Response to Reviewer Comments

"Origins of Multi-decadal Variability in Sudden Stratospheric Warmings" by Oscar Dimdore-Miles et al.

We thank all the reviewers for providing their comments on our analysis. Their questions and suggestions have helped us to consider the role of ENSO and the PDO in multi-decadal SSW signals more closely as well as make our description of our wavelet methodology and the interpretation of wavelet plots clearer to the intended reader.

Summary of major changes

- Additional analysis using multi-linear techniques (new section 3.2) to explore the comparative roles of ENSO, AL and QBO forcing (Tables 1-3) ; the results support the wavelet results and strengthen our conclusions on the role of the QBO amplitude modulation.
- Section 2 (previously 'Model and Data') renamed 'Methodology' to better reflect its content; now includes an improved description and justification of the wavelet technique;.
- All wavelet analysis figures have been replotted using consistent colour scales for easier comparisons.
- Better justification of which months have been analysed, including an extra supporting figure showing results for NDJFM.
- Improved discussion of the potential role of the PDO and an additional supporting figure.

The authors analyze the possible causes that lead to extended periods with and without Stratospheric sudden warmings. They find that such events vary on multi-decadal timescales of period between 60 and 90 years, and that signals on these timescales are present for approximately 450 years of a 1000-year long simulation. While tropical sea surface temperatures and Aleutian Low variability seem relatively unimportant, the amplitude of the stratospheric quasi biennial oscillation (QBO) westerlies in the mid stratosphere between 15hPa and 30hPa seems more important. This paper has some interesting results, and it is very likely that the paper will be suitable for publication after revision. However I must admit that I am not an expert in the methods the authors use, and I had some questions concerning the manner in which the authors draw conclusions from their results.

Major comments: 1. This reviewer has limited experience with the chief methodology used in this paper, and I suspect that most of the intended audience has a similar lack of familiarity. While advanced methods can help uncover relationships that would otherwise be missed, the interpretability of the resulting effect is often missing. While I appreciate the effort the authors put it to interpreting the results of the wavelet analysis, there are some (perhaps very basic) questions I had.

We agree that many will be less familiar with wavelet analysis compared with multi-linear regression. To address this, we have added a new section (3.2) where we show some multi-linear regression analysis (Tables 1) and discuss its limitations, as a lead-in to the section on wavelet analysis. We have also added further regression analysis (Tables 2-3) that provide supporting evidence for the wavelet analysis. Additional text in our introduction and at the end of this new section 3.2 further motivates the use of wavelet methods.

At present the arguments concerning the relative importance of ENSO and the QBO are hand-wavy. The authors claim near line 325 that the ENSO signal is weak in Figure 8. However by eye, the ENSO signal in Figure 8 near the 90 year periodicity is actually little different from the QBO signal in Figure 12. There is no colorbar on either figure, but the shade of red shown is similar, and is located in a similar location. Maybe if I squint I can see the authors point, but there must be a better way to quantify the relatively importance. Is there a way to compute some sort of different plot between figure 12 and figure 8 to more robustly make the point?

Apologies for this lack of clarity - each spectra was normalised by the variance of the time series which was not clear from our text (this is now rectified in section 2.3), and our interpretation was based primarily on the region of statistical significance. We have clarified the text and improved the figures so that all wavelet spectra use the same shading levels (with the colorbar added). It is now much more evident that ENSO exhibits significantly weaker power at the relevant periods compared to the QBO metric.

The simplest way to evaluate this point using more classical techniques is with simple regression. You compute the regression coefficient (including uncertainty) between the QBO signal and SSW_5yr, and between ENSO and SSW_5yr, and simply compare the regression coefficients.

As described above, we now include multilinear regression analyses comparing the contributions of ENSO, the AL and the deep QBO amplitude to SSW_5yr – see the new tables 1-3 and accompanying discussion.

What exactly is the ENSO signal at a periodicity of 90 years? In addition to the modes of variability considered by the authors, there is an additional oceanic mode: the PDO or IPO (Pacific decadal oscillation or Interdecadal Oscillation). The periodicity in observations is a bit shorter than 90 years, but perhaps in the authors' model the periodicity is longer. The PDO and IPO project strongly onto ENSO, and indeed one leading forcing of the PDO is simply a low-passed version of ENSO (Newman et al 2016). The PDO has been linked to vortex variability (Rao et al 2019, but see the papers cited therein). Please clarify whether the link seen in Figure 8 is just the PDO.

Thank you for raising this possibility - the PDO was not discussed sufficiently in our analysis. We have added text to note the possibility that the ENSO signal at 90 years may be a manifestation of the PDO (lines 398-401). We have added a supporting figure (A3) to discuss its importance. However, while we agree that it merits further investigation, we believe this to be outside the remit of the current study, given that it does not account for a large portion of SSW_5yr power in our model (and not as much as the QBO metrics can).

I found the QBO index the authors choose a bit strange. The strongest HT relationship on interannual timescales is with the QBO near 40hPa or 50hPa and the vortex. If one focuses on winds near 20hPa, the HT relationship more or less goes away entirely on interannual timescales. The authors claim that on longer timescales, the most important QBO phase is winds between 15hPa and 30hPa, not winds lower in the stratosphere. While it is possibly conceivable that the phase on short and long timescales isn't identical, this much of a mismatch in phases is disturbing, at least to this reviewer. Do the authors have any ideas on *possible causes for such a shift in phase? This mismatch casts some doubt on the robustness of the results, as at present the search for the "best" phase seems too much like a data-fishing exercise.*

Our rationale for focusing on the deep QBO metric follows from the work of Gray et al. (2018) and Andrews et al. (2019) who demonstrated a stronger surface response to the QBO using this metric compared to single levels (and ultimately, we are interested to understand not only the source of long-term vortex variability but also long-term surface variability). We have added some additional text to section 3.1 to note that there is no reason why the deep QBO composite should look like the average of the corresponding single-level composites. We do not believe that the lack of a clear HT response to the single-level 20 hPa QBO is a problem: let us assume that a deep region of QBO easterlies between 15-30 hPa is important for development of an SSW and hence weakened vortex. Selecting years simply based on the 20 hPa will indeed select those years with a deep easterly QBO (and these years will indicate a weakened vortex) but it will also select years with strong vertical shear above and/or below that level, which is likely to obscure the vortex response (since our assumption is that a weak vortex response requires the absence of vertical shear), thus explaining the null response to this single level. (We also note that even though there is no strong HT signal from the 20 hPa level, the 20 hPa timeseries exhibits substantial amplitude variability in the westerly phase, hence our use of the Hilbert amplitude).

Minor comments:

Line 21: there are earlier studies showing an impact of SSW on cold snaps. See Thompson et al 2002 and Lehtonen and Karpechko 2016.

Both references added.

Line 32: Using a subset of the simulations from Garfinkel et al 2017, Garfinkel et al 2015 actually found a slight strengthening of the vortex in late winter due to SSTs over the 30-year period from 1980 to 2009.

Reference to this work and the possible role of SST forcing of the vortex on decadal timescales included (lines 35-37).

Line 52: Garfinkel et al 2018 also showed results regarding the HT effect in GloSea5, which is based on HadGEM3 GC2.0 if I understand the S2S website correctly. The model performed well as compared to reanalysis.

Thank you, reference added (line 57).

Line 75: The importance of the Aleutian low for stratospheric wave driving was shown earlier than 2015. See Manzini et al 2006, Taguchi and Hartmann 2006, and Garfinkel et al 2010 The former two studies didn't focus on the Aleutian Low per se, but show this effect.

References added (line 96)

Line 93-94: A competition between Pacific and Indian Ocean SSTs for El Nino was shown earlier by Fletcher and Kushner

References added (line 100)

Line 214-215 Is the Aleutian index defined for autumn only?

This was a typo and should read Dec-Mar as this is the month range used in the figures. We tested a range of Autumn and Winter month ranges and the spectra are relatively robust across the season on the timescales we consider (60-90 years). Dec-Mar covers the same months used to define SSWs and was a natural choice to use encompassing mid-late winter. We have added a statement of this testing of different month ranges (lines 253-256).

Line 339 states you used DJFM. Please clarify
Mention of Sep-Nov on lines 214-215 is incorrect - Dec-Mar was used throughout – corrected.

Line 229: Actually the seasonal evolution of SSWs doesn't match observations all that well (non-overlapping error bars), with too many SSWs later in winter as compared to DJF. Such a model bias is fairly common however, and Horan and Reichler 2017 argue that the observed seasonal distribution may reflect sampling uncertainty.
Additional discussion of the seasonal evolution has been added, including November SSWs, and reference included (lines 270-277).

Caption of Figure A4 and the figure itself don't seem to match. Please correct. For now, the claim near line 352-353 stands unsupported.
The caption has been rewritten and the statement on 431 in the revised manuscript is now supported.