General comments:

The paper is improved regarding the basis for the study and a description of the results. However, the English in the text still needs improving – in many places, it is clumsy, with vague statements made. As a result, it is difficult for the reader to follow the arguments made by the authors. In my view, until the authors address this key point for a scientific publication, the paper should not be accepted for publication in ACP.

Below please find examples of what the authors need to address.

L. 11: Important for what aspect of ozone? The distribution?

L. 14: search for -> study.

L. 21: Would "central areas" be better than "peak areas"?

L. 22: Remove "quite".

L. 25: Do you need "well-pronounced"?

L. 28: on -> at. Make this change elsewhere in the paper, so it should read: "…significant at the…level…".

L. 40: transportation -> transport.

L. 45: Perhaps provide here a reference to Manney et al. (2011) describing the unprecedented ozone loss in the Arctic for 2011.

L. 50: Introduce the acronym for BDC. Acronyms should be introduced when first used, both in the abstract and in the main text.

L. 53: Provide a reference for the concept of age of air, e.g., Waugh and Hall (2002).

L. 70: "Many studies have been working with…". This is clumsy - reword.

L. 75: Introduce the Dobson Units (DU).

L. 77: "…some effects…". This is vague – describe what they are.

L. 79: What are these reasons?

L. 83: What do you mean by a substantial distribution?

L. 93: Changed its trend from what to what?

L. 94: I suggest: "…main contributor to heating of the…".

L. 95: "...more or less…". This is vague – describe the effects on the behaviour.

L. 110: Results of what?

L. 127: Do you mean direct wind measurements from satellites? If so, specify.

L. 130: reanalyzes -> reanalyses. The plural of the noun "reanalysis" is "reanalyses". Change this elsewhere in the paper. You can use "reanalyze" as the US spelling of the verb.

L. 138: Specify the lower level of the model, presumably the ground. For example: "…28 levels from the ground to the top of the model at…". Do the same for L. 142 and L. 145.

L. 139: Do you need "efficiently"?

L. 151: Quantify the "very close"?

L. 153: I presume sufficient to use any of the three reanalyses for the period from 1970 to 2012.

L. 160: in -> at.

L. 167: Make sure the two-core structure has been introduced before in the text.

L. 169: "…at 10 hPa…".

L. 177-178: Clumsy sentence – reword.

L. 186-188: Clumsy sentence – reword.

L. 189: The comparison using different time periods is (a priori) not consistent. Please justify.

L. 202: What do you mean by "…distribution is much less compact"?

L. 216-217: The frontside and backside of something depend on the way one is facing. In the absence of a definition of what is the front and what is the back, this usage is ambiguous. By contrast, the concepts "westward" and "eastward" are well defined for the Earth, and are not ambiguous.

L. 226: "…and not the eastern.,."

L. 238: columnar –> column.

L. 239: Remove "evident".

L. 242: factor -> factors.

L. 245: Why do you use the NCEP/NCAR reanalysis?

L. 256: "…as well as the meridional…".

L. 261, 262: "…of the QBO…"; "…on the QBO…"; a bit -> slightly.

L. 286: predominately -> predominantly.

L. 292: Is using different time periods consistent? Justify this.

L. 289-324: Very long paragraph – consider splitting into 2 or 3 paragraphs.

L. 296-299: This sentence is not clear to me – reword.

L. 321: "…found an impact from that longitudinal…".

L. 322: Provide details of the impact.

L. 323: "…with the model…".

L. 324: This is vague – provide details of how this asymmetry is demonstrated.

L. 333: What do you mean by "sector not affected…structure"? In the context of the main sentence, what you mean is not clear to me.

L. 336: Provide details of how this is confirmed.

L. 338: from -> is.

L. 339: Provide details of this changed dynamical behaviour.

L. 347: "…for years with major SSWs than…".

L. 348: I suggest "abnormal" -> "unusual".

L. 349: "…during a SSW…".

L. 350: Have you introduced the concepts of split and displacement SSWs? If not, do so, and provide a reference, e.g., Charlton and Polvani (2007).

L. 358: Provide examples/references of these other authors.

L. 360: "…larger at higher…".

L. 361: Provide references for these previous studies.

L. 368: "…a well pronounced…".

L. 370: "…one in each…".

L. 376: What do you mean by the ozone content?

L. 378-379: How do your results illustrate the limitations of the zonal mean approach?

L. 383: "…in the cores…".

L. 384-385: "…it recovers.".

L. 385: What is the influence of the solar cycle?

References

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and Modeling Benchmarks. J. Atmos. Sci., 20, 449-469.

Manney, G. L., et al., 2011: Unprecedented Arctic loss in 2011. Nature, 478, 469–475, doi:10.1038/nature10556.

Waugh, D.W., and T.M. Hall, 2002: Age of stratospheric air: Theory, observations, and models. Rev. Geophys., 40, no. 4, 1010, doi:10.1029/2000RG000101.

Overall: The authors have made an effort to address my latest comments. However, I still think the paper is not suitable for ACP. I think the paper still suffers from the following shortcomings:

(i) Confusion – there is a lot of information, e.g., in the Introduction, but there is poor organization, the focus is unclear, and the arguments are difficult to follow;

(ii) Discussion over a narrow latitude band – despite the extensions made in the last version, the paper is still mainly about features over a very narrow latitude band, and which is not convincingly justified in the text;

(iii) Trends – fine, 95% is a standard significance level in meteorology, however, one wishes to test that the quantity estimated, in this case trends, is robust. One way of doing this is testing the significance level at 99%. In my view, the results for the trends are generally not robust;

(iv) The English – much better, but still there is use of vague and qualitative language.

I would add that there is evidence that the authors, in their response, agree (at least partially) with comments (ii) and (iii) above:

Narrow latitude band:

L. 153: The latitude band is 49N – 56N, a range of 7 degrees. Is this representative of mid latitudes?

A: *Title is modified to "higher midlatitudes" and some results from 42.5 and 62.5$_o$N are added (new Figure 2). The text is modified; you are right that such a relative narrow band need not be sufficiently representative for middle latitudes.*

My answer: I agree with you regarding the representativeness of the relative narrow latitude band.

Trends:

L. 243: Only 4 trends (out of 192, I understand) are significant at the 99% level. And as mentioned by the authors, this is likely due to the limited length of the datasets. Which begs the question, why calculate trends with a dataset limited in length?

A: *Well, any dataset is limited in length. The longer data series is not available and reliable (before 1970). We started in 1970 and we split this period in the mid-1990s to see possible impact of overturning of stratospheric ozone trend, which does exist. The significance level of 95 % is usually used in meteorology.*

My answer: Yes – any dataset is limited in length. Because of this, the length may not be enough to provide robust results for trends.

Table 1: Why show this table if significance at 99% only occurs for 4 cases?

A: *The standard significance level for analyses in meteorology (wind, temperature, etc.) is 95%. That is why we use this threshold. It was added in section 2 Data and methods. Data series are probably too short for getting significance at 99% in noisy atmosphere but due to change of trends in the mid-1990s they cannot be longer.*

My answer: If the dataset is not long enough, it may be the case that robust trends are difficult (if not impossible) to derive.

To summarize, before the paper is suitable for ACP, the authors should address the above issues, as well as the specific issues below.

Specific comments (not exhaustive, but illustrative):

L. 13: Introduce acronym for NCEP/NCAR – in the abstract and in the main text. Do the same for all acronyms.

L. 17, 22: "seems to", "appears to" are vague and, in my view, unsuitable for a scientific paper. Please avoid such language.

L. 79-85: No mention of trends.

L. 106: "slightly better" is qualitative. Please quantify your statements.

L. 177: "are similar in tendency" – no, they are not. There are negative and positive tendencies.

L. 202: "support a tendency…trends". Rephrase to give sense.

L. 234: At what level are the differences significant? I presume it is 95%, but this information should be provided to the reader. There are other parts of the text where this information is not provided.

Table 3: I suggest to the authors that this is the only robust result of the paper (significance at both 95% and 99% levels).

Fig. 3: Indicate in the caption what the straight lines are.