Response to Reviewer 3 Comments

Overall Comment: *I appreciate the authors providing detailed responses to my previous comments, which have helped me to understand the paper in more depth. The proposed analytical technique has the potential to become a new definition of SSW. However, to confirm this method as a robust and standard definition of SSW, at least several major/minor SSW events must be additionally tested. Based on the responses to Points 2, 4, 16, and 20, this study is inconclusive, where further refinements are required, namely, by investigating a longer-data record. I am concerned, for example, with the response to Point 4, as the selection of TEA parameters, the corresponding height ranges, and their thresholds may be modified depending on each SSW event, and thus could produce ambiguity in defining major, minor, and final warmings.*

I would like to suggest that the authors continue the analysis of several additional major/minor SSW events to make this method more concrete and to demonstrate fully the usefulness of the novel GNSS-RO data. However, if the authors are insistent about introducing this method to the SSW community in its present version, I look forward to a follow-up study in which the method is applied to a long-term dataset.

Response to overall comment: We thank Reviewer 3 for this overall comment. And yes, we agree that any such new method should be tested with longer-term records. This is why we did refer in this initial intro paper's title to "a new approach demonstrated by the 2009 event", and we confirm we have successor work on-going for the follow-on paper that completes the introduction as a consolidated method, based on long-term RO and ERA5 records with a whole ensemble of SSW events over the recent decades. We found that an "all-in-one single paper" would be really lengthy and we therefore indeed prefer to be "insistent" with this two-papers approach.

Regarding the possible "tuning needs" of various TEA parameter settings, etc., we confirm in view of our preliminary follow-on paper results on a range of SSW events in the RO timeframe 2006 to 2019 that the new approach is robust and that detailed parameter settings can be consolidated without changing the method's basic design. This most up-to-date knowledge we meanwhile have from the follow-on work is in line with what we state in the conclusions of this paper, i.e., that we expect such refinements to work and that we will do such consolidation using long-term records.

Specific comments to the responses

Point 1: *In addition to the statements cited in the author's response, Butler et al. (2015)*

also refers to some useful points about the SSW definition, including the technique's statistical application to SSWs, consistency across observational and modeling studies, and detection of historical SSWs. These suggest that any new method should be tested with long-term records.

The following statement in the concluding section (Butler et al., 2015) provides a major summary of the new SSW definition.

"We believe a new definition should include, at a minimum, guidelines for determining a) the independence of closely timed events; b) the classification of split-type versus displacement-type events; and c) precise distinctions among major, minor, final, and Canadian SSWs."

Have these requirements been met in the current study by considering only the single SSW event from 2009?

Response to Point 1: We repeat that we agree that any new such method should be tested with long-term records; see our response to the overall comment above. An all-in-one paper would be lengthy and hence we prefer this two-papers approach. We also agree also that full compliance with these Butler et al. requirements partly needs the results from the longer-term analysis from multiple events; as we ourselves also say already in this manuscript. So overall the requirements are only met by both papers together (hence the first one is introducing the "new approach").

More specifically, regarding these requirements: a) "the independence of closely timed events", this is analyzed more in long-term processing, and we find as part of the follow-on work that our trailing-metric is quite helpful to this end; 2) "the classification of split-type versus displacement-type events", we find split-type events usually show larger anomalies, also this is part of the long-term analysis; 3) "precise distinctions among major, minor, final, and Canadian SSWs", we expect this to be quite robustly possible based on the refined TEA parameter settings and then using our three core indicators (Main-phase duration, Main-phase area, Main-phase strength) as well as optionally the trailing metric.

We have introduced a new small paragraph at the end of Section 3 to explicitly address now the point also that these aspects of the Butler et al. (2015) requirements will (only) be assessed in the follow-on work, but that we are aware of and intend to do so.

Point 3: *I agree that the zonal wind in a narrow latitude band at a specific height may not be sufficient to define SSW. However, the effects of SSW have been commonly investigated in terms of the behavior of the polar vortex and its effects on stratosphere–troposphere coupling. Therefore, examination of the wind field, including the*

circulation reversal and the characteristics of planetary waves, is critical and should probably be included in a SSW definition.

Response to Point 3: Thanks for the comment. We agree that polar vortex is strongly related to the SSW and that wind data as available can be very useful. However, we also know from many previous studies that the middle stratospheric anomalies, e.g., temperature anomalies, wind anomalies and also geopotential height anomalies are strongly related, and all related to the state and dynamics of the polar vortex. The polar vortex strength is a key factor for stratospheric downward effects on tropospheric circulations and is related to the strength of both temperature and wind anomalies.

Therefore, we see our approach with its temperature anomaly-related ansatz one good method that also from the data (availability) side has advantages over wind field use. Moreover, since ERA5 data also have wind field information, we may as part of the long-term analysis also analyze the relationship between anomalies in thermodynamic variables as used in our method and the polar vortex changes and wind reversals.

Points 4 and 16: *I understand that the thresholds of TEAs have yet to be confirmed, and they should be adjusted by repeating the analysis for a long-term dataset. This study utilizes various parameters obtained from GNSS-RO measurements over a wide height range, which are summarized as a total of 12 values in Table 1. I am not fully convinced by the selection of only five of the 12 parameters. Are the other parameters useless?*

Response to Points 4 and 16: In fact, Table 1 shows the methodology flow and that is why we list there quite a number of parameters. (1) to (3) shows how we calculated anomaly profiles for temperature, density and bending angle. Based on (1) to (3), we then have our five TEA values shown from (4) to (8). Then based on these five TEA values, we have our primary, secondary and trailing metrics ((9) to (11)) for SSW monitoring. (12) to (13) are temperature only metrics for readers who prefer only temperature for SSW detection. Based on metrics (9) to (11), we then offer three overall metrics shown in (14) to (16) as the main metrics for SSW definition and classification. Hence, all parameters along the sequence are useful, and the final ones are those metrics that are mainly used to track the SSWs characteristics in the long-term records.

Points 5 and 12: *Unfortunately, I do not fully understand the response. Provided GNSS-RO data is assimilated into the ECMWF, they both naturally provide nearly the same results. Do the authors suggest employing ECMWF for application of this method to the larger re-analysis data from the past decades, which is much longer than the GNSS-RO records?*

Response to Points 5 and 12: Yes, formally the method can be applied to any reliable gridded data and we used in this paper the RO and ECMWF data to demonstrate this. And specifically to the question: yes, as we briefly describe in the paper's conclusions, in our follow-on paper work we use both the GNSS-RO data and the longer-period (ERA5) re-analysis data, where the latter indeed strongly back-extend the record.

Point 6: *I am interested in the retrieval technique when the model atmosphere is not used. Please show the citation for the non-optimized method.*

Response to Points 6: Thanks for this comment. Based on Reviewer's suggestions in previous first review, we already used the non-optimized bending angle in our revised manuscript. Therefore, the model atmosphere is not influencing now anymore.

We again thank Reviewer 3 for his/her valuable comments.