General comments:

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.