Weather Clim. Dynam. Discuss.,
https://doi.org/10.5194/wcd-2019-16-RC2, 2020

Interactive comment on "The role of North Atlantic-European weather regimes in the surface impact of sudden stratospheric warming events" *by* Daniela I. V. Domeisen et al.

Anonymous Referee #2

Received and published: 17 February 2020

Summary

The study asks whether the North Atlantic European (NAE) weather regime present at the onset of an SSW has a bearing on the subsequent evolution of the tropospheric state. The topic fits within the scope of WCD and a study on this topic fits more broadly into a body of literature that has investigated possible factors to explain why some SSWs appear to couple to the troposphere and others do not.

The study uses ERA-Interim reanalysis data (26 SSWs) and views these through the lens of 7 NAE weather regime types introduced in earlier work by one of the co-authors. The authors conclude that European Blocking at the time of SSW onset favours Greenland Blocking in the subsequent weeks, while Greenland blocking at onset favours an Atlantic Trough following the SSW. One major limitation of the study is the small sample of observed SSWs, which are then subdivided across the 7 regimes. This leaves only small samples for each subcategory. Earlier studies provide a cautionary tale about interpreting small subsets of SSWs (e.g., Mitchell et al. (2013) and Maycock and Hitchcock (2015)), and the authors fall foul to some of these issues.

The authors undertake bootstrap analyses to test for the significance of results, but this is mainly comparing to samples drawn from non-SSW periods. If the purpose is to test whether knowledge of the NAE state at the onset of an SSW can provide additional knowledge over and above knowledge of an SSW, the null hypothesis should be either that the tropospheric state following SSWs with a given day-0 regime is not distinguishable from that for all SSWs and/or that it is not distinguishable from SSWs with a different day-0 regime. This requires calculating differences (and their significance) between the regime subsets.

The authors also make no attempt to rule out other confounding factors that might affect their interpretation of the role of NAE regimes. For example, studies have found a relationship between the amplitude of lower stratospheric anomalies around the onset and the subsequent tropospheric NAM response. This was also pointed out by reviewer 1, but I think it is hugely important for the interpretation of the present results. The manner of presentation implicitly assumes the differences are a consequence of the day-0 regime, but since no other factors are tested for or displayed it is impossible to determine whether this is the case. This is especially pertinent given the small sample sizes being dealt with.

Overall, while the topic itself is potentially interesting, I found the manuscript disappointing both in terms of setting out the motivation for why/how the NAE state could have a long-lasting impact on the subsequent response and in terms of weaknesses in the analysis that I did not feel support the conclusions for the added value of knowing an SSW has occurred AND the day-0 NAE regime as compared to simply knowing an

SSW has occurred. I therefore recommend to reject the manuscript in its current form.

Recommendation: Reject.

Major comments

1) Hypotheses and statistical tests.

a) Your statistical test in Fig. 2 and 3 asks whether the SSW periods are different from non-SSW periods (climatology). This is fine for Figs 2a and 3a, where you ask about the overall signal of SSWs compared to no-SSWs, but what you are asking in Fig 2b-d is whether knowledge of the day-0 NAE regime provides extra information over and above the general knowledge of an SSW. Step 1) is there an SSW? Step 2) if yes, what is the NAE regime? Therefore, to my mind the relevant test is whether panels (b-d) are different from each other and/or different from (a). The same applies to Fig. 3. See e.g., important lessons from a parallel case on whether split vs. displacement SSWs show different coupling. Mitchell et al. (2013) performed similar analysis to that here for the NAM, but instead stratifying events based on split and displacement types (rather than on NAE type); importantly they neglected to test the significance of their differences, which was later done by Maycock and Hitchcock (2015) who estimated that the difference is not significant. You could do something similar here constructing a bootstrap distribution of the difference between two sets of N SSW samples.

b) Fig. 3: These dripping paint diagrams are notoriously sensitive to sampling uncertainty and for such small sample sizes I strongly question their representativity. Charlton and Polvani (2007) stated in relation to their assessment of the impacts of split and displacements (p.462, Section 6) “We started our analysis by first constructing time–height composites of the NAM index for the two types of SSW. However, the structure of the NAM index for the two types of SSW was found to be extremely sensitive, particularly in the troposphere: the size and timing of the composite NAM index anomalies following the events could be substantially altered by adding or removing even a single event. Hence, composite time–height NAM plots could not be used to examine dif-

ferences in tropospheric impact between the vortex splits and vortex displacements.” They made this point in relation to splits and displacements which have bigger sample sizes than those considered here. A similar point was also made by Maycock and Hitchcock (see e.g., their Fig. 3). Charlton and Polvani (2007) instead use the integrated NAM index to assess differences between splits and displacements. You could try an approach along these lines instead.

c) No attempt is made to rule out other possible associations than the NAE regime at day-0. For example, what if there is an indirect relationship to some other factor, such as the amplitude and persistence of the stratospheric anomalies themselves (e.g., Karpechko et al., 2017). There is some hint in Fig. 3 that the character of the stratospheric anomalies is different for these particular subsets of events; might that not be important? The sample sizes available here are very limiting in being able to say what is going on. To my mind, other more effective studies on related topics of downward coupling have combined reanalysis and model results (Karpechko et al., 2017; Maycock and Hitchcock, 2015). Reviewer 1 talks about following this up with a study on S2S models. If the authors do plan this, my recommendation would be to combine the current results with such a model study.

d) L193-196 “The immediate positive geopotential height anomalies and the weak tropospheric response in the aftermath of the event are archetypal for SSWs with GL at the onset and not the result of cancellations in the composites. Indeed, they are also evident for individual SSWs, such as the SSW on 8-Dec-1987 that exhibited a dominant GL regime for an extended period around the SSW onset (Figs. 4a and 6a).” I find Figures 4 and 6 completely uninformative. You have chosen examples to support your proposed hypotheses, but the key information is what comes from the behaviour across all events, as shown in Fig 2 and 3. To give just one example, you could have chosen instead the GL event on 9 Feb 2010 which shows the GL regime for 3 weeks after the onset. Presumably this event is not “archetypal” but it is one of your 5 cases. I would also argue you cannot conclude something is “archetypal” when you have only 5 events. I suggest removing these arguably cherry picked case studies and providing more comprehensive evidence for a detectible difference between the subsets discussed would improve the manuscript. This also applies to the discussion of the 2018/2019 events, which I found too cursory and descriptive to provide any real insight.

2) Existence of plausible mechanism(s). L53-57: “Given the large tropospheric internal variability and the influence of other remote effects mentioned above, it appears plausible that also the tropospheric state at the time of occurrence of an SSW plays an important role in shaping the characteristics of its downward impact. For example, tropospheric jet characteristics have been suggested to affect the downward impact of SSW events (Chan and Plumb, 2009; Garfinkel et al., 2013).”

I appreciate the goal of the study is not to explain but rather to diagnose, but this point is central to the whole premise of the study. However, the studies cited here are highly idealised and explore a much wider range of basic states than is plausible for the real world in idealised models that do not produce the type of NAE regime behaviour described here. I therefore do not agree this is supporting evidence for the proposed hypothesis. Indeed, no mechanism or theory is provided to justify why, or in what way, the tropospheric NAE state at day-0 would influence the subsequent NAE state up to +60 days. If that is the motivation to pursue this analysis, then some hypothesis for a mechanism is needed to explain an effect that extends far beyond the characteristic decorrelation timescale of the NAE circulation. It appears the proposal is for a vague mechanism related to internal tropospheric dynamics. However, this seems to defy the premise of why SSWs are useful for predictability in the first place, which is that their intrinsic timescale is much longer than the ‘memory’ of the tropospheric circulation. To make a more convincing case for this, more discussion is needed around the persistence characteristics of the regimes themselves and the canonical transitions amongst the regimes to put the behaviour following SSWs into context.

3) Timescales. Related to 2), a more careful description of the relevant timescales is

needed in the introduction. Since the downward influence of SSWs may last for up to 6-8 weeks are the authors proposing that the NAE regime on day 0 bears some relevance for the response in week 6? Or are the authors talking about the downward coupling over a shorter period following the onset, e.g. in week 1? This does not become clear until one gets into the results, so some explicit statements on timescales in the abstract and introduction would clarify this and this should tie into the discussion of mechanisms.

4) Dataset. Why is ERA-Interim used and not a longer reanalysis like JRA-55 which contains more SSWs (41 compared to 26 in Butler et al (2017))? For rare events, the benefits of increased sample size can outweigh other uncertainties in the pre-satellite era (Hitchcock, 2019).

Hitchcock, P.: On the value of reanalyses prior to 1979 for dynamical studies of stratosphere–troposphere coupling, Atmos. Chem. Phys., 19, 2749–2764, https://doi.org/10.5194/acp-19-2749-2019, 2019.

5) Other studies. The introduction ignores important information on past efforts (and their degree of success) in identifying stratospheric factors that may influence downward coupling, e.g.:

Charlton, A.J. and L.M. Polvani, 2007: A New Look at Stratospheric Sudden Warmings. Part I: Climatology and Modeling Benchmarks. J. Climate, 20, 449–469,https://doi.org/10.1175/JCLI3996.1

Karpechko, A.Y., Hitchcock, P., Peters, D.H.W. and Schneidereit, A. (2017), Predictability of downward propagation of major sudden stratospheric warmings. Q.J.R. Meteorol. Soc., 143: 1459-1470. doi:10.1002/qj.3017

Mitchell, D.M., L.J. Gray, J. Anstey, M.P. Baldwin, and A.J. Charlton-Perez, 2013: The Influence of Stratospheric Vortex Displacements and Splits on Surface Climate. J. Climate, 26, 2668–2682, https://doi.org/10.1175/JCLI-D-12-00030.1

Maycock, A. C., and Hitchcock, P. (2015), Do split and displacement sudden stratospheric warmings have different annular mode signatures?, Geophys. Res. Lett., 42, 10,943– 10,951, doi:10.1002/2015GL066754.

Nakagawa, K. I., and Yamazaki, K. (2006), What kind of stratospheric sudden warming propagates to the troposphere? Geophys. Res. Lett., 33, L04801, doi:10.1029/2005GL024784.

Runde, T., Dameris, M., Garny, H., and Kinnison, D. E. (2016), Classification of stratospheric extreme events according to their downward propagation to the troposphere, Geophys. Res. Lett., 43, 6665– 6672, doi:10.1002/2016GL069569.

Seviour, W. J. M., Gray, L. J., and Mitchell, D. M. (2016), Stratospheric polar vortex splits and displacements in the highâĂŘtop CMIP5 climate models, J. Geophys. Res. Atmos., 121, 1400– 1413, doi:10.1002/2015JD024178.

Specific comments

L6: following the weeks after an SSW –> in the weeks following an SSW

L45-49 you need to add e.g., to these reference lists as they are highly selective

L54 remove 'also'

L55 occurrence of an SSW also plays

L65-67 "Prior work based on this extended regime definition revealed important differences in the surface weather response to the state of the stratosphere, which remain hidden using the canonical four NAE regimes (Papritz and Grams, 2018; Beerli and Grams, 2019)." It seems this needs expanding as this is important justification for the current approach of using seven regimes rather than four. What specifically is missed? Also what did Papritz and Grams, 2018 and Beerli and Grams, 2019 show in relation to the two questions investigated here? Did they analyse similar things? What did they find?

L113-115 These statements are not visible from Figure 1 without some specific information on frequencies given in the text or in a table. Also in Figure 2a I see a peak in AR between -20 to -10 days but I cannot see a clear higher frequency for ScTr compared to all the other states. Rather the peak in EuBL immediately before onset seems to be a clearer feature for "all events".

L138-139: All the subsets you are dealing with are small sample sizes. The no regime cases can provide a useful null hypothesis, i.e. what anomalous regime frequencies can apparently arise without a specific regime being identified at day 0?

Fig. 2 and A2: The choice to use 5-day running means and to test for significance on that basis has implications related to the intrinsic persistence characteristics of each regime. But these timescales differ – e.g., from Fig. 1 it appears the canonical persistence timescale of GL is longer than, say, EuBL. It needs to be mentioned how the authors have accounted for the intrinsic persistence of each regime in choosing the smoothing window. Also, do you need to account for autocorrelation in your statistical tests?

I find Figure A2 more informative than Figure 2 since what you wish to highlight is the anomalous frequencies associated with particular subsets of data not the absolute frequencies. This is more clearly seen in Figure A2. For example, it becomes clear that the significant anomalies in AT at lags -35 to -15 is because the frequency is anomalously low (i.e. a negative anomaly). I suggest switching them and putting Fig. A2 in the main text and Fig. 2 in the Appendix.

L163-164 "Given the strong influence of the tropospheric state at the time of the onset of an SSW on the weather regime frequencies in the subsequent days" I don't agree you have demonstrated this in Section 3. See major comment 1.

L170-171 These are weaker thresholds than one would typically associate with "robust" and "highly robust"

L 179-187 and Fig. A3: Why are the tropospheric Z anomalies so weak? Is it a matter of plotting (e.g., contour intervals)? The SSW compendium NAM composite for the same set of events in ERA-Interim (see Fig. 1 below) looks quite different from your Z anomalies (Butler et al., 2017).

Fig 5: Is any statistical testing applied to the anomalies? The caption does not mention it.

Typographical

Figures – the two shades of green for EuBL and ScBL are hard to differentiate

---

Interactive comment on Weather Clim. Dynam. Discuss., https://doi.org/10.5194/wcd-2019-16, 2020.

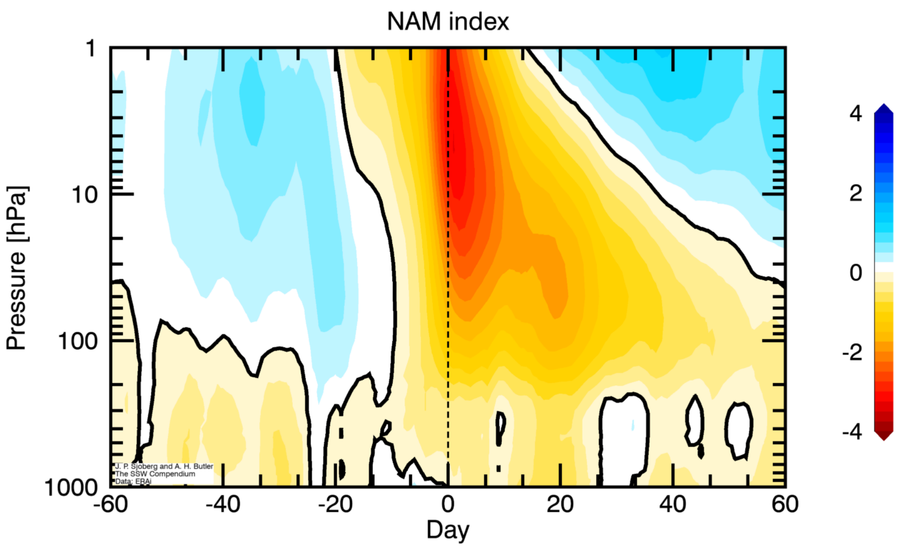

Fig. 1. SSW compendium composite NAM anomaly for SSWs in ERA-Interim (Butler et al., 2017).