *Journal of Climate, 26*(19), 7525-7540. https://doi.org/10.1175/JCLI-D-12-00738.1

Brunner, M. I., & Slater, L. J. (2022). Extreme floods in Europe: going beyond observations using reforecast ensemble pooling. *Hydrol. Earth Syst. Sci., 26*(2), 469-482. https://doi.org/10.5194/hess-26-469-2022

Chen, M., & Kumar, A. (2017). The utility of seasonal hindcast database for the analysis of climate variability: an example. *Climate Dynamics, 48*(1), 265-279. https://doi.org/10.1007/s00382-016-3073-z

Kelder, T., Müller, M., Slater, L. J., Marjoribanks, T. I., Wilby, R. L., Prudhomme, C., . . . Nipen, T. (2020). Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *npj Climate and Atmospheric Science, 3*(1), 47. https://doi.org/10.1038/s41612-020-00149-4

Kent, C., Pope, E., Thompson, V., Lewis, K., Scaife, A. A., & Dunstone, N. (2017). Using climate model simulations to assess the current climate risk to maize production. *Environmental Research Letters, 12*(5), 054012. https://doi.org/10.1088/1748-9326/aa6cb9

Monnin, E., Kretschmer, M., & Polichtchouk, I. (2022). The role of the timing of sudden stratospheric warmings for precipitation and temperature anomalies in Europe. *International Journal of Climatology, 42*(6), 3448-3462. https://doi.org/https://doi.org/10.1002/joc.7426

Spaeth, J., & Birner, T. (2021). Stratospheric Modulation of Arctic Oscillation Extremes as Represented by Extended-Range Ensemble Forecasts. *Weather Clim. Dynam. Discuss., 2021*, 1-25. https://doi.org/10.5194/wcd-2021-77

van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., & Burgers, G. (2004). Improving 104-year surge level estimates using data of the ECMWF seasonal prediction system. *Geophysical Research Letters, 31*(17). https://doi.org/https://doi.org/10.1029/2004GL020610

van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J., & Burgers, G. (2005). Estimating return periods of extreme events from ECMWF seasonal forecast ensembles. *International Journal of Climatology, 25*(10), 1345-1354. https://doi.org/https://doi.org/10.1002/joc.1155

Wang, L., Hardiman, S. C., Bett, P. E., Comer, R. E., Kent, C., & Scaife, A. A. (2020). What chance of a sudden stratospheric warming in the southern hemisphere? *Environmental Research Letters, 15*(10), 104038. https://doi.org/10.1088/1748-9326/aba8c1

Weaver, S. J., Kumar, A., & Chen, M. (2014). Recent increases in extreme temperature occurrence over land. *Geophysical Research Letters, 41*(13), 4669-4675. https://doi.org/https://doi.org/10.1002/2014GL060300