Reply to Referee #1

General Comments This paper examines the future increase/decrease of SSWs due to anthropogenic forcing in a large ensemble of models (12) from the CCMI database. Previous studies have had opposing views on this with some finding an increase in the number of SSWs whilst others have found a decrease. To check the robustness, they use three different commonly-used SSW criteria and conclude that they do not find a significant increase in the future number of SSWs over the course of the 21st century. I find the results to be interesting in that, by using a much larger number of models than previously used, the overall number of SSWs will remain approximately the same. The paper is well-written and the topic is well discussed, hence I suggest only minor corrections which I list below.

Thanks a lot for your comments. Here is our reply to the specific and technical comments in black:

Specific Comments

1) Lines51-52; not just due to anomalously large injections of wave activity from the troposphere, but can also be due to the resonance of wave activity such as that first described in Plumb (1981) and shown by Esler and Scott (2005). I would either cite this second mechanism, or just change the wording in your current sentence to be less definite.

We have modified the sentence to be less definite.

2) Line 53-54; Small point, but not all SSWs can impact the troposphere (e.g., Gerber et al. 2010, GRL and Hitchcock and Simpson 2014). I would instead just add the word 'can' into the sentence to change it from meaning all as it is currently: 'SSWs can also impact the. . .'

Done

3) Line 151; which criteria are absolute and which are relative? Can you somewhere distinguish between the two – from what I can see, the U6090N and WMO are absolute and ZPOL is relative.

It is stated at the end of Section 2.2 (old Line 136) which criteria are absolute and which one is relative. Nevertheless, we have also included a reminder in Section 3.1 (Line 163 in the marked-up manuscript).

4) Line 184-185; at which confidence level do the models simulate statistically significant differences between the SSW duration in the past and in the future? Can you make a comment here on the 90% level? Perhaps just say how many models would become significant if this level was used. My guess is that the HADGEM3-ES and MRIESM-1r1 would be significant at close-to the 95% level as the error bars are nearly separated.

In the case of the multimodel mean, the future change in the duration of SSWs is only statistically significant at the 80% confidence level. When examining each individual model, only HadGEM3-ES shows a statistically significant change in the SSW duration at the 90% confidence level. Apart from HadGEM3-ES, the referee also suggests that the future change for MRI-ESM1r1 might be also statistically significant at the 90% confidence level. However, in that case, this difference would be statistically significant at the 71% confidence level. Probably, the referee got that impression because the error bar of the past extends up to the mean value of the future period and so, it is not easy to identify the end of the first bar.

Following referee's suggestion, we have made a comment on the model that simulates statistically significant differences in the SSW duration between the past and the future at the 90% confidence level (L199-200 in the marked-up manuscript).

5) Line 191; the eddy heat flux (v'T') is what you plot right? This is a proxy for the injection of wave activity. Can you make this clearer? Further, I gather from figure 3 that you use 100hPa (I think this should be included here in the text also), but given the recent paper by de la Camara et

al. (2017) and Birner and Albers (2017), who suggest that 100hPa is not an ideal surface to use as it is already in the bottom of the vortex, how sensitive are you results to the level of choice? Do you get more significant results if you use a slightly lower level?

The referee is correct. Figure 3 shows anomalous eddy heat flux at 100hPa and averaged between 45º-75ºN (aHF100). We have described that in Section 2.3 and included a small comment in new Section 3.4 to make it clearer (L217-218 in the marked-up manuscript).

We acknowledge that these 2 recent papers show that 300hPa is maybe a better level to use than 100hPa. However, 100hPa is the traditional metric and our choice is in line with all previous work. More importantly, we do not have output at 300hPa for some models and so, we cannot do these calculations.

6) Section 2.2 and Line 197; I would be interested to see what the changes between the number of splits and displacements are between the past and the future. From figure 3 it appears that the wave 1 and wave 2 forcing through 100hPa doesn't change too much between each period and so the number of splits and displacements may not change too much either. But given the relatively short length of the paper as it is, this would be an interesting addition.

Thanks for the suggestion. We had indeed started looking at the number of split and displacements SSWs in the past by applying a similar algorithm to Charlton and Polvani (2007). However, the results showed a bias of most models towards an unrealistically high number of displacements events, probably due to a too strong climatological wavenumber 1 wave component in December and January. Thus, given that models could not realistically reproduce the distribution of split and displacements SSWs in the past, we decided not to explore this further. However, figure 3 of the manuscript suggests a null change in the number of splits and displacements as the referee indicates. In the revised version and based on referee's suggestion, we have included a short comment about this when describing figure 3 (L224-225 in the marked-up manuscript).

Technical Comments/Grammatical Errors

1) Line 56; on → at
Done.

2) Line 82; The recent paper by Kim et al. (2017) which you cite later may be a good citation here.
Included!

3) Line 111; Could you clarify this sentence to say whether both reanalyses extend back to 1979, or that both extend back to before 1979 (and if the latter, then which year: 1960?).
They extend back to before 1979, in particular JRA-55 data starts in January 1958 and the ERA-40 reanalysis data is available since September 1957. We have included this information in the manuscript to clarify the mentioned sentence.

4) Line 125; Is the Polar Cap area weighted? I think it should be and it would be good to include a sentence here saying so.
Yes, the polar cap is area-weighted. We have included a comment about that in L134 in the marked-up manuscript.

5) Line 191; 'in the course' → 'during the course'. Also, aHF100 in the text should correspond to the figure title of HF100.
We have modified both things.

6) Line 193; None of the individual models show significantly different results?

We have only found some slight significant changes in two models (GEOS-CCM and EMAC-L47), but only for a few days preceding the SSWs. Thus, we think that it was not worth important to report in the paper.

7) Line 209; 'no statistically significant changes'

Modified!

8) Line 220; Can you give references to the figure which shows this? Figure 1?

Yes, it is Figure 1. We have included the reference to that figure.

9) Line 221; 'across' → 'using'

We prefer using 'across' instead 'using' as the slight future increase in frequency of SSWs is detected in all cases when applying each criterion separately.

10) Line 260; 'in the last years' → 'in recent years'

Changed

11) Line 261; 'metrics' → 'metric'

Changed

References

Charlton, A. J., and Polvani, L. M.: A new look at stratospheric sudden warmings. Part I: Climatology and modelling benchmarks, J. Climate, 20, 449-469, 2007.